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. 63/570,065, 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 063586-541001WO_Seq-Listing_ST26.xml and is 1,678,063 bytes in size.

BACKGROUND

RNA-guided nucleases (*e.g.*, CRISPR-Cas nucleases) are a group of nucleases that can target specific nucleotide sequences mediated by an RNA guide, which interacts with the specific nucleotide sequence through base-pairing. Gene editing systems comprising such an RNA-guided nuclease are useful programmable tools for gene editing and therapeutic purposes. 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-RNA-guided nuclease gene editing systems for use in disease treatment.

SUMMARY OF THE INVENTION

The present disclosure provides gene editing systems mediated by reverse transcriptase-RNA-guided nuclease, which successfully introduced designed nucleotide substitutions into target genetic sites. In some embodiments, the gene editing systems disclosed herein involve a fusion polypeptide comprising an RNA-guided nuclease fragment and a reverse transcriptase (RT) fragment, and optionally one or more nuclear localization signals (NLS) and/or peptide linkers. The RNA-guided 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 an RNA-guided 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 RNA-guided nuclease may be the reference Nuclease BT of SEQ ID NO: 1 (or SEQ ID NO: 140, with N-terminus M residue removed) or a variant thereof, such as those disclosed herein. The RT polypeptide can be any of those provided herein (*see*, **Table 25** below), for example, an MMLV-RT polypeptide.

In some embodiments, the RNA-guided 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 nickase 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; or (iii) a combination of (i) and (ii).

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 H231, H232, H255, D67, E176, and/or D329 of SEQ ID NO: 1. In some examples, the one or more nickase mutations are located at positions H231, H232, and H255 relative to SEQ ID NO: 1. For example, the mutation at H231 is an amino acid substitution of H231A, H231G, H231L, or H231S; the mutation at H232 is an amino acid substitution of H232A, H232G, H232L, or H232S; and the mutation at H255 is an amino acid substitution of H255A, H255G, H255L, or H255S. In some specific examples, the mutation is at position H232 (*e.g.*, H232A or H232L substitution). In one specific example, the mutation is at position H232, *e.g.*, H232A.

The RNA-guided nuclease polypeptide disclosed herein comprises a PLMP domain, 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, which may be 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 V433, E105, Q102, E432, T113, I272, I206, V256, E99,

E191, E401, N239, T106, and A277 in SEQ ID NO: 1.

In some instances, the variant RNA-guided nuclease polypeptide contains up to 20 arginine and/or lysine substitutions relative to the reference RNA-guided nuclease. In specific examples, the RNA-guided nuclease polypeptide contains up to 15 arginine and/or lysine substitutions relative to the reference RNA-guided nuclease.

In some examples, the variant RNA-guided nuclease polypeptide comprises arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at the following positions relative to SEQ ID NO: 1:

a) Q102, I206, and V433;

b) Q102, V256, and V433;

c) Q102, E191, I272, and V433;

d) Q102, V256, and V433;

e) Q102, E191, I206, and V433;

f) Q102, I272, and V433;

g) E105, V256, and V433;

h) E105, E191, V256, and V433;

i) E105, E191, I272, E401, and 433; or

j) E105, E191, N239, I272, and V433.

In one specific example, the variant RNA-guided nuclease polypeptide comprises arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at Q102, I206, and V433.

In some examples, the variant RNA-guided nuclease polypeptide contains the arginine substitutions of (a) Q102R, I206R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (b) Q102R, V256R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (c) Q102R, E191R, I272R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (d) Q102R, V256R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (e) Q102R, E191R, I206R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (f) Q102R, I272R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (g) E105R, V256R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (h) E105R, E191R, V256R, and V433R. In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (i) E105R, E191R, I272R, E401R, and 433R.

In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (j) E105R, E191R, N239R, I272R, and V433R. In some specific examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of Q102R, I206R, and V433R.

In some examples, the variant RNA-guided nuclease polypeptide disclosed herein may comprise a combination of arginine/lysine substitutions and nickase mutations.

In some instances, the variant RNA-guided nuclease polypeptide disclosed herein may comprise: (a) the one or more mutations in the HNH nuclease domain (*e.g.*, at position H232, H231, and/or H255 relative to SEQ ID NO: 1); and (b) the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at positions Q102R, I206R, and V433R relative to SEQ ID NO: 1.

In some instances, the variant RNA-guided nuclease polypeptide disclosed herein may comprise: (a) the one or more mutations in the HNH nuclease domain (*e.g.*, at position H232, H231, and/or H255 relative to SEQ ID NO: 1); and (b) the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at positions Q102R, I206R, and V433R relative to SEQ ID NO: 1. By way of non-limiting examples, a variant RNA-guided nuclease polypeptide may comprise: (a) the nickase mutation of H232A relative to SEQ ID NO: 1; and (b) the arginine substitutions of Q102R, I206R, and V433R relative to SEQ ID NO: 1.

Any of the RNA-guided 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 RNA-guided nuclease polypeptide may comprise an amino acid sequence at least 98% identical to SEQ ID NO: 1. Exemplary RNA-guided nuclease polypeptides are listed in **Table 1** and **Table 11**, each of which is within the scope of the present disclosure. In some examples, the RNA-guided nuclease polypeptide may have the native N-terminus M residue removed, for example, when the RNA-guided nuclease polypeptide is fused to a functional fragment (*e.g.*, those disclosed herein) at its N-terminus or when the RNA-guided nuclease polypeptide is located at the C-terminal portion of the fusion polypeptide.

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: 56 or SEQ ID NO: 57 (with no N-terminus M residue). Other exemplary MMLV-RT polypeptides are provided in **Table 25** below, each of which can be used in the fusion polypeptides disclosed herein. In some examples, the RT polypeptide may have the native N-terminus M residue removed, for example, when the RT polypeptide is fused to a functional fragment (*e.g.*, those disclosed herein) at its N-terminus or when the RT polypeptide is

located at the C-terminal portion of the fusion polypeptide.

In some embodiments, the fusion polypeptide comprises one or more NLS upstream or downstream to the RNA-guided nuclease polypeptide, the RT polypeptide, or both. For example, the fusion polypeptide, from N-terminus to C-terminus, comprises a first NLS, the RNA-guided 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 RNA-guided nuclease polypeptide and the RT polypeptide.

In some examples, the fusion polypeptide, from N-terminus to C-terminus, may comprise: a first NLS, the RNA-guided nuclease polypeptide, a peptide linker, the RT polypeptide, and a second NLS. In other examples, the fusion polypeptide, from N-terminus to C-terminus, may comprise a first NLS, the RT polypeptide, a peptide linker, the RNA-guided nuclease polypeptide, and a second NLS. In specific examples, the fusion polypeptide in any of the gene editing systems provided herein may comprise the amino acid sequence set forth in **Table 12**, **Table 17, Table 20,** or **Table 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 RNA-guided 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). A PAM can also be referred to as a target adjacent motif (TAM). In some embodiments, the PAM is 5'-RRT-3' or 5'-NRT-3', in which R represents A or G and N represents any nucleotide.

In some embodiments, the spacer sequence in the gRNA can be 14-30-nucleotide in length. In one example, the spacer sequence may be 14-20-nucleotide in length, 15-20 nucleotide in length, or 16-22-nucleotide in length.

In some embodiments, the scaffold sequence may comprise a nucleotide sequence at least 70% (*e.g.*, at least 75%) 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. For example, the scaffold sequence may comprise: (a) one or more mutations within nucleotides 11-27; (b) one or more mutations within nucleotides 96-128; (c) a deletion within nucleotides 155-160 (*e.g.*, deletion at position 160); (d) a deletion within nucleotides 180-182, or (e) a combination of any of (a)-(d), relative to SEQ ID NO: 2. In some examples, the scaffold sequence comprises the mutations set forth in (a), which may comprise a

combination of deletions and nucleotide substitutions to shorten the P1 stem-loop depicted in **FIG. 3A**. Alternatively or in addition, the scaffold sequence may comprise or further comprise mutations set forth in (b), which may comprise a combination of deletions and nucleotide substitutions to shortening or eliminate the P4a and/or P4b stem-loop structures depicted in **FIG. 3A**. In some instances, the P4a and P4b stem-loop structures may be replaced with a small stem-loop structure (*e.g.*, CCAGAAAUGG; SEQ ID NO: 213).

In some specific examples, the scaffold sequence comprises the nucleotide sequence of any one of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, and SEQ ID NO: 146-148 and 150-152. See **Table 5B** below.

In some instances, the RNA-guided nuclease polypeptide comprised in the fusion polypeptide disclosed herein is Nuclease BT of SEQ ID NO: 1 (or a counterpart thereof lacking the N-terminus M residue; SEQ ID NO: 140) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2. In some instances, the RNA-guided nuclease polypeptide comprised in the fusion polypeptide disclosed herein can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2. In some instances, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 25. In some instances, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 26. In some instances, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, V433, and H232 (*e.g.*, Q102R, I206R, V433R, and H232A) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 25. In some instances, the RNA-guided nuclease polypeptide in fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, V433, and H232 (*e.g.*, Q102R, I206R, V433R, and H232A) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 26.

In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide disclosed herein can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 146. In another example, the

RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 147. In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 148. In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 149. In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 150. In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 151. In another example, the RNA-guided nuclease polypeptide in the fusion polypeptide can be a variant of SEQ ID NO: 1, comprising mutations at positions Q102, I206, and V433 (*e.g.*, arginine substitutions Q102R, I206R, and V433R) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 152.

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 nuclease-RT fusion polypeptide or the encoding nucleic acid and/or the RNA molecule 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 encoding nucleic acid and/or the RNA molecule 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 or the encoding nucleic acid. In some examples, the gene editing system comprises an AAV vector encoding both the fusion polypeptide and the RNA molecule of the gene editing system.

In some examples, the gene editing system disclosure herein may comprise a vector (e.g., a viral vector such as an AAV vector or an AdV vector) comprising the first nucleic acid

encoding the fusion polypeptide and the second nucleic acid encoding the RNA molecule. In other examples, the gene editing system disclosed herein may be formulated with one or more lipid excipients to form a composition. Such a gene editing system may comprise a mRNA encoding the fusion polypeptide and the RNA molecule as disclosed herein.

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 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 RNA-guided 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 set forth in **Table 12**, **Table 17, Table 20,** or **Table 22**.

In addition, the present disclosure provides a nucleic acid encoding the fusion polypeptide disclosed herein. Such a nucleic acid may comprise the nucleotide sequence set forth in **Table 12**, **Table 17, Table 20, or Table 22**. 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 RNA-guided nuclease of SEQ ID NO: 1.

FIGS. 2A-2D are gel images and quantification of nuclease activity. **FIG. 2A** is a gel image captured using a 700 nm channel showing *in vitro* cleavage of the target strand (labelled

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

FIGs. 3A-3J include diagrams illustrating predicted secondary structures of guide RNA scaffold sequences recognizable by the RNA-guided nucleases disclosed herein. **FIG. 3A**: Reference scaffold. **FIG. 3B**: Scaffold 1. **FIG. 3C**: Scaffold 1a. **FIG. 3D**: Scaffold 2. **FIG. 3E**: Scaffold 3. **FIG. 3F**: Scaffold 4. **FIG. 3G**: Scaffold 5. **FIG. 3H**: Scaffold 6. **FIG. 3I**: Scaffold 7. **FIG. 3J**: Scaffold 8.

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure provides a gene editing system involving both an RNA-guided 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 RNA-guided 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.*, an RNA-guided 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 an RNA-guided 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, an RNA-guided 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. In some instances, the RNA-guided nuclease may be located N-terminal to the RT polypeptide in such a fusion polypeptide. Alternatively, the RNA-guided nuclease may be located C-terminal to the RT polypeptide in the fusion polypeptide. 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 RNA-guided nuclease, the RT, or the fusion polypeptide comprising such. In one example, the gene editing system may comprise one AAV vector for producing both the RNA-guided nuclease-RT fusion polypeptide and the gRNA-RT donor RNA molecule. In other examples, the gene editing system may comprise one or more mRNA molecules coding for the RNA-guided nuclease, the RT, or the fusion polypeptide comprising such.

In some embodiments, the RNA-guided 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) <u>RNA-guided nuclease Polypeptides</u>

The RNA-guided nuclease polypeptide for use in the gene editing systems disclosed herein may be the reference RNA-guided Nuclease BT comprising the amino acid sequence of

SEQ ID NO: 1 (or SEQ ID NO: 140 with N-terminus M residue removed), or a variant thereof.

As used herein, the term "RNA-guided 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. An RNA-guided 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.

Nuclease BT of SEQ ID NO: 1 (see **Table 1** below) is an RNA-guided nuclease that comprises both a RuvC nuclease domain (located at residues 60-96, 148-181, and 272-341 of SEQ ID NO: 1) and a HNH domain (located at residues 182-271 of SEQ ID NO: 1). The RuvC nuclease domain and the HNH nuclease domain coordinate cleavage of the DNA strand adjacent to the 5'-RRT-3' PAM or 5'-NRT-3' motif, in which R represents A or G and N represents any nucleotide. Positions D67, E176, and D329 are deemed the active sites in the RuvC domain and positions H231, H232, and H255 are deemed the active sites in the HNH domain. In addition to the nuclease domains, the reference RNA-guided nuclease of SEQ ID NO: 1 also includes a PLMP domain (residues 1-59 of SEQ ID NO: 1), a BH domain (residues 97-125 of SEQ ID NO: 1), a REC domain (residues 126-147 of SEQ ID NO: 1), a PLL domain (residues 342-355 of SEQ ID NO: 1), a WED domain (residues 356-426 of SEQ ID NO: 1), and a PID domain (residues 427-484 of SEQ ID NO: 1).

Compared to the CRISPR Cas9 nuclease, Nuclease BT of SEQ ID NO: 1 disclosed herein is smaller. Its cognate scaffold comprises a distinctive structure compared to the Cas9 nuclease scaffold. The distinctive scaffold is expected to allow for decreased size of some domains (*e.g.*, the REC domain) and thus contribute to the smaller size of Nuclease BT. 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 Nuclease BT of SEQ ID NO: 1 to increase indel activity. Additionally, since Nuclease BT of SEQ ID NO: 1 comprises a RuvC domain and an HNH domain, nickase variants can be engineered via disruption of the nuclease activity of one of the two domains. The cutting pattern of and PAMs capable of being recognized by Nuclease BT of SEQ ID NO: 1 are different than those of Cas9, allowing for additional gene targets to be edited. Also, unlike Cas9, Nuclease BT of SEQ ID NO: 1 comprises a PLMP domain. The PLMP

domain is expected to bind a helix on the 3' end of the scaffold and provide an increased binding affinity of Nuclease BT to its cognate scaffold.

Variants of RNA-Guided Nuclease Polypeptide

The variant RNA-guided nuclease polypeptides provided herein are derived from Nuclease BT of SEQ ID NO: 1, e.g., *via* introducing one or more mutations to the reference RNA-guided nuclease to modulate (*e.g.*, enhance or reduce) one or more activities of the nuclease. As used herein, the term "variant RNA-guided nuclease polypeptide" refers to an RNA-guided nuclease polypeptide comprising an alteration, *e.g.*, a substitution, insertion, deletion and/or fusion, at one or more residue positions, compared to Nuclease BT (SEQ ID NO: 1).

The variant RNA-guided nuclease polypeptides provided herein are expected to exhibit one or more modulated activities (*e.g.*, enhanced or reduced) relative to Nuclease BT. 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.*, an RNA-guided nuclease) to an RNA guide and/or target nucleic acid. In some examples, the variant RNA-guided nuclease polypeptides disclosed herein have an enhanced binding to a cognate guide RNA (gRNA) as compared with the reference RNA-guided 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 RNA-guided nuclease. A cognate gRNA refers to a gRNA having a scaffold recognizable by the RNA-guided nuclease.

In some examples, the variant RNA-guided nuclease polypeptides disclosed herein have an enhanced enzymatic activity relative to the reference RNA-guided 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 RNA-guided nuclease. In other examples, the variant RNA-guided nuclease polypeptides disclosed herein have a decreased enzymatic activity relative to the reference RNA-guided nuclease, e.g., having an enzymatic activity at least 20%, 30%, 40%, 50%, 60%, or 70% lower than that of the reference RNA-guided 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 RNA-guided nuclease polypeptides disclosed herein have enhanced indel activity relative to the reference RNA-guided nuclease. As used herein, the term "indel activity" refers to the ability of an RNA-guided 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 RNA-guided nuclease polypeptide provided herein share a high sequence homology relative to the reference RNA-guided nuclease. For example, the variant RNA-guided 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 RNA-guided nuclease polypeptide may comprise an amino acid sequence at least 90% identical to SEQ ID NO: 1. In some instances, the variant RNA-guided nuclease polypeptide may comprise an amino acid sequence at least 95% identical to SEQ ID NO: 1. In other instances, the variant RNA-guided 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 RNA-guided nuclease polypeptide provided herein can contain one or more alterations relative to Nuclease BT 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 RNA-guided 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 RNA-guided 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 some instances, the variant RNA-guided nuclease polypeptide may contain up to 15 arginine and/or lysine substitutions, for example, up to 15 arginine substitutions. In specific examples, the variant RNA-guided 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 RNA-guided 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 E99, Q102, E105, T106, T113, E191, I206, N239, I272, V256, A277, E401, E432, and V433 in SEQ ID NO: 1. In some specific examples, the variant RNA-guided nuclease polypeptide may comprise arginine and/or lysine substitutions (*e.g.*, arginine substitutions) at the following positions relative to SEQ ID NO: 1: a) Q102, I206, and V433; b) Q102, V256, and V433; c) Q102, E191, I272, and V433; d) Q102, V256, and V433; e) Q102, E191, I206, and V433; f) Q102, I272, and V433; g) E105, V256, and V433; h) E105, E191, V256, and V433; i) E105, E191, I272, E401, and 433; and j) E105, E191, N239, I272, and V433. In one specific example, the variant RNA-guided nuclease

polypeptide comprises arginine and/or lysine substitutions at Q102, I206, and V433.

In some examples, the variant RNA-guided nuclease polypeptide may contain one or more of the following arginine substitutions relative to SEQ ID NO: 1: V433R, E105R, Q102R, E432R, T113R, I272R, I206R, V256R, E99R, E191R, E401R, N239R, T106R, and A277R. By way of non-limiting examples, the one or more arginine substitutions may include a) Q102R, I206R, and V433R; b) Q102R, V256R, and V433R; c) Q102R, E191R, I272R, and V433R; d) Q102R, V256R, and V433R; e) Q102R, E191R, I206R, and V433R; f) Q102R, I272R, and V433R; g) E105R, V256R, and V433R; h) E105R, E191R, V256R, and V433R; i) E105R, E191R, I272R, E401R, and 433R; or j) E105R, E191R, N239R, I272R, and V433R. In some specific examples, the variant RNA-guided nuclease polypeptide comprises the arginine substitutions of Q102R, I206R, and V433R.

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 D67, E176, and/or D329 in the RuvC domain). In other examples, the arginine and/or lysine substitutions may be within the HNH domain, e.g., at positions H231, H232, and/or H255 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 D67, E176, and/or D329 in the RuvC domain and/or at positions H231, H232, and/or H255 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, including that these positions are not tolerable to mutations with respect to nuclease activity: T217R, T343R, L104R, E346R, Y216R, P227R, G144R, D323R, N83R, D342R, L338R, I128R, P133R, Y378R, V402R, N321R, Q238R, F307R, G394R, A222R, N82R, D189R, L154R, V168R, P169R, D397R, V339R, A165R, S142R, W295R, E319R, S349R, G301R, F93R, T80R, G444R, Q354R, L53R, N393R, S274R, V160R, D170R, W63R, P475R, S54R, L220R, E297R, H94R, Q56R, F435R, G462R, N34R, V210R, L230R, I74R, P421R, Q55R, V379R, V226R, V305R, H347R, W306R, D471R, H131R, F40R, Q269R, E372R, L247R, V171R, L345R, V447R, T61R, Y314R, I276R, D58R, D352R, E359R, S454R, V473R, D242R, D448R, G126R, I443R, H3R, F50R, Q48R, I309R, I79R,

A134R, T463R, V355R, P60R, F456R, G12R, S90R, P129R, L264R, D24R, Q32R, V484R, V315R, N187R, C481R, I356R, Q196R, V234R, I44R, W479R, G192R, H364R, E155R, M186R, L88R, Q360R, I41R, N15R, F117R, V37R, I38R, Q4R, I16R, M91R, H152R, H18R, Y426R, H243R, W29R, P396R, V482R, M45R, I64R, Q194R, L89R, T453R, L376R, V122R, F399R, I164R, A208R, L320R, V85R, F430R, I173R, P188R, V10R, D461R, D469R, N14R, E348R, V49R, N336R, L23R, L467R, L268R, L200R, F228R, L296R, I161R, V472R, D70R, L6R, P148R, M25R, G202R, L17R, A311R, D11R, F358R, V46R, T325R, L146R, C327R, F201R, A35R, L30R, Y474R, F136R, S149R, V150R, F8R, W145R, H212R, Y177R, P19R, Y203R, S156R, I100R, N138R, L76R, G42R, T71R, M470R, S281R, L265R, Y331R, S174R, T282R, N391R, L322R, C221R, D245R, Q229R, C75R, Y62R, F465R, T459R, V7R, G425R, Y110R, E437R, T440R, L292R, I72R, P235R, Y195R, N328R, V47R, L417R, T95R, Y476R, Y460R, V26R, S370R, D278R, N205R, T147R, I316R, Y458R, G240R, M185R, Q363R, G310R, D312R, Q183R, I441R, L9R, L371R, V66R, C308R, I287R, G478R, L250R, A439R, F279R, I288R, E92R, A334R, D67R, T249R, G73R, Y209R, G477R, M289R, V449R, I407R, I335R, N246R, P68R, E176R, T20R, C218R, D329R, G452R, L175R, Y377R, V258R, V77R, C251R, H232R, G457R, G198R, I333R, A178R, H255R, A389R, H157R, N286R, S283R, L285R, S313R, H259R, H326R, G422R, G69R, F153R, S97R, G65R, L450R, G438R, C332R, H231R, I190R, D181R, C254R, A330R, D214R, I182R, and F180R. In some embodiments, a variant CRISPR nuclease provided herein exhibits nuclease activity and may not have mutations at the above positions.

Alternatively or in addition, the variant RNA-guided 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, the variant RNA-guided nuclease disclosed herein is a nickase comprising one or more nickase mutations in the HNH domain to deactivate the HNH domain. In other embodiments, a variant RNA-guided nuclease can be a nickase comprising one or more nickase mutations in the RuvC domain to deactivate the RuvC domain. The variant RNA-guided nuclease polypeptides may share a high sequence homology relative to the reference RNA-guided 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 D67, E176, and D329 are identified as putative catalytic residues in the RuvC domain, and positions H231, H232, and H255 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 D67, E176, and/or D329 in the RuvC domain). In some specific examples, the variant RNA-guided nuclease polypeptide may contain substitution(s) of D67A, E176A, and/or D329A. In other examples, any of D67, E176, and D329 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 H231, H232, and/or H255 in the HNH domain). In some specific examples, the variant RNA-guided nuclease polypeptide may contain substitutions of H231A, H232A, and/or H255A. Alternatively, any of H231, H232, and H255 may be replaced with an amino acid residue similar to A, for example, G, L, or S. In one specific example, the variant RNA-guided nuclease polypeptide may contain substitution of H232A or H232L.

In some embodiments, the variant RNA-guided 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 RNA-guided nuclease polypeptide may comprise arginine/lysine substitutions (*e.g.*, arginine substitutions) at positions Q102, I206, and/or V433, and further comprise an amino acid substitution at H232 (*e.g.*, H1232A or H232L) in SEQ ID NO: 1.

In some examples, the variant RNA-guided nuclease polypeptide that contains one or more of the mutations noted herein (*e.g.*, arginine and/or lysine substitutions, and/or nickase mutation) may share a sequence identity at least 90% (*e.g.*, 95%, 97%, 98%, 99%, 99.5%, or greater) with SEQ ID NO: 1. In some specific examples, a variant RNA-guided nuclease polypeptide may comprise (*e.g.*, consist of) an amino acid sequence of any one of SEQ ID NOs: 31, 33, 35, 37, 39, 41, and 43. In one example, a variant RNA-guided nuclease polypeptide may

comprise (*e.g.*, consist of) SEQ ID NO: 33. In another example, a variant RNA-guided nuclease polypeptide may comprise (*e.g.*, consist of) SEQ ID NO: 35.

Exemplary RNA-guided nuclease polypeptides for use in the gene editing systems provided herein are disclosed in **Table 1** or **Table 11** below, each of which is within the scope of the present disclosure.

In some embodiments, the RNA-guided nuclease polypeptide in the gene editing systems disclosed herein (*e.g.*, Nuclease BT of SEQ ID NO: 1 or the counterpart with no N-terminus M residue, or any of the variants thereof as disclosed herein) may be a fusion polypeptide comprising an RNA-guided 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 RNA-guided nuclease portion in the fusion polypeptide may be the reference RNA-guided nuclease of SEQ ID NO: 1. Alternatively, the RNA-guided nuclease portion in the fusion polypeptide may be a variant RNA-guided 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 RNA-guided nuclease). In some embodiments, the fusion polypeptide may comprise an NLS within a flexible loop of the RNA-guided nuclease.

In some embodiments, the gene editing system provided herein may comprise the RNA-guided 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 RNA-guided nuclease polypeptide. In some examples, the nucleotide sequence encoding the RNA-guided 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 RNA-guided nuclease polypeptides as disclosed herein can be an mRNA molecule, which can be codon optimized. Exemplary codon-optimized nucleotide sequences encoding exemplary RNA-guided 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 adeno-associated viral (AAV) vector, an adenoviral (AdV) vector, or a retroviral vector) encoding the RNA-guided 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 RNA-guided 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 RNA-guided 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 one of those listed in **Table 25**, a variant thereof, or an ortholog thereof, any of which can be used in the gene editing system disclosed herein.

Table 25. Exemplary Reverse Transcriptase Sequences

Reverse Transcriptase	Sequence	SEQ ID NO
MMLV-RT	See Table 14 below	56 or 57
MMLV-RT_2	MTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIK PHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPT VPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISG QLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEL DCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKET VMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQ QKAYQEILQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVA YLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQ PPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDIL AEAHGTRPDL	58 With no N-terminus M residue (SEQ ID NO: 59)
MMTV-RT	MFTLWGRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLN QWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLLQDLR AVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFS VPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAILTVRDKYQDSYIVHYM	60 With no N-

	DDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDS VSYQKLQIRTDKLRTLNDFQKLLGNINWIRPFLKLTTGELKPLFEILNGDSNP ISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQD GVVEWIHLPHISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQ FDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFN LYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQEKFYIGHIRGHT GLPGPLAQGNAYADSLTRIL	terminus M residue (SEQ ID NO: 61)
MarathonRT	MDTSNLMEQILSSDNLNRAYLQVVRNKGAEGVDGMKYTELKEHLAKNGETIKG QLRTRKYKPQPARRVEIPKPDGGVRNLGVPTVTDRFIQQAIAQVLTPIYEEQF HDHSYGFRPNRCAQQAILTALNIMNDGNDWIVDIDLEKFFDTVNHDKLMTLIG RTIKDGDVISIVRKYLVSGIMIDDEYEDSIVGTPQGGNLSPLLANIMLNELDK EMEKRGLNFVRYADDCIIMVGSEMSANRVMRNISRFIEEKLGLKVNMTKSKVD RPSGLKYLGFGFYFDPRAHQFKAKPHAKSVAKFKKRMKELTCRSWGVSNSYKV EKLNQLIRGWINYFKIGSMKTLCKELDSRIRYRLRMCIWKQWKTPQNQEKNLV KLGIDRNTARRVAYTGKRIAYVCNKGAVNVAISNKRLASFGLISMLDYYIEKC VTC	62 With no N-terminus M residue (SEQ ID NO: 63)
RTX reverse transcriptase	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVKK ITAERHGTVVTVKRVEKVQKKFLGRPVEVWKLYFTHPQDVPAIMDKIREHPAV IDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIETLYHEGEEFAEGPILM ISYADEEGARVITWKNVDLPYVDVVSTEREMIKRFLRVVKEKDPDVLITYNGD NFDFAYLKKRCEKLGINFALGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVIR RTINLPTYTLEAVYEAVFGQPKEKVYAEEITTAWETGENLERVARYSMEDAKV TYELGKEFLPMEAQLSRLIGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNK PDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLN REGCKEYDVAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERK LLDYRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYLTMTIKEIEE KYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALELEYEGFY KRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDV EKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKDYKATGPHVAVAKRLAAR GVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDPTKHKYDAEYYIEKQVLPAVE RILRAFGYRKEDLRYQKTRQVGLSARLKPKGT	64 With no N-terminus M residue (SEQ ID NO: 65)
RTX-exoMinus	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVKK ITAERHGTVVTVKRVEKVQKKFLGRPVEVWKLYFTHPQDVPAIMDKIREHPAV IDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLAFDIETLYHEGEEFAEGPILM ISYADEEGARVITWKNVDLPYVDVVSTEREMIKRFLRVVKEKDPDVLITYDGD NFDFAYLKKRCEKLGINFALGRDGSEPKIQRMGDRFAVEVKGRIHFDLYPVIR RTINLPTYTLEAVYEAVFGQPKEKVYAEEITTAWETGENLERVARYSMEDAKV TYELGKEFLPMEAQLSRLIGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNK PDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLN REGCKEYDVAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERK LLDYRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYLTMTIKEIEE KYGFKVIYSDTDGFFATIPGADAETVKKKAMEFLKYINAKLPGALELEYEGFY KRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDV EKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKDYKATGPHVAVAKRLAAR GVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDPTKHKYDAEYYIEKQVLPAVE RILRAFGYRKEDLRYQKTRQVGLSARLKPKGT	66 With no N-terminus M residue (SEQ ID NO: 67)
B11 RT	MILDTDYITEDGKPVIRIFKKENGEFKIEYDRTFEPYLYALLKDDSAIEEVKK ITAERHSTVVTVKRVEKVQKKFLGRSVEVWKLYFTHPQDVPAIMDKIREHPAV IDIYEYDIPFAIRYLIDKGLVPMEGDEELKLLALDIGTPCHEGEVFAEGPILM ISYADEEGTRVITWRNVDLPYVDVLSTEREMIQRFLRVVKEKDPDVLITYNGD NFDFAYLKKRCEKLGINFTLGREGSEPKIQRMGDRFAVEVKGRIHFDLYPVIR RTVNLPIYTLEAVYEAVFGQPKEKVYAEEITTAWETGENLERVARYSMEDAKV TYELGKEFMPMEAQLSRLIGQSLWDVSRSSTGNLVEWFLLRKAYERNELAPNK PDEKELARRHQSHEGGYIKEPERGLWENIVYLDFRSLYPSIIITHNVSPDTLN REGCKEYDVAPQVGHRFCKDFPGFIPSLLGDLLEERQKIKKRMKATIDPIERK LLDYRQRAIKILANSLYGYYGYARARWYCKECAESVIAWGREYITMTIKEIEE KYGFKLIYSDTDGFFATIPGAEAETVKKKAMEFLKYINAKLPGALELEYEGFY KRGLFVTKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDV EKAVRIVKEVTEKLSKYEVPPEKLVIHKQITRDLKDYKATGPHVAVAKRLAAR	68 With no N-terminus M residue (SEQ ID NO: 69)

	GVKIRPGTVISYIVLKGSGRIVDRAIPFDEFDPTKHKYDAEYYIENQVLPAVE RILRAYGYRKEDLWYQKTRQVGLSARLKPKGT	
RT CaMV	MDHLLLKTQTQTEQVMNVTNPNSIYIKGRLYFKGYKKIELHCFVDTGASLCIA SKFVIPEEHWVNAERPIMVKIADGSSITISKVCKDIDLIIAGEIFRIPTVYQQ ESGIDFIIGNNFCQLYEPFIQFTDRVIFTKNKSYPVHIAKLTRAVRVGTEGFL ESMKKRSKTQQPEPVNISTNKIENPLEEIAILSEGRRLSEEKLFITQQRMQKI EELLEKVCSENPLDPNKTKQWMKASIKLSDPSKAIKVKPMKYSPMDREEFDKQ IKELLDLKVIKPSKSPHMAPAFLVNNEAEKRRGKKRMVVNYKAMNKATVGDAY NLPNKDELLTLIRGKKIFSSFDCKSGFWQVLLDQESRPLTAFTCPQGHYEWNV VPFGLKQAPSIFQRHMDEAFRVFRKFCCVYVDDILVFSNNEEDHLLHVAMILQ KCNQHGIILSKKKAQLFKKKINFLGLEIDEGTHKPQGHILEHINKFPDTLEDK KQLQRFLGILTYASDYIPKLAQIRKPLQAKLKENVPWRWTKEDTLYMQKVKKN LQGFPPLHHPLPEEKLIIETDASDDYWGGMLKAIKINEGTNTELICRYASGSF KAAEKNYHSNDKETLAVINTIKKFSIYLTPVHFLIRTDNTHFKSFVNLNYKGD SKLGRNIRWQAWLSHYSFDVEHIKGTDNHFADFLSREFNKVNSSGGS	70 With no N-terminus M residue (SEQ ID NO: 71)
RT retron	MGIHGVPAAMKSAEYLNTFRLRNLGLPVMNNLHDMSKATRISVETLRLLIYTA DFRYRIYTVEKKGPEKRMRTIYQPSRELKALQGWVLRNILDKLSSSPFSIGFE KHQSILNNATPHIGANFILNIDLEDFFPSLTANKVFGVFHSLGYNRLISSVLT KICCYKNLLPQGAPSSPKLANLICSKLDYRIQGYAGSRGLIYTRYADDLTLSA QSMKKVVKARDFLFSIIPSEGLVINSKKTCISGPRSQRKVTGLVISQEKVGIG REKYKEIRAKIHHIFCGKSSEIEHVRGWLSFILSVDSKSHRRLITYISKLEKK YGKNPLNKAKT	72 With no N-terminus M residue (SEQ ID NO: 73)
HIV2	MEKEGQLEEAPPTNPYNTPTFAIKKKDKNKWRMLIDFRELNKVTQDFTEIQLG IPHPAGLAKKRRITVLDVGDAYFSIPLHEDFRPYTAFTLPSVNNAEPGKRYIY KVLPQGWKGSPAIFQHTMRQVLEPFRKANKDVIIIQYMDDILIASDRTDLEHD RVVLQLKELLNGLGFSTPDEKFQKDPPIHWMGYELWPTKWKLQKIQLPQKEIW TVNDIQKLVGVLNWAAQLYPGIK	74 With no N-terminus M residue (SEQ ID NO: 75)
HIV1	MEEEGKISKIGPENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLG IPHPAGLKKKKSVTVLDVGDAYFSVPLDESFRKYTAFTIPSINNETPGIRYQY NVLPQGWKGSPAIFQSSMTKILEPFRIKNPEIVIYQYMDDLYVGSDLEIGQHR TKIEELRAHLLSWGFTTPDKKHQKEPPFLWMGYELHPDRWTVQPIDLPEKDSW TVNDIQKLVGKLNWASQIYAGIK	76 With no N-terminus M residue (SEQ ID NO: 77)
HIV-1 RT p51	MPISPIETVPVKLKPGMDGPKVKQWPLTEEKIKALVEICTEMEKEGKISKIGP ENPYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAGLKKKKS VTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPA IFQSSMTKILEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLR WGLTTPDKKHQKEPPFLWMGYELHPDKWTVQPIVLPEKDSWTVNDIQKLVGKL NWASQIYPGIKVRQLCKLLRGTKALTEVIPLTEEAELELAENREILKEPVHGV YYDPSKDLIAEIQKQGQGQWTYQIYQEPFKNLKTGKYARMRGAHTNDVKQLTE AVQKITTESIVIWGKTPKFKLPIQKETWETWWTEYWQATWIPEWEFVNTPPLV KLWYQLEKEPIVGAETF	78 With no N-terminus M residue (SEQ ID NO: 79)
MMLV RT lacking the RNase H domain	STLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPL KATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPG TNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCL RLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFR IQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQK QVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGF AEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFV DEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDA GKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVV ALNPATLLP	80 With N-terminus M residue (SEQ ID NO: 81)

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 ºC; however, mutations may be introduced that allow for the reverse transcription activity at higher temperatures of over 48ºC, including 49ºC, 50ºC, 51ºC, 52ºC, 53ºC, 54ºC, 55ºC, 56ºC, 57ºC, 58ºC, 59ºC, 60ºC, 61ºC, 62ºC, 63ºC, 64ºC, 65ºC, 66ºC, 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 RNA-guided 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 11** 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 RNA-guided 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 RNA-guided nuclease polypeptide downstream to the RT polypeptide. In other embodiments, the fusion polypeptide may comprise the RNA-guided nuclease polypeptide at its N-terminus and the RT polypeptide downstream to the RNA-guided nuclease polypeptide. In some embodiments, the RT polypeptide may be fused with the RNA-guided nuclease polypeptide at an intramolecular position within the RT polypeptide, for example, the RNA-guided nuclease polypeptide may be within a loop of the reverse transcriptase polypeptide.

Any of the RNA-guided nuclease polypeptide disclosed herein and any of the RT polypeptides disclosed herein may be used for constructing the fusion polypeptides. In some instances, the RNA-guided nuclease polypeptide may be Nuclease BT of SEQ ID NO: 1 (or SEQ ID NO: 140). Alternatively, the RNA-guided nuclease polypeptide may be a variant 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 RNA-guided nuclease, for example, the nickase of SEQ ID NO: 33 (or SEQ ID NO: 161). 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. Exemplary variants of SEQ ID NO: 1 for use in constructing the fusion polypeptides disclosed here are provided in **Table 1** and **Table 11**. In some instances, the native N-terminus M residue of Nuclease BT or the variant thereof may be removed in the fusion polypeptide, for example, when a functional fragment such as NLS is fused to the N-terminus of the nuclease or when the nuclease is located at the C-terminal portion of the fusion polypeptide. Similarly, the native N-terminus M residue of an RT enzyme may be removed in the fusion polypeptide, for example, when a functional fragment such as NLS is fused to the N-terminus of the RT enzyme or when the RT enzyme is located at the C-terminal portion of the fusion polypeptide.

In some embodiments, any of the RNA-guided 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 RNA-guided 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 **Table 1** below. See also **Examples 1** and **2**.

In some examples, the RNA-guided nuclease-RT fusion polypeptide provided herein comprises a peptide linker located between the RNA-guided nuclease polypeptide and the RT polypeptide. In some instances, the RNA-guided 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 RNA-guided nuclease polypeptide and the RT polypeptide. For example, an additional peptide linker and an NLS may be place between the RNA-guided nuclease polypeptide and the RT polypeptide, in addition to the first peptide linker. In some specific examples, the peptide linker between the RNA-guided 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 RNA-guided 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 RNA-guided nuclease-RT fusion polypeptide disclosed herein may comprise, from N-terminus to C-terminus, a first NLS, the RNA-guided nuclease, the RT

polypeptide, and a second NLS. In some examples, the RNA-guided nuclease-RT fusion polypeptide disclosed herein may comprise, from N-terminus to C-terminus, a first NLS, the RNA-guided nuclease, a peptide linker, the RT polypeptide, and a second NLS. Examples of the RNA-guided nuclease-RT polypeptides are provided in **Tables 12**, **17**, **Table 20**, and **Table 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 RNA-guided 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 one specific example, the gene editing system disclosed herein may comprise an mRNA molecule encoding the fusion polypeptide and the RNA molecule as disclosed herein.

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 RNA-guided nuclease-RT fusion polypeptide.

(iv). Preparation of Protein Components

The RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide disclosed herein may be prepared by conventional methods or the methods disclosed herein. For example, the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide are not particularly limited as long as they can produce the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide. In some

embodiments, a vector disclosed herein includes a nucleotide sequence encoding RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide described herein.

In some embodiments, a host cell described herein is used to express the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide. After expression, the host cells can be collected and the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, a mature RT polypeptide, or a mature RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide to a host cell in a subject (*e.g.*, a human subject) wherein the polyribonucleotide encodes the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided nuclease-RT fusion polypeptide and expressing the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 nuclei 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 RNA-guided 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 RNA-guided nuclease.

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 (target sequence) and a scaffold, which is recognizable by the variant RNA-guided nuclease polypeptide contained in the gene editing system. The target sequence can be adjacent to a protospacer adjacent motif (PAM) of 5'-RRT-3' or 5'-NRT-3', in which R represents A or G and 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 RNA-guided 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 (upstream to or 5' to the PAM motif). In other specific examples, the target sequence and the PAM 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 about 14 nucleotides to about 30 nucleotides. For example, the spacer can have a length of about 14 nucleotides to about 20 nucleotides, from about 14 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 14 and 25 nucleotides (*e.g.*, 14, 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 16-22 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 RNA-guided 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 Nuclease BT of SEQ ID NO: 1.

GGGACAUGACCAGAUGCGAAAGUAACUGUCAUGUCCUCAUGCCAAUGGAC
UAUGGUGAUGCACCUAGGGGUGUCGUUCCAGCUCCUAGCUCUGCGAAGGA
GCAUUCGUUAUCGGAUGCGUUUGUACUUCAAGCCAUAGCCCAGGCCAGGG
ACGUAUUACUGCUCCUGCCCUAUAGGAGCAUA (SEQ ID NO: 2)

The projected secondary structure of the cognate scaffold (the reference scaffold sequence of SEQ ID NO: 2) is depicted in **FIG. 3A**. This scaffold sequence includes multiple stem-loop structures, including step-loop P1, step-loop P2 (including P2A and P2b), stem-loop P3 (including P3a and P3b), step-loop P4 (including P4a and P4b), step P5, and step-loop P6. The P5 stem is formed via base-pairing between the loop sequence in P3b and the segment connecting step-loop P2a and step-loop P6.

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 70% (e.g., at least 75%, 80%, 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 RNA-guided 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 variant of SEQ ID NO: 2 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, a variant of SEQ ID NO: 2 may contain a truncated P1 stem-loop relative to that in SEQ ID NO: 2. Such a variant may comprise deletions and/or nucleotide substitutions within the region of the P1 stem-loop (*e.g.*, within the region of 11-27 in SEQ ID NO: 2) so as to result in a shortened stem-loop P1. Alternatively or in addition, a variant of SEQ ID NO: 2 may comprise or further comprise truncated P4 stem-loop structures, for example, truncated P4a, truncated P4b, or a combination thereof. Such a variant may include deletions and optionally nucleotide substitutions within the P4 stem-loops, for example, within the region 96-128 of SEQ ID NO: 2. In some examples, the whole P4a and P4b stem-loops can be deleted and replaced with a smaller step-loop (e.g., a stem-loop with 15 or less nucleotides, *e.g.*, 12 or less nucleotides). In one example, the replacement step-loop may have the sequence of CCAGAAAUGG (SEQ ID NO: 213). Alternatively or in addition, a variant of SEQ ID NO: 2 may have a deletion within the segment connecting P2a and P6, for example, within the region of 155-160 of SEQ ID NO: 2. In one example, the nucleotide at position 160 of SEQ ID NO: 2 is deleted. Further, the 3' end nucleotide(s), for example, positions 180-182 may be deleted in a variant of SEQ ID NO: 2.

In one specific examples, the scaffold sequence comprises (*e.g.*, consists of) the nucleotide sequence of SEQ ID NO: 25. In another specific example, the scaffold sequence comprises (*e.g.*, consists of) SEQ ID NO: 26. Other examples are provided in **Table 5B** below. The projected secondary structures of these variant scaffold sequences are provided in **FIGs. 3B-3J**.

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 RNA-guided 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 RNA-guided 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 an RNA-guided 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 RNA-guided 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 RNA-guided nuclease polypeptide and synthesis of a DNA fragment from a free 3'end of a free DNA strand generated by the RNA-guided 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 RNA-guided 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 RNA-guided 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 RNA-guided 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 RNA-guided 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 H232 mutation, *e.g.*, H232A). Alternatively or in addition, the RNA-guided 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.*, V433, E105, Q102, E432, T113, I272, I206, V256, E99, E191, E401, N239, T106, and A277 in SEQ ID NO: 1). In some specific examples, the RNA-guided nuclease polypeptide in the fusion polypeptide may comprise a combination of arginine and/or lysine substitutions at positions provided herein, *e.g.*, Q102, I206, and V433 of SEQ ID NO: 1 (*e.g.*, Q102R, I206R and V433R).

In some embodiments, the RT polypeptide in the fusion polypeptide may be an MMLV variant, for example, SEQ ID NO: 56 or SEQ ID NO: 57 provided in **Example 6** below.

In some examples, the fusion polypeptide provided herein may comprise an N-terminal RNA-guided 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 RNA-guided 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 RNA-guided nuclease-RT fusion polypeptides provided in **Tables 12**, **17**, **20**, 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 RNA-guided nuclease-RT fusion polypeptides provided in **Tables 12**, **17**, **20**, 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 RNA-guided nuclease-RT fusion polypeptides provided in **Tables 12** and **17** 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 RNA-guided nuclease-RT fusion polypeptides provided in **Tables 12**, **17**, **20**, 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 RNA-guided 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/RNA-guided 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 RNA-guided 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 an RNA-guided nuclease polypeptide are delivered in separate compositions. In some embodiments, an RNA guide/RT donor RNA and an RNA encoding a RNA-guided 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 RNA-guided nuclease polypeptide/RT polypeptide delivered in separate compositions are delivered using different delivery technologies. In one example, an AAV vector may be used to deliver both protein components (the RNA-guided nuclease polypeptide, the RT polypeptide, or the fusion polypeptide comprising both) and the RNA components (the RNA guide, the RT donor RNA, or the single RNA molecule comprising both).

In some embodiments, one or more of the protein components and one or more of the RNA components are delivered together. For example, the RNA-guided 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 RNA-guided nuclease and/or RT polypeptides and the RNA guide and/or RT donor RNA are delivered separately. For example, the RNA-guided nuclease and/or RT polypeptides and the RNA guide and/or RT donor RNA are packaged into separate AAV particles. In another example, the RNA-guided 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 an RNA-guided 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.

In some examples, the gene editing system provided herein may comprise a nucleic acid

encoding the fusion polypeptide and a nucleic acid encoding the RNA molecule, for example, a single vector (e.g., a viral vector such as an AAV or AdV vector) comprising both nucleic acids for delivering both gene editing components to host cells, wherein the target gene can be edited by the gene editing components.

In some examples, the gene editing system disclosed herein may comprise a mRNA molecule encoding the fusion polypeptide and the RNA molecule. Such a gene editing system can be delivered to host cells via lipid nanoparticles (LNPs) in some instances.

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 RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided nuclease polypeptide, and an RT polypeptide, or a fusion polypeptide thereof. In

some embodiments, the kits include the single RNA molecule and the RNA-guided nuclease-RT fusion polypeptide. In some embodiments, the kits include a polynucleotide that encodes the RNA-guided nuclease polypeptide, the RT polypeptide, or the RNA-guided 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 RNA-guided 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 RNA-guided 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 RNA-guided 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.

EXAMPLES

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

Example 1: Editing of Human Target Genes by an RNA-Guided Nuclease in HEK293T Cells

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

The RNA-guided nuclease was tagged with an N-terminal SV40 nuclear localization sequence (NLS) and a C-terminal XTEN linker directly upstream of a nucleoplasmin NLS, and its coding sequence was converted to a human codon-optimized DNA sequence, synthesized, and cloned into a pcDNA3.1 vector (Invitrogen), containing a CMV promoter for expression. The general architecture of the nuclease with NLS sequences was: SV40 NLS – Nuclease – Linker – nucleoplasmin NLS. The reference (Nuclease BT), NLS and linker sequences used are shown in **Table 1**.

Table 1. Sequences of RNA-Guided Nuclease Constructs

Component	Nucleotide Sequence (Codon Optimized for Human Cells)	Amino Acid sequence
Nuclease BT	atgagacaccagagactcgtgtttctcgtggacggta gaaacaacatcctgcatcctaccagaaagctggatat ggtcagaagatggctgagacaaacaaacgccaaggt gatcagattcatcggaaaaatcatggtggtccaggtgt tcaagatcctgtctcagcagaaagacgaccctacata ctggatcggcgtggacccaggcgacaccatcggcat ctgcctggtcagaatcacagccaacaacagagttaag cggctgctgagcatggaattccacaccagatctaagg aaatcaagcagctgctggaaacccggagaatgtacc ggtgcacacggagatactttagaagaaagcgggtga agaagaaagggagaattcccaagcacagacctgcta gatttgagaacagaaggcggagcgacggctggctg acccctagcgtgcgccacttccttgagagccacaagt ccgtgatccggaagatcgccgagtgtgtgccagatgt gaacatcagcctggagtacgccaagttcgacatccaa aagatgatgaaccccgatatcgagggcgagcagtac cagagaggcaaactgttcggatatcagaatatccgcg cctacgtgctgcatagagataagtacacctgtcagctg tgcgccaagaaaaacgtgcccttccagctgcaccac gtggtgcctagaagccagaacggcaaggaccaccc cgataacctgatcaccctgtgtcctgagtgccacgtga aagtgcacacagacaccaagctgctgaaaaaactgc agcggaagattaagagcaagattgctgatttcaccag caccagcagactgaatattatcatgagcttcctgaagc actggctggaaaacgagttcggcaaggatcgggtgt ggttctgcatcggagccgactcttatgtgatccggaag gaactgaacctggacaagacacactgcaacgacgct tactgcatcgccatcaacaagctggtgagatttgacac cgacctggaacacgaaagcgtggtggataagcaagt gatccggttcgagcagaaaagacagcacaaccgga gaatgaaatctctggagagaaacaggctgtactacgt ggtcaaggacaagaagcgcgtgcgggtggccacca acagaaatggcaatcctgatgccttccacgaggtgta cgacacaaagatcctgcctctgctgaagcagggctat cggctgaaagtgaaccctggcgttaagggctacagg aagaagttcagcgaggtgggatttaaggaaggcgcc acaatcagaatcggaaatcgggtggacgtgctgaag ggcacctccggcttcggctacacctacgacggcacc atcttcaagctcaaggacatggacgtggtgtacccta cggcggctggaagtgcgtgaaggtg (SEQ ID NO: 3)	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYW IGVDPGDTIGICLVRITANNRV KRLLSMEFHTRSKEIKQLLETR RMYRCTRRYFRRKRVKKKGR IPKHRPARFENRRRSDGWLTP SVRHFLESHKSVIRKIAECVPD VNISLEYAKFDIQKMMNPDIE GEQYQRGKLFGYQNIRAYVL HRDKYTCQLCAKKNVPFQLH HVVPRSQNGKDHPDNLITLCP ECHVKVHTDTKLLKKLQRKIK SKIADFTSTSRLNIIMSFLKHW LENEFGKDRVWFCIGADSYVI RKELNLDKTHCNDAYCIAINK LVRFDTDLEHESVVDKQVIRF EQKRQHNRRMKSLERNRLYY VVKDKKRVRVATNRNGNPDA FHEVYDTKILPLLKQGYRLKV NPGVKGYRKKFSEVGFKEGA TIRIGNRVDVLKGTSGFGYTY DGTIFKLKDMDVVYPYGGWK CVKV (SEQ ID NO: 1) *With no N-terminus M residue (SEQ ID NO: 140)*
SV40 NLS	aaacggacagccgacggaagcgagttcgagtcacc aaagaagaagcggaaagtc (SEQ ID NO: 4)	KRTADGSEFESPKKKRKV (SEQ ID NO: 5)

Component	Nucleotide Sequence (Codon Optimized for Human Cells)	Amino Acid sequence
Nucleoplasmin NLS	aaaaggccggcggccacgaaaaaggccggccagg caaaaaagaaaaag (SEQ ID NO: 6)	KRPAATKKAGQAKKKK (SEQ ID NO: 7)
XTEN linker	tctggaggatctagcggaggatcctctggcagcgag acaccaggaacaagcgagtcagcaacaccagagag cagtggcggcagcagcggcggcagcagc (SEQ ID NO: 8)	SGGSSGGSSGSETPGTSESATP ESSGGSSGGSS (SEQ ID NO: 9)
NLS-tagged Nuclease BT	atgaaacggacagccgacggaagcgagttcgagtc accaaagaagaagcggaaagtcagacaccagagac tcgtgtttctcgtggacggtagaaacaacatcctgcat cctaccagaaagctggatatggtcagaagatggctga gacaaacaaacgccaaggtgatcagattcatcggaa aaatcatggtggtccaggtgttcaagatcctgtctcag cagaaagacgaccctacatactggatcggcgtggac ccaggcgacaccatcggcatctgcctggtcagaatca cagccaacaacagagttaagcggctgctgagcatgg aattccacaccagatctaaggaaatcaagcagctgct ggaaacccggagaatgtaccggtgcacacggagat actttagaagaaagcgggtgaagaagaaagggaga attcccaagcacagacctgctagatttgagaacagaa ggcggagcgacggctggctgacccctagcgtgcgc cacttccttgagagccacaagtccgtgatccggaaga tcgccgagtgtgtgccagatgtgaacatcagcctgga gtacgccaagttcgacatccaaaagatgatgaacccc gatatcgagggcgagcagtaccagagaggcaaact gttcggatatcagaatatccgcgcctacgtgctgcata gagataagtacacctgtcagctgtgcgccaagaaaaa cgtgcccttccagctgcaccacgtggtgcctagaagc cagaacggcaaggaccaccccgataacctgatcacc ctgtgtcctgagtgccacgtgaaagtgcacacagaca ccaagctgctgaaaaaactgcagcggaagattaaga gcaagattgctgatttcaccagcaccagcagactgaa tattatcatgagcttcctgaagcactggctggaaaacg agttcggcaaggatcgggtgtggttctgcatcggagc cgactcttatgtgatccggaaggaactgaacctggac aagacacactgcaacgacgcttactgcatcgccatca acaagctggtgagatttgacaccgacctggaacacg aaagcgtggtggataagcaagtgatccggttcgagc agaaaagacagcacaaccggagaatgaaatctctgg agagaaacaggctgtactacgtggtcaaggacaaga agcgcgtgcgggtggccaccaacagaaatggcaat cctgatgccttccacgaggtgtacgacacaaagatcc tgcctctgctgaagcagggctatcggctgaaagtgaa ccctggcgttaagggctacaggaagaagttcagcga ggtgggatttaaggaaggcgccacaatcagaatcgg aaatcgggtggacgtgctgaagggcacctccggctt cggctacacctacgacggcaccatcttcaagctcaag	MKRTADGSEFESPKKKRKVR HQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKI MVVQVFKILSQQKDDPTYWIG VDPGDTIGICLVRITANNRVKR LLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVN ISLEYAKFDIQKMMNPDIEGE QYQRGKLFGYQNIRAYVLHR DKYTCQLCAKKNVPFQLHHV VPRSQNGKDHPDNLITLCPEC HVKVHTDTKLLKKLQRKIKSK IADFTSTSRLNIIMSFLKHWLE NEFGKDRVWFCIGADSYVIRK ELNLDKTHCNDAYCIAINKLV RFDTDLEHESVVDKQVIRFEQ KRQHNRRMKSLERNRLYYVV KDKKRVRVATNRNGNPDAFH EVYDTKILPLLKQGYRLKVNP GVKGYRKKFSEVGFKEGATIR IGNRVDVLKGTSGFGYTYDGT IFKLKDMDVVYPYGGWKCVK VSGGSSGGSSGSETPGTSESAT PESSGGSSGGSSKRPAATKKA GQAKKKK (SEQ ID NO: 11)

Component	Nucleotide Sequence (Codon Optimized for Human Cells)	Amino Acid sequence
	gacatggacgtggtgtacccctacggcggctggaag tgcgtgaaggtgtctggaggatctagcggaggatcct ctggcagcgagacaccaggaacaagcgagtcagca acaccagagagcagtggcggcagcagcggcggca gcagcaaaaggccggcggccacgaaaaaggccgg ccaggcaaaaaagaaaaag (SEQ ID NO: 10)	

RNA guides were designed and cloned into a pUC19 plasmid following the U6 PolIII promoter and terminated with a 6X polyT sequence. RNA guides were designed to be specific to target sequences within the coding exons of AAVS1, EMX1, and VEGFA with 5'-RRT-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** below.

Table 2. Target and RNA Guide Sequences

Guide	PAM	Target Sequence	gRNA Guide Sequence
AAVS1-T2	AAT	AGGTGAGGG ACTTGGGGG GT (SEQ ID NO: 12)	AGGUGAGGGACUUGGGGGGUgggacaugaccaga ugcgaaaguaacugucauguccucaugccaauggacuaugguga ugcaccuaggggugucguuccagcuccuagcucugcgaaggagc auucguuaucggaugcguuuguacuucaagccauagcccaggcc agggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 13)
AAVS1-T6	AAT	GAGCGGGAG AGGCTACAA AG (SEQ ID NO: 14)	GAGCGGGAGAGGCUACAAAGgggacaugaccaga ugcgaaaguaacugucauguccucaugccaauggacuaugguga ugcaccuaggggugucguuccagcuccuagcucugcgaaggagc auucguuaucggaugcguuuguacuucaagccauagcccaggcc agggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 15)
EMX1-T3	GAT	GATAGCATA GGTACACAT TA (SEQ ID NO: 16)	GAUAGCAUAGGUACACAUUAgggacaugaccaga ugcgaaaguaacugucauguccucaugccaauggacuaugguga ugcaccuaggggugucguuccagcuccuagcucugcgaaggagc auucguuaucggaugcguuuguacuucaagccauagcccaggcc agggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 17)

EMX1-T7	GAT	TTAGCAAGG GACTATTCA GG (SEQ ID NO: 18)	UUAGCAAGGGACUAUUCAGGgggacaugaccaga ugcgaaaguaacugucauguccucaugccaauggacuaugguga ugcaccuaggggugucguuccagcuccuagcucugcgaaggagc auucguuaucggaugcguuuguacuucaagccauagcccaggcc agggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 19)
VEGFA-T3	AAT	AGCCGATTC CAGGTGTGG AC (SEQ ID NO: 20)	AGCCGAUUCCAGGUGUGGACgggacaugaccaga ugcgaaguaacugucauguccucaugccaauggacuauggugau gcaccuaggggugucguuccagcuccuagcucugcgaaggagca uucguuaucggaugcguuuguacuucaagccauagcccaggcca gggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 21)
VEGFA-T6	AGT	TGGGGACCA AGGAAGTGT CA (SEQ ID NO: 22)	UGGGGACCAAGGAAGUGUCAgggacaugaccaga ugcgaaaguaacugucauguccucaugccaauggacuaugguga ugcaccuaggggugucguuccagcuccuagcucugcgaaggagc auucguuaucggaugcguuuguacuucaagccauagcccaggcc agggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 23)

* 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 RNA-guided nuclease plasmid (NLS-tagged), RNA guide plasmid, and Opti-MEM™ (Solution 2). In the case of negative controls, the RNA-guided nuclease plasmid was excluded. Solutions 1 and 2 were mixed by pipetting up and down, then incubated at room temperature for 25 minutes. Following incubation, the Solution 1 and 2 mixture was added dropwise to each well of a 96-well plate containing the cells. Approximately 72 hours post transfection, cells were trypsinized by adding TrypLE™ (Thermo Fisher Scientific) to the center of each well and incubating at 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, 68 ºC for 15 minutes, and 98 ºC 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 kit such as 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 QC standard for the minimum number of reads passing filters was 10,000.

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

As shown in **FIG. 1**, each of the six targets tested demonstrated a greater level of indels observed when the RNA-guided nuclease-containing plasmid was added.

Example 2: Effectiveness of Variant RNA-Guided Nucleases for Targeting of Exemplary Mammalian Genes

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

Arginine scanning mutagenesis was performed to individually substitute select non-arginine and non-lysine residues of the reference RNA-guided nuclease (SEQ ID NO: 1) to arginine. SEQ ID NO: 1 is referred to herein as the reference sequence. This resulted in 381 single arginine substitution variants. Nucleic acids encoding the reference RNA-guided nuclease and each RNA-guided nuclease variant were then individually cloned into a pcDNA3.1 backbone (Invitrogen™), and the plasmids were mini-prepped and normalized. The plasmids comprised a CMV promoter, a first NLS (KRTADGSEFESPKKKRKV; SEQ ID NO: 5) upstream of the coding sequence, and an XTEN linker (SGGSSGGSSGSETPGTSESATPESSGGSSGGSS; SEQ ID NO: 9) followed by a second

NLS (KRPAATKKAGQAKKKK; SEQ ID NO: 7) downstream of the coding sequence. See also **Example 1** above.

Exemplary RNA guides of AAVS1-T6, EMX1-T7, and VEGFA-T6 were used in this study. Details of these gRNAs are provided in **Table 2** above. RNA guides were cloned into a pUC19 backbone (New England Biolabs®). The plasmids were purified using a maxi-prep kit and diluted. Cells were transfected, and samples were prepared for NGS as described in **Example 1**. Indel ratios, referring to the fraction of NGS reads containing indels, were calculated for the reference and for each variant. The indel ratios used for fold change calculations were the average of two technical replicates. To then calculate fold change in indel ratios, the indel ratio for each variant was divided by the indel ratio for the reference. **Table 3** shows fold change in indel ratios for each target tested. Numbering is relative to the reference nuclease of SEQ ID NO: 1 (*i.e.*, without an NLS).

As shown in **Table 3**, 9 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 all targets (right column).

Table 3. Fold Change in Indel Ratios*

Substitution	AAVS1	EMX1	VEGFA	Average Fold Change
V433R	1.83	1.56	2.99	2.13
E105R	1.90	2.34	1.63	1.96
Q102R	1.68	2.25	1.74	1.89
P244R	0.45	0.61	4.39	1.82
E432R	1.54	1.40	2.34	1.76
T113R	2.23	0.73	2.12	1.69
I272R	1.42	1.80	1.78	1.67
I206R	1.40	2.03	1.41	1.61
V256R	1.30	1.33	1.88	1.51

* Variant indel ratio/Reference indel ratio

74 variants were analyzed as having indel ratios 1-1.5X of the reference indel ratios, when averaged across all targets: E99R, E191R, E401R, T106R, Y415R, S431R, N239R, I248R, T33R, E299R, G434R, L410R, L103R, Q413R, M368R, P252R, Y116R, A277R, D404R, E193R, G455R, V419R, T260R, E137R, S290R, H294R, F291R, N298R, N374R, I464R, M109R, V386R, L211R, N365R, V423R, Y403R, D59R, L411R, H400R, P409R, N445R, D344R, D143R, T262R, S237R, N172R, F300R, N420R, Q219R, V351R, T405R, C167R, N225R, V388R, V233R, S159R, Q204R, E253R, A398R, F341R, V380R, T280R,

D303R, T390R, D261R, V350R, L408R, N395R, E166R, A81R, C112R, G414R, D382R, and I52R.

The remaining variants (298 variants) resulted in decreased indel ratios relative to the reference indel ratios (fold change in indel ratios of less than 1.0): T217R, T343R, L104R, E346R, Y216R, P227R, G144R, D323R, N83R, D342R, L338R, I128R, P133R, Y378R, V402R, N321R, Q238R, F307R, G394R, A222R, N82R, D189R, L154R, V168R, P169R, D397R, V339R, A165R, S142R, W295R, E319R, S349R, G301R, F93R, T80R, G444R, Q354R, L53R, N393R, S274R, V160R, D170R, W63R, P475R, S54R, L220R, E297R, H94R, Q56R, F435R, G462R, N34R, V210R, L230R, I74R, P421R, Q55R, V379R, V226R, V305R, H347R, W306R, D471R, H131R, F40R, Q269R, E372R, L247R, V171R, L345R, V447R, T61R, Y314R, I276R, D58R, D352R, E359R, S454R, V473R, D242R, D448R, G126R, I443R, H3R, F50R, Q48R, I309R, I79R, A134R, T463R, V355R, P60R, F456R, G12R, S90R, P129R, L264R, D24R, Q32R, V484R, V315R, N187R, C481R, I356R, Q196R, V234R, I44R, W479R, G192R, H364R, E155R, M186R, L88R, Q360R, I41R, N15R, F117R, V37R, I38R, Q4R, I16R, M91R, H152R, H18R, Y426R, H243R, W29R, P396R, V482R, M45R, I64R, Q194R, L89R, T453R, L376R, V122R, F399R, I164R, A208R, L320R, V85R, F430R, I173R, P188R, V10R, D461R, D469R, N14R, E348R, V49R, N336R, L23R, L467R, L268R, L200R, F228R, L296R, I161R, V472R, D70R, L6R, P148R, M25R, G202R, L17R, A311R, D11R, F358R, V46R, T325R, L146R, C327R, F201R, A35R, L30R, Y474R, F136R, S149R, V150R, F8R, W145R, H212R, Y177R, P19R, Y203R, S156R, I100R, N138R, L76R, G42R, T71R, M470R, S281R, L265R, Y331R, S174R, T282R, N391R, L322R, C221R, D245R, Q229R, C75R, Y62R, F465R, T459R, V7R, G425R, Y110R, E437R, T440R, L292R, I72R, P235R, Y195R, N328R, V47R, L417R, T95R, Y476R, Y460R, V26R, S370R, D278R, N205R, T147R, I316R, Y458R, G240R, M185R, Q363R, G310R, D312R, Q183R, I441R, L9R, L371R, V66R, C308R, I287R, G478R, L250R, A439R, F279R, I288R, E92R, A334R, D67R, T249R, G73R, Y209R, G477R, M289R, V449R, I407R, I335R, N246R, P68R, E176R, T20R, C218R, D329R, G452R, L175R, Y377R, V258R, V77R, C251R, H232R, G457R, G198R, I333R, A178R, H255R, A389R, H157R, N286R, S283R, L285R, S313R, H259R, H326R, G422R, G69R, F153R, S97R, G65R, L450R, G438R, C332R, H231R, I190R, D181R, C254R, A330R, D214R, I182R, and F180R.

Based on this experiment, the following substitutions were selected for further engineering: V433R, E105R, Q102R, E432R, T113R, I272R, I206R, V256R, E99R, E191R, E401R, N239R, T106R, and A277R.

Example 3: Effectiveness of Combination RNA-Guided Nuclease Variants for Targeting of Mammalian Genes

This Example describes indel assessment on mammalian targets using RNA-guided nuclease variants comprising two or more substitutions identified as increasing indel activity in **Example 2**. 29 combinations of RNA-guided nuclease variants were tested.

Each RNA-guided nuclease variant and RNA guide was cloned as described in **Example 2**. Exemplary RNA guides of AAVS1-T6, EMX1-T7, and VEGFA-T6 were used in this study. Details of these gRNAs are provided in **Table 2** above. HEK293T cells were further transfected, followed by NGS analysis, as described in **Example 2**. For each target, the percentage of NGS reads comprising indels were calculated for the reference RNA-guided nuclease (SEQ ID NO: 1) and for each variant RNA-guided nuclease. The amino acid sequences of the variant RNA-guided nucleases tested herein are provided in **Table 4A** (substitution positions relative to SEQ ID NO: 1) and the percent of reads containing indels achieved by these variants are shown in **Table 4B**, which were calculated as the average of three technical replicates unless indicated otherwise.

Table 4A: Amino Acid Sequences of Exemplary RNA-Guided Nuclease Variants

Variants	Amino Acid Sequences	SEQ ID NO
Q102R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	82 *With N-terminus M residue (SEQ ID NO: 83)*
E105R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	84 *With N-terminus M residue (SEQ ID NO: 85)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
Q102R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	86 *With N-terminus M residue (SEQ ID NO: 87)*
Q102R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	88 *With N-terminus M residue (SEQ ID NO: 89)*
Q102R, E191R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	90 *With N-terminus M residue (SEQ ID NO: 91)*
Q102R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	92 *With N-terminus M residue (SEQ ID NO: 93)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
E105R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	94 *With N-terminus M residue (SEQ ID NO: 95)*
E105R, E191R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	96 *With N-terminus M residue (SEQ ID NO: 97)*
E105R, E191R, I272R, E401R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHRVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	98 *With N-terminus M residue (SEQ ID NO: 99)*
E105R, E191R, N239R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQRGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	100 *With N-terminus M residue (SEQ ID NO: 101)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	102 *With N-terminus M residue (SEQ ID NO: 103)*
E105R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	104 *With N-terminus M residue (SEQ ID NO: 105)*
E105R, T106R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRRRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	106 *With N-terminus M residue (SEQ ID NO: 107)*
E105R, E191R, I272R, E432R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	108 *With N-terminus M residue (SEQ ID NO: 109)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSRRGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
Q102R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	110 *With N-terminus M residue (SEQ ID NO: 111)*
E105R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	112 *With N-terminus M residue (SEQ ID NO: 113)*
I206R, I272R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	114 *With N-terminus M residue (SEQ ID NO: 115)*
E105R, E191R, I272R, A277R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIRDFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	116 *With N-terminus M residue (SEQ ID NO: 117)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
Q102R, E105R, E191R, I206R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	118 *With N-terminus M residue (SEQ ID NO: 119)*
Q102R, E105R, E191R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	120 *With N-terminus M residue (SEQ ID NO: 121)*
I206R, V256R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	122 *With N-terminus M residue (SEQ ID NO: 123)*
Q102R, E105R, E191R, I206R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	124 *With N-terminus M residue (SEQ ID NO: 125)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
E105R, T113R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCRRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	126 *With N-terminus M residue (SEQ ID NO: 127)*
I272R, V256R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	128 *With N-terminus M residue (SEQ ID NO: 129)*
Q102R, E105R, E191R, I206R, V256R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	130 *With N-terminus M residue (SEQ ID NO: 131)*
Q102R, E105R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY	132 *With N-terminus M residue (SEQ ID NO: 133)*

Variants	Amino Acid Sequences	SEQ ID NO
	RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	
Q102R, E105R, E191R, V256R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	134 *With N-terminus M residue (SEQ ID NO: 135)*
Q102R, E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	136 *With N-terminus M residue (SEQ ID NO: 137)*
E99R, E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKRIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	138 *With N-terminus M residue (SEQ ID NO: 139)*
Nuclease BT	See **Table 1** above	1 or 140

Table 4B. Indel Percentages for Mammalian Targets

Variants	AAVS1 Average Indel Percentage	EMX1 Average Indel Percentage	VEGFA Average Indel Percentage	Average Indel Percentage (All Targets)
Q102R, I206R, V433R	1.27	3.10	2.92	2.43
E105R, I206R, V433R	1.36	2.99	2.63	2.33
Q102R, E191R, I272R, V433R	1.22	2.91	2.73	2.29
Q102R, V256R, V433R	1.04	3.08	2.65	2.26
Q102R, E191R, I206R, V433R	1.26	2.74	2.66	2.22
Q102R, I272R, V433R	1.00	3.03	2.43	2.15
E105R, V256R, V433R	1.19	2.90	2.22	2.10
E105R, E191R, V256R, V433R	1.21	2.72	2.36	2.10
E105R, E191R, I272R, E401R, V433R	1.14	2.85	2.25	2.08
E105R, E191R, N239R, I272R, V433R	1.14	2.76	2.12	2.01
E105R, E191R, I272R, V433R	0.94	2.76	2.25	1.98
E105R, I272R, V433R	0.92	2.68	2.16	1.92
E105R, T106R, E191R, I272R, V433R	1.07	2.19	2.18	1.81
E105R, E191R, I272R, E432R, V433R	0.88	2.56	1.89	1.78
Q102R, V433R	0.77	2.59	1.96	1.77
E105R, V433R	1.00	2.36	1.96	1.77
I206R, I272R	0.91	1.61	2.46	1.66
E105R, E191R, I272R, A277R, V433R	0.74	2.06	2.06	1.62
Q102R, E105R, E191R, I206R, V256R, V433R	0.73	2.41	1.63	1.59
Q102R, E105R, E191R, I206R, V433R	0.83	2.38	1.55	1.58

Variants	AAVS1 Average Indel Percentage	EMX1 Average Indel Percentage	VEGFA Average Indel Percentage	Average Indel Percentage (All Targets)
I206R, V256R	1.03	1.60	2.12	1.58
Q102R, E105R, E191R, I206R, I272R, V433R	0.72	2.56	1.44	1.57
E105R, T113R, E191R, I272R, V433R	1.09	2.33	1.28	1.57
I272R, V256R	0.63*	1.71	2.27	1.54
Q102R, E105R, E191R, I206R, V256R, I272R, V433R	0.73	2.49	1.36	1.53
Q102R, E105R, V433R	0.65	2.19	1.65	1.50
Q102R, E105R, E191R, V256R, I272R, V433R	0.64	2.45	1.35	1.48
Q102R, E105R, E191R, I272R, V433R	0.57	2.48	1.33	1.46
E99R, E105R, E191R, I272R, V433R	0.38	1.55	0.85	0.93
Nuclease BT	0.56	1.09	1.02	0.89

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

As shown in **Table 4B**, each of the RNA-guided nuclease variants with combinations of amino acid substitutions exhibited higher indel activity than the reference RNA-guided nuclease (SEQ ID NO: 1). 10 RNA-guided nuclease variants resulted in indel percentages of over 2% when averaged across all three targets. These 10 RNA-guided nuclease variants comprised the following substitution combinations: a) Q102R, I206R, and V433R; b) Q102R, V256R, and V433R; c) Q102R, E191R, I272R, and V433R; d) Q102R, V256R, and V433R; e) Q102R, E191R, I206R, and V433R; f) Q102R, I272R, and V433R; g) E105R, V256R, and V433R; h) E105R, E191R, V256R, and V433R; i) E105R, E191R, I272R, E401R, and 433R; j) E105R, E191R, N239R, I272R, and V433R. 18 RNA-guided nuclease variants resulted in indel percentages between 1% and 2% when averaged across all three targets. 1 RNA-guided nuclease variants resulted in indel percentages below 1% when averaged across all three targets. The average indel ratio across all three targets exceeded that of the reference for all variants tested.

Based on this experiment, the top-performing RNA-guided nuclease variant comprising substitutions Q102R, I206R, and V433R was selected for further testing. This RNA-guided

nuclease variant exhibited a 2.7-fold increase in indel activity compared to the reference RNA-guided nuclease (Nuclease BT).

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 an exemplary target gene, VEGFA, by the RNA-guided nuclease variant comprising substitutions Q102R, I206R, and V433R relative to SEQ ID NO: 1 and the N-terminal M was removed (the amino acid sequence of the variant is shown below) and RNA guides comprising alternative scaffold sequences and spacer lengths. **Table 5A** below provides amino acid sequences of the RNA-guided nuclease variants used in this example.

Table 5A. Amino Acid Sequence of Exemplary RNA-Guided Nuclease Variants

Variants	Amino Acid Sequences	SEQ ID NO
Q102R, I206R, V433R (with no N-terminus M residue)	See **Table 4A** above	82
NLS-tagged variant	MKRTADGSEFESPKKKRKVRHQRLVFLVDGRNNILHPTRKLDMV RRWLRQTNAKVIRFIGKIMVVQVFKILSQQKDDPTYWIGVDPGD TIGICLVRITANNRVKRLLSMEFHTRSKEIKRLLETRRMYRCTR RYFRRKRVKKKGRIPKHRPARFENRRRSDGWLTPSVRHFLESHK SVIRKIAECVPDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFG YQNRRAYVLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDN LITLCPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMS FLKHWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNRLYY VVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGYRLKVNPG VKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGFGYTYDGTIFK LKDMDVVYPYGGWKCVKVSGGSSGGSSGSETPGTSESATPESSG GSSGGSSKRPAATKKAGQAKKKK	141

RNA guides were designed using the RNA scaffold sequences of **Table 5B** and cloned into a pUC19 plasmid following the U6 PolIII promoter and terminated with a 6x polyT sequence. The structures of the reference RNA scaffold (SEQ ID NO: 2) and the structures of the scaffold variants are illustrated in FIGs. 3A-3J. RNA guides were designed to be specific to the VEGFA-T6, AAVS1-T6, and EMX1-T7 target sequences (*see* **Example 1**) and utilized spacers with lengths between 11-nucleotides and 20-nucleotides. A 16-nucleotide spacer was used most

commonly. 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 exemplary spacer sequences for each target in **Table 5B**. Exemplary RNA guide sequences utilizing a 16-nt VEGFA-T6 spacer are shown in **Table 6**.

Table 5B. RNA Guide Component Sequences

Component	Sequence
20-nt VEGFA-T6 spacer	UGGGGACCAAGGAAGUGUCA (SEQ ID NO: 142)
16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCA (SEQ ID NO: 24)
20-nt AAVS1-T6 spacer	GAGCGGGAGAGGCUACAAAG (SEQ ID NO: 143)
16-nt AAVS1-T6 spacer	GGGAGAGGCUACAAAG (SEQ ID NO: 144)
16-nt EMX1-T7 spacer	CAAGGGACUAUUCAGG (SEQ ID NO: 145)
Reference Scaffold	SEQ ID NO: 2
Scaffold 1	gggacaugaccaggaaacugucauguccucaugccaauggacuauggugaugcacc uaggggugucguuccagcuccuagcucugcgaaggagcauucggaaacggaugcgu uuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauagga gc (SEQ ID NO: 25)
Scaffold 1a	gggacaugaccaggaaacugucauguccucaugccaauggacuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 146)
Scaffold 2	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 26)
Scaffold 3	gggacaugaccaggaaacugucauguccucauggcaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccagc ccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 147)
Scaffold 4	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 148)
Scaffold 5	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacgua (SEQ ID NO: 149)
Scaffold 6	gggacaugacggaaacgucauguccucaugccaaugggcuauggugaugcaccuag gggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccaggcca gggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 150)
Scaffold 7	gggacaugacggaaacgucauguccucauggcaaugggcuauggugaugcaccuag gggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccagccca gggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 151)
Scaffold 8	gggacaugacggaaacgucauguccucauggcaaugggcuauggugaugcaccuag cggugucguuccagcugcuagcucugcgccagaaauggcaagccauagcccagccca gggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 152)

Table 6. Exemplary RNA Guide Sequences

Description	Sequence
Reference scaffold with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccagaugcgaaaguaacugucaug uccucaugccaauggacuauggugaugcaccuaggggugucguuccagcuccuagcu cugcgaaggagcauucguuaucggaugcguuuguacuucaagccauagcccaggcca gggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 27)
Scaffold 1 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaauggacuauggugaugcaccuaggggugucguuccagcuccuagcucugcgaag gagcauucggaaacggaugcguuuguacuucaagccauagcccaggccagggacgua uuacugcuccugcccuauaggagc (SEQ ID NO: 28)
Scaffold 1a with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaauggacuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 153)
Scaffold 2 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 29)
Scaffold 3 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug gcaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccagcccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 154)
Scaffold 4 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 155)
Scaffold 5 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacgua (SEQ ID NO: 156)
Scaffold 6 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucaugcca augggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgccagaa auggcaagccauagcccaggccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 157)
Scaffold 7 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucauggca augggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgccagaa auggcaagccauagcccagcccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 158)
Scaffold 8 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucauggca augggcuauggugaugcaccuagcgguucguuccagcugcuagcucugcgccagaa auggcaagccauagcccagcccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 159)

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

The RNA-guided nuclease comprising Q102R, I206R, and V433R substitutions and RNA guides of **Table 6** were introduced into HEK293T cells by lipid-based transient transfection as described in **Example 1**. Transfections were performed across multiple experiments, representing distinct engineering rounds. The first round of engineering focused

on identifying the ideal spacer length for the RNA guide. The VEGFA-T6 RNA and AAVS1-T6 guides from **Example 1**, which each utilize a 20-nucleotide spacer, were used as a control in the first engineering round. Subsequent engineering rounds focused on optimizing the scaffold sequence and utilized one or more RNA guides from prior engineering rounds as bridging 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 **Tables 7-10** were calculated as the average of three technical replicates, unless indicated otherwise.

Table 7. Q102R, I206R, V433R Variant CRISPR Nuclease Indel Activity (RNA Engineering Round 1)

RNA Guide	VEGFA-T6 Average Indel Percentage	AAVS1-T6 Average Indel Percentage
Reference scaffold with 20-nt spacer	2.91	1.47
Reference scaffold with 19-nt spacer	2.25	1.35
Reference scaffold with 18-nt spacer	3.30	2.33
Reference scaffold with 17-nt spacer	4.64	1.57
Reference scaffold with 16-nt spacer	5.24*	2.04
Reference scaffold with 15-nt spacer	5.22	2.58
Reference scaffold with 14-nt spacer	0.40	2.00*
Reference scaffold with 13-nt spacer	0.24	1.38
Reference scaffold with 12-nt spacer	0.24	0.31
Reference scaffold with 11-nt spacer	0.26	0.32
No-guide control	0.27	0.29

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

As shown in **Table 7**, the RNA guide comprising the reference RNA scaffold and spacer of a length greater than or equal to 15-nucleotides introduced indels above background levels at VEGFA-T6. Spacer lengths greater than or equal to 13-nucleotides were sufficient to introduce indels at AAVS1-T6. Across both targets, spacers of approximately 15-nucleotides to 16-nucleotides in length introduced greater levels of indels than other spacer lengths tested. Based on these results, RNA guides were designed with 16-nucleotide spacers for subsequent engineering rounds. These subsequent rounds focused on optimizing the scaffold sequence through iterative rational mutagenesis. Putative secondary structures of the modified scaffolds are depicted and annotated in **FIGs. 3B-3J**. Results are shown in **Tables 8-10**.

Table 8. Q102R, I206R, V433R Variant CRISPR Nuclease Indel Activity (RNA Engineering Round 2)

RNA Guide	VEGFA-T6 Average Indel Percentage
Reference scaffold with 16-nt spacer	8.93
Scaffold 1 with 16-nt spacer	9.27
Scaffold 1a with 16-nt spacer	14.77
Scaffold 2 with 16-nt spacer	18.03
No-guide control	0.06

As shown in **Table 8**, all three new scaffolds maintained or improved activity relative to the Reference Scaffold sequence at VEGFA-T6. Scaffold 1 introduced a truncated P1 helix, replaced the loop of P4b with a GNRA tetraloop (*e.g.* GAAA tetraloop; SEQ ID NO: 214), and removed 3 nucleotides from the 3' end of the scaffold. This resulted in similar activity to the reference scaffold. Scaffold 1a replaced the P4 helix of Scaffold 1 with a significantly truncated form. This change led to a 1.65X increase in indels relative to the reference scaffold. Scaffold 2 further introduced an AC-to-GC base pair mutation into the P2b helix of Scaffold 1a. This led to a 2X improvement in indels relative to the reference scaffold. Scaffold 2 was used as the parent sequence for the next round of optimizations.

Table 9. Q102R, I206R, V433R Variant CRISPR Nuclease Indel Activity (RNA Engineering Round 3)

RNA Guide	VEGFA-T6 Average Indel Percentage	AAVS1-T6 Average Indel Percentage	EMX1-T7 Average Indel Percentage
Reference scaffold with 16-nt spacer	7.72*	2.44**	-
Scaffold 2 with 16-nt spacer	12.77*	5.12	11.84
Scaffold 3 with 16-nt spacer	13.61	5.68	11.27
Scaffold 4 with 16-nt spacer	13.95	-	-
Scaffold 5 with 16-nt spacer	5.17	1.13	3.41
Scaffold 6 with 16-nt spacer	13.66	-	-
No-guide control	0.10	0.06	0.51

"-" = Not tested

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

***Two technical replicates were excluded because fewer than 10,000 NGS reads were generated*

As shown in **Table 9**, three of the new scaffolds, led to a modest increase in activity relative to Scaffold 2 at one or more target sites. These were Scaffolds 3, 4, and 6, which each introduced small modifications to Scaffold 2. Scaffolds 3 introduced a CG-to-GC mutation into the P2a helix, Scaffold 4 removed a single nucleotide from just upstream of the P6 helix, and Scaffold 6 introduced a further truncated P1 helix. Although it showed reduced activity, Scaffold 5 is notable as it showed activity above background levels despite the removal of P6. This indicates that the P6 helix is beneficial but not strictly required for activity. Based on these results, Scaffold 3 was chosen as the parent sequence for the final round of engineering.

Table 10. Q102R, I206R, V433R Variant CRISPR Nuclease Indel Activity (RNA Engineering Round 4)

RNA Guide	VEGFA-T6 Average Indel Percentage	AAVS1-T6 Average Indel Percentage	EMX1-T7 Average Indel Percentage
Scaffold 3 with 16-nt spacer	15.17	4.32	8.13*
Scaffold 7 with 16-nt spacer	16.52	5.30	8.02
Scaffold 8 with 16-nt spacer	16.89	4.93	8.20*
No-guide control	0.16	0.14	0.58**

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

***Two technical replicates were excluded because fewer than 10,000 NGS reads were generated*

As shown in **Table 10**, both Scaffolds 7 and 8 either resulted in similar or slightly improved levels of indels relative to Scaffold 3 at the three targets tested. Scaffold 7 combined the individual modifications described for Scaffolds 3, 4, and 6. Scaffold 8 included these modifications as well as a GC-to-CG mutation in the P3b helix.

This Example thus shows that spacers of approximately 15-nucleotides or longer are sufficient for RNA-guided nuclease activity by this nuclease. Spacers around 15-to-16-nucleotides in length can yield increased indel activity compared to longer spacers, such as a 20-nucleotide spacer. Additionally, scaffolds 1-8 are capable of being recognized by the RNA-guided nuclease for targeting mammalian genes. Scaffolds 1a, 2, 3, 4, 6, 7, and 8 exhibit increased indel activity over the reference scaffold sequence.

Example 5: Engineering and Effectiveness of Nickase Variants of RNA-Guided Nuclease for Targeting Mammalian Genes

This Example describes introducing mutations into the RNA-guided nuclease of SEQ ID NO: 1 that disrupt either the HNH or RuvC domains to produce a functional nickase. H231, H232, and H255 were identified as putative catalytic residues of the HNH domain. D67, E176,

and D329 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 RNA-guided 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 H232. 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. RNA-Guided Nuclease and Nickase Sequences

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
Nuclease BT	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttgcaccacgtggtcccgcgttcgca	SEQ ID NO: 1 or SEQ ID NO: 140 (with N-terminal M residue removed) See **Table 1** above.

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	aaacggtaaggaccacccggacaaccttatcacgctgt gcccggaatgccatgttaaggttcataccgataccaaact cttgaagaaactgcagcggaaaattaaaagcaaaatcgc tgattttacctccacctctcgtctgaacatcattatgagcttt ctgaagcattggctggagaacgaattcggcaaagaccg tgtgtggttctgcataggcgcggatagctatgtgatccgc aaggagctgaatctcgataagacccattgtaatgatgcgt actgcattgcgattaataaattggttcgtttcgatactgatct ggagcatgaatctgttgtggataagcaagtgatccgtttc gagcaaaagcgccagcataatagacgtatgaaatccct ggaacgtaaccgcctgtattacgtagtaaaagacaagaa gcgtgtgcgcgtggcgacgaaccgcaacggtaacccc gacgcctttcacgaggtttacgacaccaagatcctgccg ctgctgaaacagggttaccgtttgaaggtaaatccaggtg tgaagggttaccgtaagaagttcagcgaagttggtttcaa agagggtgctaccatccgcatcggcaaccgtgtggacg tgctgaagggtacttcaggttttggttacacgtacgacggt accatttttaagcttaaggacatggatgtggtgtatccgtat ggcggttggaaatgcgttaaggtg (SEQ ID NO: 30)	
H231A (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttg**GCG**cacgtggtcccgcgttc gcaaaacggtaaggaccacccggacaaccttatcacgc tgtgcccggaatgccatgttaaggttcataccgataccaa actcttgaagaaactgcagcggaaaattaaaagcaaaat cgctgattttacctccacctctcgtctgaacatcattatgag ctttctgaagcattggctggagaacgaattcggcaaaga ccgtgtgtggttctgcataggcgcggatagctatgtgatc cgcaaggagctgaatctcgataagacccattgtaatgat gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY TCQLCAKKNVPFQL**A**HVVPRS QNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 31) *With no N-terminus M residue (SEQ ID NO: 160)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt ggacgtgctgaagggtacttcaggttttggttacacgtac gacggtaccatttttaagcttaaggacatggatgtggtgta tccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 32)	
H232A (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttgcac**GCG**gtggtcccgcgttc gcaaaacggtaaggaccacccggacaaccttatcacgc tgtgcccggaatgccatgttaaggttcataccgataccaa actcttgaagaaactgcagcggaaaattaaaagcaaaat cgctgattttacctccacctctcgtctgaacatcattatgag ctttctgaagcattggctggagaacgaattcggcaaaga ccgtgtgtggttctgcataggcgcggatagctatgtgatc cgcaaggagctgaatctcgataagacccattgtaatgat gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt ggacgtgctgaagggtacttcaggttttggttacacgtac gacggtaccatttttaagcttaaggacatggatgtggtgta tccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 34)	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY TCQLCAKKNVPFQLH**A**VVPRS QNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 33) *With no N-terminus M residue (SEQ ID NO: 161)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
H232L (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttgcac**CTG**gtggtcccgcgttcg caaaacggtaaggaccacccggacaaccttatcacgct gtgcccggaatgccatgttaaggttcataccgataccaaa ctcttgaagaaactgcagcggaaaattaaaagcaaaatc gctgattttacctccacctctcgtctgaacatcattatgagc tttctgaagcattggctggagaacgaattcggcaaagac cgtgtgtggttctgcataggcgcggatagctatgtgatcc gcaaggagctgaatctcgataagacccattgtaatgatg cgtactgcattgcgattaataaattggttcgtttcgatactg atctggagcatgaatctgttgtggataagcaagtgatccg tttcgagcaaaagcgccagcataatagacgtatgaaatc cctggaacgtaaccgcctgtattacgtagtaaaagacaa gaagcgtgtgcgcgtggcgacgaaccgcaacggtaac cccgacgcctttcacgaggtttacgacaccaagatcctg ccgctgctgaaacagggttaccgtttgaaggtaaatcca ggtgtgaagggttaccgtaagaagttcagcgaagttggtt tcaaagagggtgctaccatccgcatcggcaaccgtgtg gacgtgctgaagggtacttcaggttttggttacacgtacg acggtaccatttttaagcttaaggacatggatgtggtgtat ccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 36)	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY TCQLCAKKNVPFQLH**L**VVPRS QNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 35) *With no N-terminus M residue (SEQ ID NO: 162)*
H255A (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttgcaccacgtggtcccgcgttcgca aaacggtaaggaccacccggacaaccttatcacgctgt gcccggaatgc**GCG**gttaaggttcataccgataccaa actcttgaagaaactgcagcggaaaattaaaagcaaaat cgctgattttacctccacctctcgtctgaacatcattatgag ctttctgaagcattggctggagaacgaattcggcaaaga ccgtgtgtggttctgcataggcgcggatagctatgtgatc cgcaaggagctgaatctcgataagacccattgtaatgat gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt ggacgtgctgaagggtacttcaggttttggttacacgtac gacggtaccatttttaagcttaaggacatggatgtggtgta tccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 38)	TCQLCAKKNVPFQLHHVVPRS QNGKDHPDNLITLCPEC**A**VKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 37) *With no N-terminus M residue (SEQ ID NO: 163)*
D67A (putative RuvC-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt g**GCG**ccgggcgatacgatcggcatttgcctggttcgt atcaccgcaaataaccgcgtcaagcgcttgctgtccatg gaattccacaccagaagcaaagagatcaaacaactgtta gagacacgccgtatgtatcgttgtacccgtcgttatttccg ccggaagcgcgtgaagaagaagggcagaattccgaaa caccgtccggcacgtttcgaaaaccgtcgccgctctgat ggttggttgaccccaagcgtgcgccacttttggagagc cataaatcggtcattcgtaagatcgccgaatgtgttccgg acgttaatatcagcttggagtatgctaaattcgacatccag aaaatgatgaacccggatattgaaggtgaacaatatcaa cgtggcaaactgtttggctaccaaaacattcgtgcgtacg ttctgcaccgcgataagtatacctgtcagctgtgcgcgaa aaaaaatgttccgtttcagttgcaccacgtggtcccgcgtt cgcaaaacggtaaggaccacccggacaaccttatcacg ctgtgcccggaatgccatgttaaggttcataccgatacca	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GV**A**PGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY TCQLCAKKNVPFQLHHVVPRS QNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	aactcttgaagaaactgcagcggaaaattaaaagcaaaa tcgctgattttacctccacctctcgtctgaacatcattatga gctttctgaagcattggctggagaacgaattcggcaaag accgtgtgtggttctgcataggcgcggatagctatgtgat ccgcaaggagctgaatctcgataagacccattgtaatgat gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt ggacgtgctgaagggtacttcaggttttggttacacgtac gacggtaccatttttaagcttaaggacatggatgtggtgta tccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 40)	SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 39) *With no N-terminus M residue (SEQ ID NO: 164)*
E176A (putative RuvC-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttg**GCG**tatgctaaattcgacatccagaa aatgatgaacccggatattgaaggtgaacaatatcaacgt ggcaaactgtttggctaccaaaacattcgtgcgtacgttct gcaccgcgataagtatacctgtcagctgtgcgcgaaaaa aaatgttccgtttcagttgcaccacgtggtcccgcgttcgc aaaacggtaaggaccacccggacaacccttatcacgctgt gcccggaatgccatgttaaggttcataccgataccaaact cttgaagaaactgcagcggaaaattaaaagcaaaatcgc tgattttacctccacctctcgtctgaacatcattatgagcttt ctgaagcattggctggagaacgaattcggcaaagaccg tgtgtggttctgcataggcgcggatagctatgtgatccgc aaggagctgaatctcgataagacccattgtaatgatgcgt actgcattgcgattaataaattggttcgtttcgatactgatct ggagcatgaatctgttgtggataagcaagtgatccgtttc gagcaaaagcgccagcataatagacgtatgaaatccct ggaacgtaaccgcctgtattacgtagtaaaagacaagaa gcgtgtgcgcgtggcgacgaaccgcaacggtaacccc	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SL**A**YAKFDIQKMMNPDIEGEQ YQRGKLFGYQNIRAYVLHRDK YTCQLCAKKNVPFQLHHVVPR SQNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 41) *With no N-terminus M residue (SEQ ID NO: 165)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	gacgcctttcacgaggtttacgacaccaagatcctgccg ctgctgaaacagggttaccgtttgaaggtaaatccaggtg tgaagggttaccgtaagaagttcagcgaagttggtttcaa agagggtgctaccatccgcatcggcaaccgtgtggacg tgctgaagggtacttcaggttttggttacacgtacgacggt accatttttaagcttaaggacatggatgtggtgtatccgtat ggcggttggaaatgcgttaaggtg (SEQ ID NO: 42)	
D329A (putative RuvC-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttgcaccacgtggtcccgcgttcgca aaacggtaaggaccacccggacaaccttatcacgctgt gcccggaatgccatgttaaggttcataccgataccaaact cttgaagaaactgcagcggaaaattaaaagcaaaatcgc tgattttacctccacctctcgtctgaacatcattatgagcttt ctgaagcattggctggagaacgaattcggcaaagaccg tgtgtggttctgcataggcgcggatagctatgtgatccgc aaggagctgaatctcgataagacccattgtaat**GCG**g cgtactgcattgcgattaataaattggttcgtttcgatactg atctggagcatgaatctgttgtggataagcaagtgatccg tttcgagcaaaagcgccagcataatagacgtatgaaatc cctggaacgtaaccgcctgtattacgtagtaaaagacaa gaagcgtgtgcgcgtggcgacgaaccgcaacggtaac cccgacgcctttcacgaggtttacgacaccaagatcctg ccgctgctgaaacagggttaccgtttgaaggtaaatcca ggtgtgaagggttaccgtaagaagttcagcgaagttggtt tcaaagagggtgctaccatccgcatcggcaaccgtgtg gacgtgctgaagggtacttcaggttttggttacacgtacg acggtaccatttttaagcttaaggacatggatgtggtgtat ccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 44)	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQNIRAYVLHRDKY TCQLCAKKNVPFQLHHVVPRS QNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCN**A**AYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SEVGFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 43) *With no N-terminus M residue (SEQ ID NO: 166)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
H232A + Q102R, I206R, and V433R	atgagacaccagagactcgtgtttctcgtggacggtaga aacaacatcctgcatcctaccagaaagctggatatggtc agaagatggctgagacaaacaaacgccaaggtgatca gattcatcggaaaaatcatggtggtccaggtgttcaagat cctgtctcagcagaaagacgaccctacatactggatcgg cgtggacccaggcgacaccatcggcatctgcctggtca gaatcacagccaacaacagagttaagcggctgctgagc atggaattccacaccagatctaaggaaatcaag**CGG**c tgctggaaacccggagaatgtaccggtgcacacggaga tactttagaagaaagcgggtgaagaagaaagggagaat tcccaagcacagacctgctagatttgagaacagaaggc ggagcgacggctggctgacccctagcgtgcgccacttc cttgagagccacaagtccgtgatccggaagatcgccga gtgtgtgccagatgtgaacatcagcctggagtacgccaa gttcgacatccaaaagatgatgaaccccgatatcgaggg cgagcagtaccagagaggcaaactgttcggatatcaga at**CGG**cgcgcctacgtgctgcatagagataagtacac ctgtcagctgtgcgccaagaaaaacgtgcccttccagct gcac**GCC**gtggtgcctagaagccagaacggcaagg accaccccgataacctgatcaccctgtgtcctgagtgcc acgtgaaagtgcacacagacaccaagctgctgaaaaaa ctgcagcggaagattaagagcaagattgctgatttcacc agcaccagcagactgaatattatcatgagcttcctgaagc actggctggaaaacgagttcggcaaggatcgggtgtgg ttctgcatcggagccgactcttatgtgatccggaaggaac tgaacctggacaagacacactgcaacgacgcttactgca tcgccatcaacaagctggtgagatttgacaccgacctgg aacacgaaagcgtggtggataagcaagtgatccggttc gagcagaaaagacagcacaaccggagaatgaaatctct ggagagaaacaggctgtactacgtggtcaaggacaaga agcgcgtgcgggtggccaccaacagaaatggcaatcct gatgccttccacgaggtgtacgacacaaagatcctgcct ctgctgaagcagggctatcggctgaaagtgaaccctgg cgttaagggctacaggaagaagttcagcgag**CGG**g gatttaaggaaggcgccacaatcagaatcggaaatcgg gtggacgtgctgaagggcacctccggcttcggctacac ctacgacggcaccatcttcaagctcaaggacatggacgt ggtgtacccctacggcggctggaagtgcgtgaaggtg (SEQ ID NO: 216)	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GVDPGDTIGICLVRITANNRVK RLLSMEFHTRSKEIK**R**LLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY QRGKLFGYQN**R**RAYVLHRDK YTCQLCAKKNVPFQLH**A**VVPR SQNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SE**R**GFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 215) *With no N-terminus M residue (SEQ ID NO: 167)*
D67A + Q102R, I206R, and V433R	atgagacaccagagactcgtgtttctcgtggacggtaga aacaacatcctgcatcctaccagaaagctggatatggtc agaagatggctgagacaaacaaacgccaaggtgatca gattcatcggaaaaatcatggtggtccaggtgttcaagat cctgtctcagcagaaagacgaccctacatactggatcgg cgtg**GCC**ccaggcgacaccatcggcatctgcctggt cagaatcacagccaacaacagagttaagcggctgctga gcatggaattccacaccagatctaaggaaatcaag**CG** **G**ctgctggaaacccggagaatgtaccggtgcacacgg	MRHQRLVFLVDGRNNILHPTR KLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWI GV**A**PGDTIGICLVRITANNRVK RLLSMEFHTRSKEIK**R**LLETRR MYRCTRRYFRRKRVKKKGRIP KHRPARFENRRRSDGWLTPSV RHFLESHKSVIRKIAECVPDVNI SLEYAKFDIQKMMNPDIEGEQY

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	agatactttagaagaaagcgggtgaagaagaaaggga gaattcccaagcacagacctgctagatttgagaacagaa ggcggagcgacggctggctgacccctagcgtgcgcca cttccttgagagccacaagtccgtgatccggaagatcgc cgagtgtgtgccagatgtgaacatcagcctggagtacgc caagttcgacatccaaaagatgatgaaccccgatatcga gggcgagcagtaccagagaggcaaactgttcggatatc agaat**CGG**cgcgcctacgtgctgcatagagataagta cacctgtcagctgtgcgccaagaaaaacgtgcccttcca gctgcaccacgtggtgcctagaagccagaacggcaag gaccaccccgataacctgatcaccctgtgtcctgagtgc cacgtgaaagtgcacacagacaccaagctgctgaaaaa actgcagcggaagattaagagcaagattgctgatttcac cagcaccagcagactgaatattatcatgagcttcctgaag cactggctggaaaacgagttcggcaaggatcgggtgtg gttctgcatcggagccgactcttatgtgatccggaaggaa ctgaacctggacaagacacactgcaacgacgcttactgc atcgccatcaacaagctggtgagatttgacaccgacctg gaacacgaaagcgtggtggataagcaagtgatccggtt cgagcagaaaagacagcacaaccggagaatgaaatct ctggagagaaacaggctgtactacgtggtcaaggacaa gaagcgcgtgcgggtggccaccaacagaaatggcaat cctgatgccttccacgaggtgtacgacacaaagatcctg cctctgctgaagcagggctatcggctgaaagtgaaccct ggcgttaagggctacaggaagaagttcagcgag**CG** **G**ggatttaaggaaggcgccacaatcagaatcggaaatc gggtggacgtgctgaagggcacctccggcttcggctac acctacgacggcaccatcttcaagctcaaggacatggac gtggtgtacccctacggcggctggaagtgcgtgaaggt g (SEQ ID NO: 170)	QRGKLFGYQN**R**RAYVLHRDK YTCQLCAKKNVPFQLHHVVPR SQNGKDHPDNLITLCPECHVKV HTDTKLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLENEFGK DRVWFCIGADSYVIRKELNLDK THCNDAYCIAINKLVRFDTDLE HESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVR VATNRNGNPDAFHEVYDTKILP LLKQGYRLKVNPGVKGYRKKF SE**R**GFKEGATIRIGNRVDVLKG TSGFGYTYDGTIFKLKDMDVV YPYGGWKCVKV (SEQ ID NO: 168) *With no N-terminus M residue (SEQ ID NO: 169)*

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 VEGFA with a 5'- RRT-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: 47. The sequence of the encoded RNA guide and its individual components are shown in **Table 12**.

Table 12. RNA Sequences

Component	RNA sequence	SEQ ID NO
Reference sgRNA scaffold	*See above*	2
VEGFA-T6 spacer	UGGGGACCAAGGAAGUGUCA	45
T7Te terminator	*ggcucaccuucgggugggccuuucugcg*	46
T7 transcribed VEGFA-T6 sgRNA	gUGGGGACCAAGGAAGUGUCAgggacaugaccagaugcg aaaguaacugucauguccucaugccaauggacuauggugaugcaccuag gggugucguuccagcuccuagcucugcgaaggagcauucguuaucggau gcguuuguacuucaagccauagcccaggccagggacguauuacugcucc ugcccuauaggagcaua*ggcucaccuucgggugggccuuucugcg*	47

A DNA target was designed and ordered as a synthesized linear DNA fragment. The target sequence from VEGFA and 10 bases upstream and downstream within the exon was flanked by 200 bases of unrelated sequence upstream and 100 bases of unrelated sequence downstream. The extra sequence was added so that the cleaved and uncleaved products would separate well on a gel. The target and non-target strands were labelled with 5' IR700 and 5' IR800 labels, respectively, through PCR amplification using labelled primers. The sequences of the DNA target, the individual components of the DNA target, and the labelled PCR primers are in **Table 13**.

Table 13. Target gBlock and Primer Sequences

Component	DNA sequence	SEQ ID NO
VEGFA-T6 target	See **Table 2**	22
VEGFA-T6 upstream 10 nt	*TTGCCCATAC*	48
VEGFA-T6 downstream 10 nt	*AGTGCATGAG*	49
200 nt upstream sequence	cagcaaaacccgtacttttggacgttccggcggttttttgtggcgagtggtgttcgg gcggtgcgcgcaagatccatctagattatgttaaacgggcgagtttacatgaattc cgagctccaaaaccgcccgcttaaggtagggaatgcgagtccatgtctcgttata cgctgtggttcgccaaccggtccgatactact	50
100 nt downstream sequence	actactgtccgataatcgaagatatgagacgaacacccaaatcaccctcagcgcg attaagctcatagtactacggtacctccagtctaaagctttgagc	51

Full target gblock	cagcaaaacccgtacttttggacgttccggcggttttttgtggcgagtggtgttcgg gcggtgcgcgcaagatccatctagattatgttaaacgggcgagtttacatgaattc cgagctccaaaaccgcccgcttaaggtagggaatgcgagtccatgtctcgttata cgctgtggttcgccaaccggtccgatactact*TTGCCCATAC*TGGGG ACCAAGGAAGTGTCA*AGTGCATGAG*actactgtccgataat cgaagatatgagacgaacacccaaatcaccctcagcgcgattaagctcatagtac tacggtacctccagtctaaagctttgagc	52
Forward primer	/5IRD800/CAGCAAAACCCGTACTTTTGGACG	53
Reverse primer	/5IRD700/GCTCAAAGCTTTAGACTGGAGGTACCG	54

Cleavage activity of the reference RNA-guided 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 VEGFA-T6 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 700 nm and 800 nm channels to visualize the 5' IR700 and 5' IR800 labels on the target and non-target strands of the target DNA substrate. Band intensities were quantified using ImageJ software.

Gel images are shown in **FIGS. 2A-2C**, and quantification of the percent of cleaved target and non-target strands are shown in **FIG. 2D**. The uncleaved, HNH-cleaved, and RuvC-cleaved strands are indicated. **FIG. 2A** is a gel image captured using the 700 nm channel showing cleavage of the target strand. **FIG. 2B** is a gel image captured using the 800 nm channel showing cleavage of the non-target strand. **FIG. 2C** is an overlay of the gel images from **FIG. 2A** and **FIG. 2B.** As shown in **FIGS. 2A-2D**, the reference RNA-guided 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 (H231A, H232A, H232L, and H255A) showed significantly decreased activity on the target strand while retaining activity on the non-target strand. Each of

the three RuvC-knockout nickase constructs (D67A, E176A, and D329A) 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 H232A variant was further chosen to install edits into human gene targets, as described in **Example 6**.

Example 6: Fusion of RNA-Guided Nuclease and RNA-Guided Nickase to a Reverse Transcriptase

In this Example, a reverse transcriptase polypeptide was fused to the C-terminus of the Nuclease BT of SEQ ID NO: 1 and the Q102R, I206R, V433R variant of SEQ ID NO: 1 (see **Example 3**). The H232A nickase substitution was also installed into the Q102R, I206R, V433R RNA-guided nuclease-reverse transcriptase polypeptide.

A sequence encoding the RNA-guided nuclease-reverse transcriptase fusion polypeptide was cloned into a pcDNA3.1 vector (Invitrogen) comprising a CMV promoter. The fusion comprised the following components arranged from N- to C-termini: 1) SV40 NLS, 2) Nuclease BT of SEQ ID NO: 1 or the RNA-guided nuclease variant comprising substitutions Q102R, I206R, and V433R relative to 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 RNA-guided nuclease of SEQ ID NO: 1, the XTEN linker, and nucleoplasmin NLS are shown in **Table 1** in **Example 1**, and the nucleotide and amino acid sequences of the variant MMLV are shown in **Table 14** below. The variant MMLV reverse transcriptase was a human codon-optimized DNA sequence.

Table 14. Reverse Transcriptase (RT) Sequences

Component	Nucleotide Sequence	Amino Acid Sequence
Variant MMLV-RT	atgaccctaaatatagaagatgagtatcggctacatgagacctca aaagagccagatgtttctctagggtccacatggctgtctgattttc ctcaggcctgggcggaaaccgggggcatgggactggcagttc gccaagctcctctgatcatacctctgaaagcaacctctacccccg tgtccataaaacaataccccatgtcacaagaagccagactgggg atcaagccccacatacagagactgttggaccagggaatactggt accctgccagtccccctggaacacgcccctgctacccgttaaga aaccagggactaatgattataggcctgtccaggatctgagagaa gtcaacaagcgggtggaagacatccaccccaccgtgcccaac ccttacaacctcttgagcgggctcccaccgtcccaccagtggta cactgtgcttgatttaaaggatgccttttttctgcctgagactccacc	MTLNIEDEYRLHETSKEP DVSLGSTWLSDFPQAWAE TGGMGLAVRQAPLIIPLKA TSTPVSIKQYPMSQEARLG IKPHIQRLLDQGILVPCQSP WNTPLLPVKKPGTNDYRP VQDLREVNKRVEDIHPTV PNPYNLLSGLPPSHQWYT VLDLKDAFFCLRLHPTSQ PLFAFEWRDPEMGISGQLT WTRLPQGFKNSPTLFNEA

	ccaccagtcagcctctcttcgcctttgagtggagagatccagaga tgggaatctcaggacaattgacctggaccagactcccacagggt ttcaaaaacagtcccaccctgtttaatgaggcactgcacagagac ctagcagacttccggatccagcacccagacttgatcctgctaca gtacgtggatgacttactgctggccgccacttctgagctagactg ccaacaaggtactcgggccctgttacaaaccctagggaacctcg ggtatcgggcctcggccaagaaagcccaaatttgccagaaaca ggtcaagtatctggggtatcttctaaaagagggtcagagatggct gactgaggccagaaaagagactgtgatggggcagcctactccg aagacccctcgacaactaagggagttcctagggaaggcaggct tctgtcgcctcttcatccctgggtttgcagaaatggcagccccct gtaccctctcaccaaaccggggactctgtttaattggggcccaga ccaacaaaaggcctatcaagaaatcaagcaagctcttctaactg ccccagccctggggttgccagatttgactaagccctttgaactctt tgtcgacgagaagcagggctacgccaaaggtgtcctaacgcaa aaactgggaccttggcgtcggccggtggcctacctgtccaaaaa gctagacccagtagcagctgggtggccccccttgcctacggatg gtagcagccattgccgtactgacaaaggatgcaggcaagctaa ccatgggacagccactagtcattctggcccccccatgcagtagag gcactagtcaaacaacccccccgaccgctggctttccaacgccc ggatgactcactatcaggccttgcttttggacacggaccgggtcc agttcggaccggtggtagccctgaacccggctacgctgctccca ctgcctgaggaagggctgcaacacaactgccttgatatcctggc cgaagcccacggaacccgacccgacctaacggaccagccgct cccagacgccgaccacacctggtacacggatggaagcagtctc ttacaagagggacagcgtaaggcgggagctgcggtgaccacc gagaccgaggtaatctgggctaaagccctgccagccgggacat ccgctcagcgggctgaactgatagcactcacccaggccctaaa gatggcagaaggtaagaagctaaatgtttatactgatagccgttat gcttttgctactgcccatatccatggagaaatatacagaaggcgt gggtggctcacatcagaaggcaaagagatcaaaaataaagacg agatcttggccctactaaaagccctctttctgcccaaaagacttag cataatccattgtccaggacatcaaaagggacacagcgccgag gctagaggcaaccggatggctgaccaagcggcccgaaaggc agccatcacagagactccagacacctctaccctcctcatagaaa attcatcaccc (SEQ ID NO: 171)	LHRDLADFRIQHPDLILLQ YVDDLLLAATSELDCQQG TRALLQTLGNLGYRASAK KAQICQKQVKYLGYLLKE GQRWLTEARKETVMGQP TPKTPRQLREFLGKAGFC RLFIPGFAEMAAPLYPLTK PGTLFNWGPDQQKAYQEI KQALLTAPALGLPDLTKPF ELFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDPV AAGWPPCLRMVAAIAVLT KDAGKLTMGQPLVILAPH AVEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFGP VVALNPATLLPLPEEGLQH NCLDILAEAHGTRPDLTD QPLPDADHTWYTDGSSLL QEGQRKAGAAVTTETEVI WAKALPAGTSAQRAELIA LTQALKMAEGKKLNVYT DSRYAFATAHIHGEIYRRR GWLTSEGKEIKNKDEILAL LKALFLPKRLSIIHCPGHQ KGHSAEARGNRMADQAA RKAAITETPDTSTLLIENSS P (SEQ ID NO: 56) *With no N-terminus M residue (SEQ ID NO: 57)*

An H232A nickase mutation (*see* **Example 5**) was installed into the Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase fusion polypeptide plasmid. The nucleotide and amino acid sequences of the RNA-guided nuclease-reverse transcriptase fusion polypeptides and the RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide are shown in **Table 15**. The codon encoding the substituted H232A 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.

Table 15. Fusion Polypeptides with C-Terminal Reverse Transcriptase (RT)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
Reference RNA-guided nuclease-RT fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa	*MKRTADGSEFESPKKKRK*
	agaagaagcggaaagtcagacaccagagactcgtgtttctc	*V*RHQRLVFLVDGRNNILH
	gtggacggtagaaacaacatcctgcatcctaccagaaagct	PTRKLDMVRRWLRQTNA
	ggatatggtcagaagatggctgagacaaacaaacgccaag	KVIRFIGKIMVVQVFKILS
	gtgatcagattcatcggaaaaatcatggtggtccaggtgttca	QQKDDPTYWIGVDPGDT
	agatcctgtctcagcagaaagacgaccctacatactggatcg	IGICLVRITANNRVKRLLS
	gcgtggacccaggcgacaccatcggcatctgcctggtcag	MEFHTRSKEIKQLLETRR
	aatcacagccaacaacagagttaagcggctgctgagcatgg	MYRCTRRYFRRKRVKKK
	aattccacaccagatctaaggaaatcaagcagctgctggaa	GRIPKHRPARFENRRRSD
	acccggagaatgtaccggtgcacacggagatactttagaag	GWLTPSVRHFLESHKSVI
	aaagcgggtgaagaagaaagggagaattcccaagcacag	RKIAECVPDVNISLEYAK
	acctgctagatttgagaacagaaggcggagcgacggctgg	FDIQKMMNPDIEGEQYQ
	ctgacccctagcgtgcgccacttccttgagagccacaagtcc	RGKLFGYQNIRAYVLHR
	gtgatccggaagatcgccgagtgtgtgccagatgtgaacat	DKYTCQLCAKKNVPFQL
	cagcctggagtacgccaagttcgacatccaaaagatgatga	HHVVPRSQNGKDHPDNL
	accccgatatcgagggcgagcagtaccagagaggcaaact	ITLCPECHVKVHTDTKLL
	gttcggatatcagaatatccgcgcctacgtgctgcatagaga	KKLQRKIKSKIADFTSTS
	taagtacacctgtcagctgtgcgccaagaaaaacgtgccctt	RLNIIMSFLKHWLENEFG
	ccagctgcaccacgtggtgcctagaagccagaacggcaag	KDRVWFCIGADSYVIRKE
	gaccaccccgataacctgatcaccctgtgtcctgagtgccac	LNLDKTHCNDAYCIAINK
	gtgaaagtgcacacagacaccaagctgctgaaaaaactgc	LVRFDTDLEHESVVDKQ
	agcggaagattaagagcaagattgctgatttcaccagcacc	VIRFEQKRQHNRRMKSL
	agcagactgaatattatcatgagcttcctgaagcactggctg	ERNRLYYVVKDKKRVRV
	gaaaacgagttcggcaaggatcgggtgtggttctgcatcgg	ATNRNGNPDAFHEVYDT
	agccgactcttatgtgatccggaaggaactgaacctggaca	KILPLLKQGYRLKVNPGV
	agacacactgcaacgacgcttactgcatcgccatcaacaag	KGYRKKFSEVGFKEGATI
	ctggtgagatttgacaccgacctggaacacgaaagcgtggt	RIGNRVDVLKGTSGFGYT
	ggataagcaagtgatccggttcgagcagaaaagacagcac	YDGTIFKLKDMDVVYPY
	aaccggagaatgaaatctctggagagaaacaggctgtacta	GGWKCVKV*SGGSSGGSS*
	cgtggtcaaggacaagaagcgcgtgcgggtggccaccaa	*GSETPGTSESATPESSGGSS*
	cagaaatggcaatcctgatgccttccacgaggtgtacgaca	*GGSS*TLNIEDEYRLHETS
	caaagatcctgcctctgctgaagcagggctatcggctgaaa	KEPDVSLGSTWLSDFPQA
	gtgaaccctggcgttaagggctacaggaagaagttcagcga	WAETGGMGLAVRQAPLII
	ggtgggatttaaggaaggcgccacaatcagaatcggaaatc	PLKATSTPVSIKQYPMSQ
	gggtggacgtgctgaagggcacctccggcttcggctacac	EARLGIKPHIQRLLDQGIL
	ctacgacggcaccatcttcaagctcaaggacatggacgtgg	VPCQSPWNTPLLPVKKP
	tgtacccctacggcggctggaagtgcgtgaaggtgtctgga	GTNDYRPVQDLREVNKR
	ggatctagcggaggatcctctggcagcgagacaccaggaa	VEDIHPTVPNPYNLLSGL
	caagcgagtcagcaacaccagagagcagtggcggcagca	PPSHQWYTVLDLKDAFF
	gcggcggcagcagcaccctaaatatagaagatgagtatcg	CLRLHPTSQPLFAFEWRD
	gctacatgagacctcaaaagagccagatgtttctctagggtc	PEMGISGQLTWTRLPQGF
	cacatggctgtctgattttcctcaggcctgggcggaaaccgg	KNSPTLFNEALHRDLADF
	gggcatgggactggcagttcgccaagctcctctgatcatacc	RIQHPDLILLQYVDDLLL
	tctgaaagcaacctctacccccgtgtccataaaacaataccc	AATSELDCQQGTRALLQ
	catgtcacaagaagccagactggggatcaagccccacata	TLGNLGYRASAKKAQIC
	cagagactgttggaccagggaatactggtaccctgccagtc	QKQVKYLGYLLKEGQR
	cccctggaacacgcccctgctacccgttaagaaaccaggg	WLTEARKETVMGQPTPK

	actaatgattataggcctgtccaggatctgagagaagtcaac aagcgggtggaagacatccaccccaccgtgcccaaccctt acaacctcttgagcgggctcccaccgtcccaccagtggtac actgtgcttgatttaaaggatgccttttttctgcctgagactccac cccaccagtcagcctctcttcgcctttgagtggagagatcca gagatgggaatctcaggacaattgacctggaccagactccc acagggtttcaaaaacagtcccaccctgtttaatgaggcact gcacagagacctagcagacttccggatccagcacccagac ttgatcctgctacagtacgtggatgacttactgctggccgcca cttctgagctagactgccaacaaggtactcgggccctgttac aaaccctagggaacctcgggtatcgggcctcggccaagaa agcccaaatttgccagaaacaggtcaagtatctggggtatctt ctaaaagagggtcagagatggctgactgaggccagaaaag agactgtgatggggcagcctactccgaagacccctcgacaa ctaagggagttcctagggaaggcaggcttctgtcgcctcttc atccctgggtttgcagaaatggcagcccccctgtaccctctc accaaaccggggactctgtttaattggggcccagaccaaca aaaggcctatcaagaaatcaagcaagctcttctaactgcccc agccctggggttgccagatttgactaagccctttgaactctttg tcgacgagaagcagggctacgccaaaggtgtcctaacgca aaaactgggaccttggcgtcggccggtggcctacctgtcca aaaagctagacccagtagcagctgggtggccccttgccta cggatggtagcagccattgccgtactgacaaaggatgcagg caagctaaccatgggacagccactagtcattctggccccc atgcagtagaggcactagtcaaacaacccccccgaccgctg gctttccaacgcccggatgactcactatcaggccttgcttttg gacacggaccgggtccagttcggaccggtggtagccctga acccggctacgctgctcccactgcctgaggaagggctgca acacaactgccttgatatcctggccgaagcccacggaaccc gacccgacctaacggaccagccgctcccagacgccgacc acacctggtacacggatggaagcagtctcttacaagaggga cagcgtaaggcgggagctgcggtgaccaccgagaccgag gtaatctgggctaaagccctgccagccgggacatccgctca gcgggctgaactgatagcactcacccaggccctaaagatg gcagaaggtaagaagctaaatgtttatactgatagccgttatg cttttgctactgcccatatccatggagaaatatacagaaggcg tgggtggctcacatcagaaggcaaagagatcaaaaataaag acgagatcttggccctactaaaagccctctttctgcccaaaag acttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggc ccgaaaggcagccatcacagagactccagacacctctacc ctcctcatagaaaattcatcacccaaaaggccggcggccac gaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 172)	TPRQLREFLGKAGFCRLF IPGFAEMAAPLYPLTKPG TLFNWGPDQQKAYQEIK QALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDP VAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLS NARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRP DLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVT TETEVIWAKALPAGTSAQ RAELIALTQALKMAEGK KLNVYTDSRYAFATAHIH GEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARG NRMADQAARKAAITETP DTSTLLIENSSP*KRPAATK* *KAGQAKKKK* (SEQ ID NO: 173)
Q102R, I206R, V433R variant RNA-guided nuclease-RT fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcagacaccagagactcgtgtttctc gtggacggtagaaacaacatcctgcatcctaccagaaagct ggatatggtcagaagatggctgagacaaacaaacgccaag gtgatcagattcatcggaaaaatcatggtggtccaggtgttca agatcctgtctcagcagaaagacgaccctacatactggatcg	*MKRTADGSEFESPKKKRK* *V*RHQRLVFLVDGRNNILH PTRKLDMVRRWLRQTNA KVIRFIGKIMVVQVFKILS QQKDDPTYWIGVDPGDT IGICLVRITANNRVKRLLS

	gcgtggacccaggcgacaccatcggcatctgcctggtcag	MEFHTRSKEIKRLLETRR
	aatcacagccaacaacagagttaagcggctgctgagcatgg	MYRCTRRYFRRKRVKKK
	aattccacaccagatctaaggaaatcaagcggctgctggaa	GRIPKHRPARFENRRRSD
	acccggagaatgtaccggtgcacacggagatactttagaag	GWLTPSVRHFLESHKSVI
	aaagcgggtgaagaagaaagggagaattcccaagcacag	RKIAECVPDVNISLEYAK
	acctgctagatttgagaacagaaggcggagcgacggctgg	FDIQKMMNPDIEGEQYQ
	ctgacccctagcgtgcgccacttccttgagagccacaagtcc	RGKLFGYQNRRAYVLHR
	gtgatccggaagatcgccgagtgtgtgccagatgtgaacat	DKYTCQLCAKKNVPFQL
	cagcctggagtacgccaagttcgacatccaaaagatgatga	HHVVPRSQNGKDHPDNL
	accccgatatcgagggcgagcagtaccagagaggcaaact	ITLCPECHVKVHTDTKLL
	gttcggatatcagaatcggcgcgcctacgtgctgcatagag	KKLQRKIKSKIADFTSTS
	ataagtacacctgtcagctgtgcgccaagaaaaacgtgccct	RLNIIMSFLKHWLENEFG
	tccagctgcaccacgtggtgcctagaagccagaacggcaa	KDRVWFCIGADSYVIRKE
	ggaccaccccgataacctgatcaccctgtgtcctgagtgcca	LNLDKTHCNDAYCIAINK
	cgtgaaagtgcacacagacaccaagctgctgaaaaaactg	LVRFDTDLEHESVVDKQ
	cagcggaagattaagagcaagattgctgatttcaccagcac	VIRFEQKRQHNRRMKSL
	cagcagactgaatattatcatgagcttcctgaagcactggctg	ERNRLYYVVKDKKRVRV
	gaaaacgagttcggcaaggatcgggtgtggttctgcatcgg	ATNRNGNPDAFHEVYDT
	agccgactcttatgtgatccggaaggaactgaacctggaca	KILPLLKQGYRLKVNPGV
	agacacactgcaacgacgcttactgcatcgccatcaacaag	KGYRKKFSERGFKEGATI
	ctggtgagatttgacaccgacctggaacacgaaagcgtggt	RIGNRVDVLKGTSGFGYT
	ggataagcaagtgatccggttcgagcagaaaagacagcac	YDGTIFKLKDMDVVYPY
	aaccggagaatgaaatctctggagagaaacaggctgtacta	GGWKCVKV*SGGSSGGSS*
	cgtggtcaaggacaagaagcgcgtgcgggtggccaccaa	*GSETPGTSESATPESSGGSS*
	cagaaatggcaatcctgatgccttccacgaggtgtacgaca	*GGSS*TLNIEDEYRLHETS
	caaagatcctgcctctgctgaagcagggctatcggctgaaa	KEPDVSLGSTWLSDFPQA
	gtgaaccctggcgttaagggctacaggaagaagttcagcga	WAETGGMGLAVRQAPLII
	gcggggatttaaggaaggcgccacaatcagaatcggaaat	PLKATSTPVSIKQYPMSQ
	cgggtggacgtgctgaagggcacctccggcttcggctaca	EARLGIKPHIQRLLDQGIL
	cctacgacggcaccatcttcaagctcaaggacatggacgtg	VPCQSPWNTPLLPVKKP
	gtgtacccctacggcggctggaagtgcgtgaaggtgtctgg	GTNDYRPVQDLREVNKR
	aggatctagcggaggatcctctggcagcgagacaccagga	VEDIHPTVPNPYNLLSGL
	acaagcgagtcagcaacaccagagagcagtggcggcagc	PPSHQWYTVLDLKDAFF
	agcggcggcagcagcaccctaaatatagaagatgagtatc	CLRLHPTSQPLFAFEWRD
	ggctacatgagacctcaaaagagccagatgtttctctagggt	PEMGISGQLTWTRLPQGF
	ccacatggctgtctgattttcctcaggcctgggcggaaaccg	KNSPTLFNEALHRDLADF
	ggggcatgggactggcagttcgccaagctcctctgatcata	RIQHPDLILLQYVDDLLL
	cctctgaaagcaacctctacccccgtgtccataaaacaatac	AATSELDCQQGTRALLQ
	cccatgtcacaagaagccagactggggatcaagccccaca	TLGNLGYRASAKKAQIC
	tacagagactgttggaccagggaatactggtaccctgccagt	QKQVKYLGYLLKEGQR
	ccccctggaacacgcccctgctacccgttaagaaaccagg	WLTEARKETVMGQPTPK
	gactaatgattataggcctgtccaggatctgagagaagtcaa	TPRQLREFLGKAGFCRLF
	caagcgggtggaagacatccaccccaccgtgcccaaccct	IPGFAEMAAPLYPLTKPG
	tacaacctcttgagcgggctcccaccgtcccaccagtggtac	TLFNWGPDQQKAYQEIK
	actgtgcttgatttaaaggatgccttttttctgcctgagactccac	QALLTAPALGLPDLTKPFE
	cccaccagtcagcctctcttcgcctttgagtggagagatcca	LFVDEKQGYAKGVLTQK
	gagatgggaatctcaggacaattgacctggaccagactccc	LGPWRRPVAYLSKKLDP
	acagggtttcaaaaacagtcccaccctgtttaatgaggcact	VAAGWPPCLRMVAAIAV
	gcacagagacctagcagacttccggatccagcacccagac	LTKDAGKLTMGQPLVILA
	ttgatcctgctacagtacgtggatgacttactgctggccgcca	PHAVEALVKQPPDRWLS

	cttctgagctagactgccaacaaggtactcgggccctgttac aaaccctagggaacctcgggtatcgggcctcggccaagaa agcccaaatttgccagaaacaggtcaagtatctggggtatctt ctaaaagagggtcagagatggctgactgaggccagaaaag agactgtgatggggcagcctactccgaagacccctcgacaa ctaagggagttcctagggaaggcaggcttctgtcgcctcttc atccctgggtttgcagaaatggcagcccccctgtacctctc accaaaccggggactctgtttaattggggcccagaccaaca aaaggcctatcaagaaatcaagcaagctcttctaactgcccc agccctggggttgccagatttgactaagccctttgaactctttg tcgacgagaagcagggctacgccaaaggtgtcctaacgca aaaactgggaccttggcgtcggccggtggcctacctgtcca aaaagctagacccagtagcagctgggtggcccccttgccta cggatggtagcagccattgccgtactgacaaaggatgcagg caagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctg gctttccaacgcccggatgactcactatcaggccttgcttttg gacacggaccgggtccagttcggaccggtggtagccctga acccggctacgctgctcccactgcctgaggaagggctgca acacaactgccttgatatcctggccgaagcccacggaaccc gacccgacctaacggaccagccgctcccagacgccgacc acacctggtacacggatggaagcagtctcttacaagaggga cagcgtaaggcgggagctgcggtgaccaccgagaccgag gtaatctgggctaaagccctgccagccgggacatccgctca gcgggctgaactgatagcactcacccaggccctaaagatg gcagaaggtaagaagctaaatgtttatactgatagccgttatg cttttgctactgcccatatccatggagaaatatacagaaggcg tgggtggctcacatcagaaggcaaagagatcaaaaataaag acgagatcttggccctactaaaagccctctttctgcccaaaag acttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggc ccgaaaggcagccatcacagagactccagacacctctacc ctcctcatagaaaattcatcacccaaaaggccggcggccac gaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 174)	NARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRP DLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVT TETEVIWAKALPAGTSAQ RAELIALTQALKMAEGK KLNVYTDSRYAFATAHIH GEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARG NRMADQAARKAAITETP DTSTLLIENSSP*KRPAATK KAGQAKKKK* (SEQ ID NO: 175)
Q102R, I206R, V433R variant RNA-guided nickase (H232A)-RT fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcagacaccagagactcgtgtttctc gtggacggtagaaacaacatcctgcatcctaccagaaagct ggatatggtcagaagatggctgagacaaacaaacgccaag gtgatcagattcatcggaaaaatcatggtggtccaggtgttca agatcctgtctcagcagaaagacgaccctacatactggatcg gcgtggacccaggcgacaccatcggcatctgcctggtcag aatcacagccaacaacagagttaagcggctgctgagcatgg aattccacaccagatctaaggaaatcaagcggctgctggaa acccggagaatgtaccggtgcacacggagatactttagaag aaagcgggtgaagaagaaagggagaattcccaagcacag acctgctagatttgagaacagaaggcggagcgacggctgg ctgacccctagcgtgcgccacttccttgagagccacaagtcc gtgatccggaagatcgccgagtgtgtgccagatgtgaacat cagcctggagtacgccaagttcgacatccaaaagatgatga	*MKRTADGSEFESPKKKRK VRHQRLVFLVDGRNNILH PTRKLDMVRRWLRQTNA KVIRFIGKIMVVQVFKILS QQKDDPTYWIGVDPGDT IGICLVRITANNRVKRLLS MEFHTRSKEIKRLLETRR MYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSD GWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAK FDIQKMMNPDIEGEQYQ RGKLFGYQNRRAYVLHR DKYTCQLCAKKNVPFQL HAVVPRSQNGKDHPDNL

	accccgatatcgagggcgagcagtaccagagaggcaaact	ITLCPECHVKVHTDTKLL
	gttcggatatcagaatcggcgcgcctacgtgctgcatagag	KKLQRKIKSKIADFTSTS
	ataagtacacctgtcagctgtgcgccaagaaaaacgtgccct	RLNIIMSFLKHWLENEFG
	tccagctgcac**GCC**gtggtgcctagaagccagaacggc	KDRVWFCIGADSYVIRKE
	aaggaccaccccgataacctgatcaccctgtgtcctgagtgc	LNLDKTHCNDAYCIAINK
	cacgtgaaagtgcacacagacaccaagctgctgaaaaaact	LVRFDTDLEHESVVDKQ
	gcagcggaagattaagagcaagattgctgatttcaccagca	VIRFEQKRQHNRRMKSL
	ccagcagactgaatattatcatgagcttcctgaagcactggct	ERNRLYYVVKDKKRVRV
	ggaaaacgagttcggcaaggatcgggtgtggttctgcatcg	ATNRNGNPDAFHEVYDT
	gagccgactcttatgtgatccggaaggaactgaacctggac	KILPLLKQGYRLKVNPGV
	aagacacactgcaacgacgcttactgcatcgccatcaacaa	KGYRKKFSERGFKEGATI
	gctggtgagatttgacaccgacctggaacacgaaagcgtgg	RIGNRVDVLKGTSGFGYT
	tggataagcaagtgatccggttcgagcagaaaagacagca	YDGTIFKLKDMDVVYPY
	caaccggagaatgaaatctctggagagaaacaggctgtact	GGWKCVKV*SGGSSGGSS*
	acgtggtcaaggacaagaagcgcgtgcgggtggccacca	*GSETPGTSESATPESSGGSS*
	acagaaatggcaatcctgatgccttccacgaggtgtacgac	*GGSS*TLNIEDEYRLHETS
	acaaagatcctgcctctgctgaagcagggctatcggctgaa	KEPDVSLGSTWLSDFPQA
	agtgaaccctggcgttaagggctacaggaagaagttcagcg	WAETGGMGLAVRQAPLII
	agcggggatttaaggaaggcgccacaatcagaatcggaaa	PLKATSTPVSIKQYPMSQ
	tcgggtggacgtgctgaagggcacctccggcttcggctaca	EARLGIKPHIQRLLDQGIL
	cctacgacggcaccatcttcaagctcaaggacatggacgtg	VPCQSPWNTPLLPVKKP
	gtgtacccctacggcggctggaagtgcgtgaaggtgtctgg	GTNDYRPVQDLREVNKR
	aggatctagcggaggatcctctggcagcgagacaccagga	VEDIHPTVPNPYNLLSGL
	acaagcgagtcagcaacaccagagagcagtggcggcagc	PPSHQWYTVLDLKDAFF
	agcggcggcagcagcaccctaaatatagaagatgagtatc	CLRLHPTSQPLFAFEWRD
	ggctacatgagacctcaaaagagccagatgtttctctagggt	PEMGISGQLTWTRLPQGF
	ccacatggctgtctgattttcctcaggcctgggcggaaaccg	KNSPTLFNEALHRDLADF
	ggggcatgggactggcagttcgccaagctcctctgatcata	RIQHPDLILLQYVDDLLL
	cctctgaaagcaacctctacccccgtgtccataaaacaatac	AATSELDCQQGTRALLQ
	cccatgtcacaagaagccagactggggatcaagccccaca	TLGNLGYRASAKKAQIC
	tacagagactgttggaccagggaatactggtaccctgccagt	QKQVKYLGYLLKEGQR
	ccccctggaacacgcccctgctacccgttaagaaaccagg	WLTEARKETVMGQPTPK
	gactaatgattataggcctgtccaggatctgagagaagtcaa	TPRQLREFLGKAGFCRLF
	caagcgggtggaagacatccaccccaccgtgcccaaccct	IPGFAEMAAPLYPLTKPG
	tacaacctcttgagcgggctcccaccgtcccaccagtggtac	TLFNWGPDQQKAYQEIK
	actgtgcttgatttaaaggatgccttttttctgcctgagactccac	QALLTAPALGLPDLTKPFE
	cccaccagtcagcctctcttcgcctttgagtggagagatcca	LFVDEKQGYAKGVLTQK
	gagatgggaatctcaggacaattgacctggaccagactccc	LGPWRRPVAYLSKKLDP
	acagggtttcaaaaacagtcccaccctgtttaatgaggcact	VAAGWPPCLRMVAAIAV
	gcacagagacctagcagacttccggatccagcacccagac	LTKDAGKLTMGQPLVILA
	ttgatcctgctacagtacgtggatgacttactgctggccgcca	PHAVEALVKQPPDRWLS
	cttctgagctagactgccaacaaggtactcgggccctgttac	NARMTHYQALLLDTDRV
	aaaccctagggaacctcgggtatcgggcctcggccaagaa	QFGPVVALNPATLLPLPEE
	agcccaaatttgccagaaacaggtcaagtatctggggtatctt	GLQHNCLDILAEAHGTRP
	ctaaaagagggtcagagatggctgactgaggccagaaaag	DLTDQPLPDADHTWYTD
	agactgtgatggggcagcctactccgaagacccctcgacaa	GSSLLQEGQRKAGAAVT
	ctaagggagttcctagggaaggcaggcttctgtcgcctcttc	TETEVIWAKALPAGTSAQ
	atccctgggtttgcagaaatggcagcccccctgtaccctctc	RAELIALTQALKMAEGK
	accaaaccggggactctgtttaattggggcccagaccaaca	KLNVYTDSRYAFATAHIH
	aaaggcctatcaagaaatcaagcaagctcttctaactgcccc	GEIYRRRGWLTSEGKEIK

	agccctggggttgccagatttgactaagccctttgaactctttg tcgacgagaagcagggctacgccaaaggtgtcctaacgca aaaactgggaccttggcgtcggccggtggcctacctgtcca aaaagctagacccagtagcagctgggtggcccccttgccta cggatggtagcagccattgccgtactgacaaaggatgcagg caagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaaccccccgaccgctg gctttccaacgcccggatgactcactatcaggccttgcttttg gacacggaccgggtccagttcggaccggtggtagccctga acccggctacgctgctcccactgcctgaggaagggctgca acacaactgccttgatatcctggccgaagcccacggaaccc gacccgacctaacggaccagccgctcccagacgccgacc acacctggtacacggatggaagcagtctcttacaagaggga cagcgtaaggcgggagctgcggtgaccaccgagaccgag gtaatctgggctaaagccctgccagccgggacatccgctca gcgggctgaactgatagcactcacccaggccctaaagatg gcagaaggtaagaagctaaatgtttatactgatagccgttatg cttttgctactgcccatatccatggagaaatatacagaaggcg tgggtggctcacatcagaaggcaaagagatcaaaaataaag acgagatcttggccctactaaaagccctctttctgcccaaaag acttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggc ccgaaaggcagccatcacagagactccagacacctctacc ctcctcatagaaaattcatcacccaaaaggccggcggccac gaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 176)	NKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARG NRMADQAARKAAITETP DTSTLLIENSSP*KRPAATK KAGQAKKKK* (SEQ ID NO: 177)

This Example describes how the reference RNA-guided nuclease-reverse transcriptase fusion polypeptide, the Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase fusion polypeptide, the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide, and the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase (N-terminal RT) fusion polypeptide constructs were cloned. The constructs of SEQ ID NOs: DD and FF were used in **Example 7** to install edits in human target genes. The construct of SEQ ID NO: BB is used in **Example 8**.

Similar to the C-terminal RT fusions described above, coding sequences for fusion polypeptides with N-terminal RT were also cloned into a pcDNA3.1 vector (Invitrogen) comprising a CMV promoter. The N-terminal RT fusion polypeptides comprise the following components arranged from N- to C-termini: 1) an SV40 NLS, 2) a variant Moloney Murine Leukemia Virus (MMLV) reverse transcriptase, 3) XTEN linker 4) Nuclease BT variant comprising substitutions Q102R, I206R, V433R, and H232A (nickase substitution) relative to SEQ ID NO: 1, and 5) the nucleoplasmin NLS.

Amino acid sequences for additional exemplary N-terminal RT fusion polypeptides are

provided in **Table 22** below.

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

This Example shows genetic modification of human genes utilizing the Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase and the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptides 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 2** 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 1 of SEQ ID NO: 25 3) reverse transcription template (RTT) encoding 6 nucleotide substitutions, 4) primer binding site (PBS), and 5) a 3' extension motif. For all editing templates tested, the spacer was 16-nucleotides in length, the RTT was 20-nucleotides in length and the PBS was 11-nucleotides in length. 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. The sequences of each component and the full-length editing template RNA sequences are shown in **Tables 16** and **17**, respectively. 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 16. Sequences of Editing Template RNA Components

Editing template RNA	Target	Spacer	Scaffold	RTT*	PBS	3' Extension
1	AAVS1-T6	GGGAGAGGCUACAAAG (SEQ ID NO: 178)	GGGACAUGACCAGGAAACUGUCAUGUCCUCAUGCCAAUGGACUAUGGUGAUGCACCUAGGGGUGUCGUUCCAGCUCCUAGCUCUGCGAAGGAGCAUUCGGAAACGGAUGCGU	UUCUCCUUGACuaagaaUGU (SEQ ID NO: 180)	AGCCUCUCCCG (SEQ ID NO: 181)	UCUCUCUCUUGACGCGGUUCUAUCUAGUUACGCGUUAAACCAACUAGAAA

			UUGUACUUCAAGCC AUAGCCCAGGCCAG GGACGUAUUACUGC UCCUGCCCUAUAGG AGC (SEQ ID NO: 179)			(SEQ ID NO: 182)
2	EMX1-T7	CAAGGG ACUAUU CAGG (SEQ ID NO: 183)	GGGACAUGACCAGG AAACUGUCAUGUCC UCAUGCCAAUGGAC UAUGGUGAUGCACC UAGGGGUGUCGUUC CAGCUCCUAGCUCU GCGAAGGAGCAUUC GGAAACGGAUGCGU UUGUACUUCAAGCC AUAGCCCAGGCCAG GGACGUAUUACUGC UCCUGCCCUAUAGG AGC (SEQ ID NO: 184)	AUCCCUGCU UCuagggaGA A (SEQ ID NO: 185)	UAGUCCC UUGC (SEQ ID NO: 186)	UCUCUCUC UUGACGCG GUUCUAUC UAGUUACG CGUUAAAC CAACUAGA AA (SEQ ID NO: 187)
3	VEGFA-T6	GACCAA GGAAGU GUCA (SEQ ID NO: 188)	GGGACAUGACCAGG AAACUGUCAUGUCC UCAUGCCAAUGGAC UAUGGUGAUGCACC UAGGGGUGUCGUUC CAGCUCCUAGCUCU GCGAAGGAGCAUUC GGAAACGGAUGCGU UUGUACUUCAAGCC AUAGCCCAGGCCAG GGACGUAUUACUGC UCCUGCCCUAUAGG AGC (SEQ ID NO: 189)	UUCACUCAU GCugaacuCAC (SEQ ID NO: 190)	UUCCUUG GUCC (SEQ ID NO: 191)	UCUCUCUC UUGACGCG GUUCUAUC UAGUUACG CGUUAAAC CAACUAGA AA (SEQ ID NO: 192)

**6-nucleotide substitution in lowercase*

Table 17. Full-length Editing Template RNA Sequences

Editing template RNA	Target	Sequence
1	AAVS1-T6	GGGAGAGGCUACAAAG**GGGACAUGACCAGGAAACUGUCAUG UCCUCAUGCCAAUGGACUAUGGUGAUGCACCUAGGGGUGUC GUUCCAGCUCCUAGCUCUGCGAAGGAGCAUUCGGAAACGGA UGCGUUUGUACUUCAAGCCAUAGCCCAGGCCAGGGACGUAU UACUGCUCCUGCCCUAUAGGAGC**CUUCUCCUUGACuaagaaUGUA GCCUCUCCCGUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGC GUUAAACCAACUAGAAA (SEQ ID NO: 193)

2	EMX1-T7	CAAGGGACUAUUCAGG**GGGACAUGACCAGGAAACUGUCAUG UCCUCAUGCCAAUGGACUAUGGUGAUGCACCUAGGGGUGUC GUUCCAGCUCCUAGCUCUGCGAAGGAGCAUUCGGAAACGGA UGCGUUUGUACUUCAAGCCAUAGCCCAGGCCAGGGACGUAU UACUGCUCCUGCCCUAUAGGAGC**AUCCCUGCUUCuagggaGAA UAGUCCCUUGCUCUCUCUCUUGACGCGGUUCUAUCUAGUUAC GCGUUAAACCAACUAGAAA (SEQ ID NO: 194)
3	VEGFA-T6	GACCAAGGAAGUGUCA**GGGACAUGACCAGGAAACUGUCAUG UCCUCAUGCCAAUGGACUAUGGUGAUGCACCUAGGGGUGUC GUUCCAGCUCCUAGCUCUGCGAAGGAGCAUUCGGAAACGGA UGCGUUUGUACUUCAAGCCAUAGCCCAGGCCAGGGACGUAU UACUGCUCCUGCCCUAUAGGAGC**UUCACUCAUGCugaacuCACU UCCUUGGUCCUCUCUCUCUUGACGCGGUUCUAUCUAGUUACGC GUUAAACCAACUAGAAA (SEQ ID NO: 195)

** The scaffold sequence is shown in bold and the 6-nucleotide substitution 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**.

Samples were prepared for NGS and analyzed as described in **Example 1**. For each target, the percentage of NGS reads containing indels were calculated for each sample and its cognate no-guide 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 18** and **Table 19**. The data presented therein are the average of three technical replicates unless indicated otherwise.

Table 18. Editing Efficiencies with the Q102R/I206R/V433R Variant RNA-guided Nuclease-Reverse Transcriptase Fusion Polypeptide

Editing Template RNAs	Target	Average % Indel	Average % 6-nt substitution installation
1	AAVS1-T6	0.77	0.04
2	EMX1-T7	2.27*	0.07*
3	VEGFA-T6	2.61	0.07
No-guide control	AAVS1-T6	0.02	0.00
No-guide control	EMX1-T7	0.27	0.00
No-guide control	VEGFA-T6	0.04	0.00

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

Table 19. Editing Efficiencies with the Q102R/I206R/V433R Variant RNA-guided Nickase (H232A)-Reverse Transcriptase Fusion Polypeptide

Editing Template RNAs	Target	Average % Indel	Average % 6-nt substitution installation
1	AAVS1-T6	0.04	0.02
2	EMX1-T7	0.30	0.04
3	VEGFA-T6	0.04	0.02
No-guide control	AAVS1-T6	0.04	0.00
No-guide control	EMX1-T7	0.34	0.00
No-guide control	VEGFA-T6	0.04	0.00

As shown in **Table 18** and **Table 19**, the Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase fusion polypeptide and the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide, respectively, introduced substitutions encoded by the tested editing template RNAs at the AAVS1, EMX1, and VEGFA target loci. For the Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 0.77% to 2.61%, while the average percentage of NGS reads comprising encoding edits ranged from 0.04% to 0.07% (**Table 18**). Editing templates RNAs 2 and 3 displayed the highest encoded edit installations. For the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide, the average percentage of NGS reads comprising indels ranged from 0.04% to 0.30%, while the average percentage of NGS reads comprising encoding edits ranged from 0.02% to 0.04% (**Table 19**). For the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide, editing template RNA 2 displayed the greatest encoded edit installation.

The background indel rate from PCR amplification artifacts and/or sequencing errors was determined by testing controls where the editing template guide was excluded. The indel levels observed for the Q102R, I206R, V433R variant RNA-guide nickase (H232A)-reverse transcriptase fusion polypeptide were within error of these background indel rates (**Table 18** and **Table 19**). Additionally, no reads were observed with the 6-nucleotide substitution installed for either control polypeptide in the absence of an editing template. This confirmed that the H232A substitution converted the Q102R, I206R, V433R variant RNA-guided nuclease to an RNA-guided nickase 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 Q102R, I206R, V433R variant RNA-guided nuclease-reverse transcriptase fusion polypeptide and the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide incorporated substitutions encoded by editing template RNAs into human genes.

Example 8: Engineering an RNA-Guided Nickase-Reverse Transcriptase Fusion Polypeptide

In this Example, an H232A nickase mutation (*see* **Example 5**) is installed into the reference RNA-guided nuclease-reverse transcriptase fusion polypeptide of SEQ ID NO: 173 as described in **Example 6**.

The nucleotide and amino acid sequences of the reference RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide is shown in **Table 20**. The codon encoding the substituted H232A 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. This fusion polypeptide is used in **Example 9** to install edits in human target genes.

Table 20. Exemplary Fusion Polypeptides

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
Reference RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcagacaccagagactcgtgtttctc gtggacggtagaaacaacatcctgcatcctaccagaaagct ggatatggtcagaagatggctgagacaaacaaacgccaag gtgatcagattcatcggaaaaatcatggtggtccaggtgttca agatcctgtctcagcagaaagacgaccctacatactggatcg gcgtggacccaggcgacaccatcggcatctgcctggtcag aatcacagccaacaacagagttaagcggctgctgagcatgg aattccacaccagatctaaggaaatcaagcagctgctggaa acccggagaatgtaccggtgcacacggagatactttagaag aaagcgggtgaagaagaaagggagaattcccaagcacag acctgctagatttgagaacagaaggcggagcgacggctgg ctgacccctagcgtgcgccacttccttgagagccacaagtcc gtgatccggaagatcgccgagtgtgtgccagatgtgaacat cagcctggagtacgccaagttcgacatccaaaagatgatga accccgatatcgagggcgagcagtaccagagaggcaaact gttcggatatcagaatatccgcgcctacgtgctgcatagaga taagtacacctgtcagctgtgcgccaagaaaaacgtgccctt ccagctgcac**GCC**gtggtgcctagaagccagaacggca aggaccaccccgataacctgatcaccctgtgtcctgagtgcc acgtgaaagtgcacacagacaccaagctgctgaaaaaact gcagcggaagattaagagcaagattgctgatttcaccagca ccagcagactgaatattatcatgagcttcctgaagcactggct	*MKRTADGSEFESPKKKRK* *V*RHQRLVFLVDGRNNILH PTRKLDMVRRWLRQTNA KVIRFIGKIMVVQVFKILS QQKDDPTYWIGVDPGDT IGICLVRITANNRVKRLLS MEFHTRSKEIKQLLETRR MYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSD GWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAK FDIQKMMNPDIEGEQYQ RGKLFGYQNIRAYVLHR DKYTCQLCAKKNVPFQL H**A**VVPRSQNGKDHPDNL ITLCPECHVKVHTDTKLL KKLQRKIKSKIADFTSTS RLNIIMSFLKHWLENEFG KDRVWFCIGADSYVIRKE LNLDKTHCNDAYCIAINK LVRFDTDLEHESVVDKQ VIRFEQKRQHNRRMKSL ERNRLYYVVKDKKRVRV

	ggaaaacgagttcggcaaggatcgggtgtggttctgcatcg gagccgactcttatgtgatccggaaggaactgaacctggac aagacacactgcaacgacgcttactgcatcgccatcaacaa gctggtgagatttgacaccgacctggaacacgaaagcgtgg tggataagcaagtgatccggttcgagcagaaaagacagca caaccggagaatgaaatctctggagagaaacaggctgtact acgtggtcaaggacaagaagcgcgtgcgggtggccacca acagaaatggcaatcctgatgccttccacgaggtgtacgac acaaagatcctgcctctgctgaagcagggctatcggctgaa agtgaaccctggcgttaagggctacaggaagaagttcagcg aggtgggatttaaggaaggcgccacaatcagaatcggaaat cgggtggacgtgctgaagggcacctccggcttcggctaca cctacgacggcaccatcttcaagctcaaggacatggacgtg gtgtacccctacggcggctggaagtgcgtgaaggtgtctgg aggatctagcggaggatcctctggcagcgagacaccagga acaagcgagtcagcaacaccagagagcagtggcggcagc agcggcggcagcagcaccctaaatatagaagatgagtatc ggctacatgagacctcaaaagagccagatgtttctctagggt ccacatggctgtctgattttcctcaggcctgggcggaaaccg ggggcatgggactggcagttcgccaagctcctctgatcata cctctgaaagcaacctctacccccgtgtccataaaacaatac cccatgtcacaagaagccagactggggatcaagccccaca tacagagactgttggaccagggaatactggtaccctgccagt ccccctggaacacgcccctgctacccgttaagaaaccagg gactaatgattataggcctgtccaggatctgagagaagtcaa caagcgggtggaagacatccaccccaccgtgcccaaccct tacaacctcttgagcgggctcccaccgtcccaccagtggtac actgtgcttgatttaaaggatgccttttttctgcctgagactccac cccaccagtcagcctctcttcgcctttgagtggagagatcca gagatgggaatctcaggacaattgacctggaccagactccc acagggtttcaaaaacagtcccaccctgtttaatgaggcact gcacagagacctagcagacttccggatccagcacccagac ttgatcctgctacagtacgtggatgacttactgctggccgcca cttctgagctagactgccaacaaggtactcgggccctgttac aaaccctagggaacctcgggtatcgggcctcggccaagaa agcccaaatttgccagaaacaggtcaagtatctggggtatctt ctaaaagagggtcagagatggctgactgaggccagaaaag agactgtgatggggcagcctactccgaagacccctcgacaa ctaagggagttcctagggaaggcaggcttctgtcgcctcttc atccctgggtttgcagaaatggcagccccctgtaccctctc accaaaccggggactctgtttaattggggcccagaccaaca aaaggcctatcaagaaatcaagcaagctcttctaactgcccc agccctggggttgccagatttgactaagccctttgaactctttg tcgacgagaagcagggctacgccaaaggtgtcctaacgca aaaactgggaccttggcgtcggccggtggcctacctgtcca aaaagctagacccagtagcagctgggtggccccctttgccta cggatggtagcagccattgccgtactgacaaaggatgcagg caagctaaccatgggacagccactagtcattctggcccccc atgcagtagaggcactagtcaaacaacccccccgaccgctg gctttccaacgcccggatgactcactatcaggccttgcttttg	ATNRNGNPDAFHEVYDT KILPLLKQGYRLKVNPGV KGYRKKFSEVGFKEGATI RIGNRVDVLKGTSGFGYT YDGTIFKLKDMDVVYPY GGWKCVKV*SGGSSGGSS* *GSETPGTSESATPESSGGSS* *GGSS*TLNIEDEYRLHETS KEPDVSLGSTWLSDFPQA WAETGGMGLAVRQAPLII PLKATSTPVSIKQYPMSQ EARLGIKPHIQRLLDQGIL VPCQSPWNTPLLPVKKP GTNDYRPVQDLREVNKR VEDIHPTVPNPYNLLSGL PPSHQWYTVLDLKDAFF CLRLHPTSQPLFAFEWRD PEMGISGQLTWTRLPQGF KNSPTLFNEALHRDLADF RIQHPDLILLQYVDDLLL AATSELDCQQGTRALLQ TLGNLGYRASAKKAQIC QKQVKYLGYLLKEGQR WLTEARKETVMGQPTPK TPRQLREFLGKAGFCRLF IPGFAEMAAPLYPLTKPG TLFNWGPDQQKAYQEIK QALLTAPALGLPDLTKPFE LFVDEKQGYAKGVLTQK LGPWRRPVAYLSKKLDP VAAGWPPCLRMVAAIAV LTKDAGKLTMGQPLVILA PHAVEALVKQPPDRWLS NARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEE GLQHNCLDILAEAHGTRP DLTDQPLPDADHTWYTD GSSLLQEGQRKAGAAVT TETEVIWAKALPAGTSAQ RAELIALTQALKMAEGK KLNVYTDSRYAFATAHIH GEIYRRRGWLTSEGKEIK NKDEILALLKALFLPKRL SIIHCPGHQKGHSAEARG NRMADQAARKAAITETP DTSTLLIENSSP*KRPAATK* *KAGQAKKKK* (SEQ ID NO: 197)

	gacacggaccgggtccagttcggaccggtggtagccctga acccggctacgctgctcccactgcctgaggaagggctgca acacaactgccttgatatcctggccgaagcccacggaaccc gacccgacctaacggaccagccgctcccagacgccgacc acacctggtacacggatggaagcagtctcttacaagaggga cagcgtaaggcgggagctgcggtgaccaccgagaccgag gtaatctgggctaaagccctgccagccgggacatccgctca gcgggctgaactgatagcactcacccaggccctaaagatg gcagaaggtaagaagctaaatgtttatactgatagccgttatg cttttgctactgcccatatccatggagaaatatacagaaggcg tgggtggctcacatcagaaggcaaagagatcaaaaataaag acgagatcttggccctactaaaagccctctttctgcccaaaag acttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggc ccgaaaggcagccatcacagagactccagacacctctacc ctcctcatagaaaattcatcacccaaaaggccggcggccac gaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 196)	

Example 9: RNA-Templated Editing of Human Genes in HEK293T Cells Using RNA-Guided Nuclease-Reverse Transcriptase and RNA-guided Nickase-Reverse Transcriptase Fusion Polypeptides

This Example shows genetic modification of human genes utilizing the reference RNA-guided nuclease-reverse transcriptase fusion polypeptide of SEQ ID NO: 173, the Q102R, I206R, V433R RNA-guided nuclease-reverse transcriptase fusion polypeptide of SEQ ID NO: 175, the reference RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide of SEQ D NO: 197, and the Q102R, I206R, V433R RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide of SEQ D NO: 177 constructed in **Examples 6 and 8**. Specifically, the RNA-guided 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 17** of **Example 7**, which comprise scaffold 1, or **shown** in **Table 21** below. The editing template RNA sequences shown in **Table 21** are cloned into plasmids as described in **Example 7** above. The editing template RNA sequences shown in **Table 21** 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 7**.

Table 21. Exemplary Full-Length Editing Template RNA Sequences

Editing template RNA	Target	Scaffold ID	Sequence
4	AAVS1-T6	Reference	GGGAGAGGCUACAAAG**GGGACAUGACCAGAUG CGAAAGUAACUGUCAUGUCCUCAUGCCAAUGG ACUAUGGUGAUGCACCUAGGGGUGUCGUUCCA GCUCCUAGCUCUGCGAAGGAGCAUUCGUUAUC GGAUGCGUUUGUACUUCAAGCCAUAGCCCAGG CCAGGGACGUAUUACUGCUCCUGCCCUAUAGG AGCAUA**UUCUCCUUGACuaagaaUGUAGCCUCUCC CGUCUCUCUCUUGACGCGGUUCUAUCUAGUUAC GCGUUAAACCAACUAGAAA (SEQ ID NO: 198)
5	AAVS1-T6	Scaffold 2	GGGAGAGGCUACAAAG**GGGACAUGACCAGGAA ACUGUCAUGUCCUCAUGCCAAUGGGCUAUGGU GAUGCACCUAGGGGUGUCGUUCCAGCUCCUAG CUCUGCGCCAGAAAUGGCAAGCCAUAGCCCAG GCCAGGGACGUAUUACUGCUCCUGCCCUAUAG GAGC**UUCUCCUUGACuaagaaUGUAGCCUCUCCCG UCUCUCUCUUGACGCGGUUCUAUCUAGUUACGC GUUAAACCAACUAGAAA (SEQ ID NO: 199)
6	EMX1-T7	Reference	CAAGGGACUAUUCAGG**GGGACAUGACCAGAUG CGAAAGUAACUGUCAUGUCCUCAUGCCAAUGG ACUAUGGUGAUGCACCUAGGGGUGUCGUUCCA GCUCCUAGCUCUGCGAAGGAGCAUUCGUUAUC GGAUGCGUUUGUACUUCAAGCCAUAGCCCAGG CCAGGGACGUAUUACUGCUCCUGCCCUAUAGG AGCAUA**AUCCCUGCUUCuagggaGAAUAGUCCCUU GCUCUCUCUCUUGACGCGGUUCUAUCUAGUUAC GCGUUAAACCAACUAGAAA (SEQ ID NO: 200)
7	EMX1-T7	Scaffold 2	CAAGGGACUAUUCAGG**GGGACAUGACCAGGAA ACUGUCAUGUCCUCAUGCCAAUGGGCUAUGGU GAUGCACCUAGGGGUGUCGUUCCAGCUCCUAG CUCUGCGCCAGAAAUGGCAAGCCAUAGCCCAG GCCAGGGACGUAUUACUGCUCCUGCCCUAUAG GAGC**AUCCCUGCUUCuagggaGAAUAGUCCCUUGC UCUCUCUCUUGACGCGGUUCUAUCUAGUUACGC GUUAAACCAACUAGAAA (SEQ ID NO: 201)
8	VEGFA-T6	Reference	GACCAAGGAAGUGUCA**GGGACAUGACCAGAUG CGAAAGUAACUGUCAUGUCCUCAUGCCAAUGG ACUAUGGUGAUGCACCUAGGGGUGUCGUUCCA GCUCCUAGCUCUGCGAAGGAGCAUUCGUUAUC GGAUGCGUUUGUACUUCAAGCCAUAGCCCAGG CCAGGGACGUAUUACUGCUCCUGCCCUAUAGG AGCAUA**UUCACUCAUGCugaacuCACUUCCUUGGU

			CCUCUCUCUCUUGACGCGGUUCUAUCUAGUUAC GCGUUAAACCAACUAGAAA (SEQ ID NO: 202)
9	VEGFA-T6	Scaffold 2	GACCAAGGAAGUGUCA**GGGACAUGACCAGGAA ACUGUCAUGUCCUCAUGCCAAUGGGCUAUGGU GAUGCACCUAGGGGUGUCGUUCCAGCUCCUAG CUCUGCGCCAGAAAUGGCAAGCCAUAGCCCAG GCCAGGGACGUAUUACUGCUCCUGCCCUAUAG GAGC**UUCACUCAUGCugaacuCACUUCCUUGGUCC UCUCUCUCUUGACGCGGUUCUAUCUAGUUACGC GUUAAACCAACUAGAAA (SEQ ID NO: 203)

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

It is anticipated that the RNA-guided 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.

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

This Example shows genetic modification of human genes utilizing fusion polypeptides with a C-terminal reverse transcriptase or an N-terminal reverse transcriptase. Specifically, the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase fusion polypeptide of SEQ ID NO: 175 and the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase (N-terminal) of SEQ ID NO: 177 constructed in **Example 6**. These fusion polypeptides were used to install sequence substitutions of 6 nucleotides into the human target genes when combined with editing template RNAs.

The amino acid sequences for the fusion polypeptides with C-terminal reverse transcriptase are provided in **Table 15** above. The amino acid sequences for the fusion polypeptides with N-terminal reverse transcriptase are provided in **Table 22** below.

Table 22. Fusion Polypeptides with N-Terminal Reverse Transcriptase (RT)

Fusion polypeptide	Nucleotide Sequence	Amino Acid Sequence
RT-Reference RNA-guided nuclease fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctaggg tccacatggctgtctgattttcctcaggcctgggcggaaacc gggggcatgggactggcagttcgccaagctcctctgatcat acctctgaaagcaacctctacccccgtgtccataaaacaata ccccatgtcacaagaagccagactggggatcaagccccac atacagagactgttggaccagggaatactggtaccctgcca	MKRTADGSEFESPKKKR KVTLNIEDEYRLHETSKE PDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIP LKATSTPVSIKQYPMSQE ARLGIKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGT NDYRPVQDLREVNKRVE

	gtccccctggaacacgcccctgctacccgttaagaaaccag	DIHPTVPNPYNLLSGLPPS
	ggactaatgattataggcctgtccaggatctgagagaagtca	HQWYTVLDLKDAFFCLR
	acaagcgggtggaagacatccaccccaccgtgcccaaccc	LHPTSQPLFAFEWRDPEM
	ttacaacctcttgagcgggctcccaccgtcccaccagtggta	GISGQLTWTRLPQGFKNS
	cactgtgcttgatttaaaggatgccttttctgcctgagactcca	PTLFNEALHRDLADFRIQ
	ccccaccagtcagcctctcttcgcctttgagtggagagatcc	HPDLILLQYVDDLLLAAT
	agagatgggaatctcaggacaattgacctggaccagactcc	SELDCQQGTRALLQTLG
	cacagggtttcaaaaacagtcccaccctgtttaatgaggcac	NLGYRASAKKAQICQKQ
	tgcacagagacctagcagacttccggatccagcacccaga	VKYLGYLLKEGQRWLTE
	cttgatcctgctacagtacgtggatgacttactgctggccgcc	ARKETVMGQPTPKTPRQ
	acttctgagctagactgccaacaaggtactcgggccctgtta	LREFLGKAGFCRLFIPGFA
	caaaccctagggaacctcgggtatcgggcctcggccaaga	EMAAPLYPLTKPGTLFN
	aagcccaaatttgccagaaacaggtcaagtatctggggtatc	WGPDQQKAYQEIKQALL
	ttctaaaagagggtcagagatggctgactgaggccagaaaa	TAPALGLPDLTKPFELFV
	gagactgtgatggggcagcctactccgaagacccctcgac	DEKQGYAKGVLTQKLGP
	aactaagggagttcctagggaaggcaggcttctgtcgcctct	WRRPVAYLSKKLDPVAA
	tcatccctgggtttgcagaaatggcagccccctgtaccctct	GWPPCLRMVAAIAVLTK
	caccaaaccggggactctgtttaattggggcccagaccaac	DAGKLTMGQPLVILAPHA
	aaaaggcctatcaagaaatcaagcaagctcttctaactgccc	VEALVKQPPDRWLSNAR
	cagccctggggttgccagatttgactaagccctttgaactcttt	MTHYQALLLDTDRVQFG
	gtcgacgagaagcagggctacgccaaaggtgtcctaacgc	PVVALNPATLLPLPEEGL
	aaaaactgggaccttggcgtcggccggtggcctacctgtcc	QHNCLDILAEAHGTRPDL
	aaaaagctagacccagtagcagctgggtggccccccttgcct	TDQPLPDADHTWYTDGS
	acggatggtagcagccattgccgtactgacaaaggatgcag	SLLQEGQRKAGAAVTTE
	gcaagctaaccatgggacagccactagtcattctggccccc	TEVIWAKALPAGTSAQRA
	catgcagtagaggcactagtcaaacaacccccgaccgct	ELIALTQALKMAEGKKL
	ggctttccaacgcccggatgactcactatcaggccttgctttt	NVYTDSRYAFATAHIHGEI
	ggacacggaccgggtccagttcggaccggtggtagccctg	YRRRGWLTSEGKEIKNK
	aacccggctacgctgctcccactgcctgaggaagggctgc	DEILALLKALFLPKRLSII
	aacacaactgccttgatatcctggccgaagcccacggaacc	HCPGHQKGHSAEARGNR
	cgacccgacctaacggaccagccgctcccagacgccgac	MADQAARKAAITETPDT
	cacacctggtacacggatggaagcagtctcttacaagaggg	STLLIENSSPSGGSSGGSS
	acagcgtaaggcgggagctgcggtgaccaccgagaccga	GSETPGTSESATPESSGGS
	ggtaatctgggctaaagccctgccagccgggacatccgctc	SGGSSRHQRLVFLVDGRN
	agcgggctgaactgatagcactcacccaggccctaaagatg	NILHPTRKLDMVRRWLR
	gcagaaggtaagaagctaaatgtttatactgatagccgttatg	QTNAKVIRFIGKIMVVQV
	cttttgctactgcccatatccatggagaaatatacagaaggcg	FKILSQQKDDPTYWIGVD
	tgggtggctcacatcagaaggcaaagagatcaaaaataaag	PGDTIGICLVRITANNRVK
	acgagatcttggccctactaaaagccctctttctgcccaaaag	RLLSMEFHTRSKEIKQLL
	acttagcataatccattgtccaggacatcaaaagggacacag	ETRRMYRCTRRYFRRKR
	cgccgaggctagaggcaaccggatggctgaccaagcggc	VKKKGRIPKHRPARFENR
	ccgaaaggcagccatcacagagactccagacacctctacc	RRSDGWLTPSVRHFLESH
	ctcctcatagaaaattcatcaccctctggaggatctagcgga	KSVIRKIAECVPDVNISLE
	ggatcctctggcagcgagacaccaggaacaagcgagtcag	YAKFDIQKMMNPDIEGE
	caacaccagagagcagtggcggcagcagcggcggcagc	QYQRGKLFGYQNIRAYV
	agcagacaccagagactcgtgtttctcgtggacggtagaaa	LHRDKYTCQLCAKKNVP
	caacatcctgcatcctaccagaaagctggatatggtcagaag	FQLHHVVPRSQNGKDHP
	atggctgagacaaacaaacgccaaggtgatcagattcatcg	DNLITLCPECHVKVHTDT
	gaaaaatcatggtggtccaggtgttcaagatcctgtctcagc	KLLKKLQRKIKSKIADFT
	agaaagacgaccctacatactggatcggcgtggacccagg	STSRLNIIMSFLKHWLEN

	cgacaccatcggcatctgcctggtcagaatcacagccaaca acagagttaagcggctgctgagcatggaattccacaccaga tctaaggaaatcaagcagctgctggaaacccggagaatgta ccggtgcacacggagatactttagaagaaagcgggtgaag aagaaagggagaattcccaagcacagacctgctagatttga gaacagaaggcggagcgacggctggctgacccctagcgt gcgccacttccttgagagccacaagtccgtgatccggaaga tcgccgagtgtgtgccagatgtgaacatcagcctggagtac gccaagttcgacatccaaaagatgatgaaccccgatatcga gggcgagcagtaccagagaggcaaactgttcggatatcag aatatccgcgcctacgtgctgcatagagataagtacacctgt cagctgtgcgccaagaaaaacgtgcccttccagctgcacca cgtggtgcctagaagccagaacggcaaggaccaccccgat aacctgatcaccctgtgtcctgagtgccacgtgaaagtgcac acagacaccaagctgctgaaaaaactgcagcggaagatta agagcaagattgctgatttcaccagcaccagcagactgaata ttatcatgagcttcctgaagcactggctggaaaacgagttcg gcaaggatcgggtgtggttctgcatcggagccgactcttatg tgatccggaaggaactgaacctggacaagacacactgcaa cgacgcttactgcatcgccatcaacaagctggtgagatttga caccgacctggaacacgaaagcgtggtggataagcaagtg atccggttcgagcagaaaagacagcacaaccggagaatga aatctctggagagaaacaggctgtactacgtggtcaaggac aagaagcgcgtgcgggtggccaccaacagaaatggcaatc ctgatgccttccacgaggtgtacgacacaaagatcctgcctc tgctgaagcagggctatcggctgaaagtgaaccctggcgtt aagggctacaggaagaagttcagcgaggtgggatttaagg aaggcgccacaatcagaatcggaaatcgggtggacgtgct gaagggcacctccggcttcggctacacctacgacggcacc atcttcaagctcaaggacatggacgtggtgtaccctacggc ggctggaagtgcgtgaaggtgaaaaggccggcggccacg aaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 204)	EFGKDRVWFCIGADSYVI RKELNLDKTHCNDAYCI AINKLVRFDTDLEHESVV DKQVIRFEQKRQHNRRM KSLERNRLYYVVKDKKR VRVATNRNGNPDAFHEV YDTKILPLLKQGYRLKV NPGVKGYRKKFSEVGFK EGATIRIGNRVDVLKGTS GFGYTYDGTIFKLKDMD VVYPYGGWKCVKVKRP AATKKAGQAKKKK (SEQ ID NO: 205)
RT-Q102R, I206R, V433R variant RNA-guided nuclease fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctaggg tccacatggctgtctgattttcctcaggcctgggcggaaacc gggggcatgggactggcagttcgccaagctcctctgatcat acctctgaaagcaacctctacccccgtgtccataaaacaata ccccatgtcacaagaagccagactggggatcaagccccac atacagagactgttggaccagggaatactggtaccctgcca gtccccctggaacacgcccctgctacccgttaagaaaccag ggactaatgattataggcctgtccaggatctgagagaagtca acaagcgggtggaagacatccaccccaccgtgcccaaccc ttacaacctcttgagcgggctcccaccgtcccaccagtggta cactgtgcttgatttaaaggatgccttttttctgcctgagactcca ccccaccagtcagcctctcttcgcctttgagtggagagatcc agagatgggaatctcaggacaattgacctggaccagactcc cacagggtttcaaaaacagtcccaccctgtttaatgaggcac tgcacagagacctagcagacttccggatccagcacccaga	MKRTADGSEFESPKKKR KVTLNIEDEYRLHETSKE PDVSLGSTWLSDFPQAW AETGGMGLAVRQAPLIIP LKATSTPVSIKQYPMSQE ARLGIKPHIQRLLDQGILV PCQSPWNTPLLPVKKPGT NDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLR LHPTSQPLFAFEWRDPEM GISGQLTWTRLPQGFKNS PTLFNEALHRDLADFRIQ HPDLILLQYVDDLLLAAT SELDCQQGTRALLQTLG NLGYRASAKKAQICQKQ VKYLGYLLKEGQRWLTE

	cttgatcctgctacagtacgtggatgacttactgctggccgcc	ARKETVMGQPTPKTPRQ
	acttctgagctagactgccaacaaggtactcgggccctgtta	LREFLGKAGFCRLFIPGFA
	caaaccctagggaacctcgggtatcgggcctcggccaaga	EMAAPLYPLTKPGTLFN
	aagcccaaatttgccagaaacaggtcaagtatctggggtatc	WGPDQQKAYQEIKQALL
	ttctaaaagagggtcagagatggctgactgaggccagaaaa	TAPALGLPDLTKPFELFV
	gagactgtgatggggcagcctactccgaagacccctcgac	DEKQGYAKGVLTQKLGP
	aactaagggagttcctagggaaggcaggcttctgtcgcctct	WRRPVAYLSKKLDPVAA
	tcatccctgggtttgcagaaatggcagccccctgtaccctct	GWPPCLRMVAAIAVLTK
	caccaaaccggggactctgtttaattggggcccagaccaac	DAGKLTMGQPLVILAPHA
	aaaaggcctatcaagaaatcaagcaagctcttctaactgccc	VEALVKQPPDRWLSNAR
	cagccctggggttgccagatttgactaagccctttgaactctttt	MTHYQALLLDTDRVQFG
	gtcgacgagaagcagggctacgccaaaggtgtcctaacgc	PVVALNPATLLPLPEEGL
	aaaaactgggaccttggcgtcggccggtggcctacctgtcc	QHNCLDILAEAHGTRPDL
	aaaaagctagacccagtagcagctgggtggcccccttgcct	TDQPLPDADHTWYTDGS
	acggatggtagcagccattgccgtactgacaaaggatgcag	SLLQEGQRKAGAAVTTE
	gcaagctaaccatgggacagccactagtcattctggccccc	TEVIWAKALPAGTSAQRA
	catgcagtagaggcactagtcaaacaacccccccgaccgct	ELIALTQALKMAEGKKL
	ggctttccaacgcccggatgactcactatcaggccttgctttt	NVYTDSRYAFATAHIHGEI
	ggacacggaccgggtccagttcggaccggtggtagccctg	YRRRGWLTSEGKEIKNK
	aacccggctacgctgctcccactgcctgaggaagggctgc	DEILALLKALFLPKRLSII
	aacacaactgccttgatatcctggccgaagcccacggaacc	HCPGHQKGHSAEARGNR
	cgacccgacctaacggaccagccgctcccagacgccgac	MADQAARKAAITETPDT
	cacacctggtacacggatggaagcagtctcttacaagaggg	STLLIENSSPSGGSSGGSS
	acagcgtaaggcgggagctgcggtgaccaccgagaccga	GSETPGTSESATPESSGGS
	ggtaatctgggctaaagccctgccagccgggacatccgctc	SGGSSRHQRLVFLVDGRN
	agcgggctgaactgatagcactcacccaggccctaaagatg	NILHPTRKLDMVRRWLR
	gcagaaggtaagaagctaaatgtttatactgatagccgttatg	QTNAKVIRFIGKIMVVQV
	cttttgctactgcccatatccatggagaaatatacagaaggcg	FKILSQQKDDPTYWIGVD
	tgggtggctcacatcagaaggcaaagagatcaaaaataaag	PGDTIGICLVRITANNRVK
	acgagatcttggccctactaaaagccctctttctgcccaaaag	RLLSMEFHTRSKEIKRLL
	acttagcataatccattgtccaggacatcaaaagggacacag	ETRRMYRCTRRYFRRKR
	cgccgaggctagaggcaaccggatggctgaccaagcggc	VKKKGRIPKHRPARFENR
	ccgaaaggcagccatcacagagactccagacacctctacc	RRSDGWLTPSVRHFLESH
	ctcctcatagaaaattcatcaccctctggaggatctagcgga	KSVIRKIAECVPDVNISLE
	ggatcctctggcagcgagacaccaggaacaagcgagtcag	YAKFDIQKMMNPDIEGE
	caacaccagagagcagtggcggcagcagcggcggcagc	QYQRGKLFGYQNRRAYV
	agcagacaccagagactcgtgtttctcgtggacggtagaaa	LHRDKYTCQLCAKKNVP
	caacatcctgcatcctaccagaaagctggatatggtcagaag	FQLHHVVPRSQNGKDHP
	atggctgagacaaacaaacgccaaggtgatcagattcatcg	DNLITLCPECHVKVHTDT
	gaaaaatcatggtggtccaggtgttcaagatcctgtctcagc	KLLKKLQRKIKSKIADFT
	agaaagacgaccctacatactggatcggcgtggacccagg	STSRLNIIMSFLKHWLEN
	cgacaccatcggcatctgcctggtcagaatcacagccaaca	EFGKDRVWFCIGADSYVI
	acagagttaagcggctgctgagcatggaattccacaccaga	RKELNLDKTHCNDAYCI
	tctaaggaaatcaagCGGctgctggaaacccggagaatg	AINKLVRFDTDLEHESVV
	taccggtgcacacggagatactttagaagaaagcgggtgaa	DKQVIRFEQKRQHNRRM
	gaagaaagggagaattcccaagcacagacctgctagatttg	KSLERNRLYYVVKDKKR
	agaacagaaggcggagcgacggctggctgaccctagcg	VRVATNRNGNPDAFHEV
	tgcgccacttccttgagagccacaagtccgtgatccggaag	YDTKILPLLKQGYRLKV
	atcgccgagtgtgtgccagatgtgaacatcagcctggagta	NPGVKGYRKKFSERGFK
	cgccaagttcgacatccaaaagatgatgaaccccgatatcg	EGATIRIGNRVDVLKGTS

	agggcgagcagtaccagagaggcaaactgttcggatatca gaatCGGcgcgcctacgtgctgcatagagataagtacac ctgtcagctgtgcgccaagaaaaacgtgcccttccagctgc accacgtggtgcctagaagccagaacggcaaggaccacc ccgataacctgatcaccctgtgtcctgagtgccacgtgaaag tgcacacagacaccaagctgctgaaaaaactgcagcggaa gattaagagcaagattgctgatttcaccagcaccagcagact gaatattatcatgagcttcctgaagcactggctggaaaacga gttcggcaaggatcgggtgtggttctgcatcggagccgact cttatgtgatccggaaggaactgaacctggacaagacacac tgcaacgacgcttactgcatcgccatcaacaagctggtgag atttgacaccgacctggaacacgaaagcgtggtggataagc aagtgatccggttcgagcagaaaagacagcacaaccggag aatgaaatctctggagagaaacaggctgtactacgtggtcaa ggacaagaagcgcgtgcgggtggccaccaacagaaatgg caatcctgatgccttccacgaggtgtacgacacaaagatcct gcctctgctgaagcagggctatcggctgaaagtgaaccctg gcgttaagggctacaggaagaagttcagcgagCGGgga tttaaggaaggcgccacaatcagaatcggaaatcgggtgga cgtgctgaagggcacctccggcttcggctacacctacgacg gcaccatcttcaagctcaaggacatggacgtggtgtaccct acggcggctggaagtgcgtgaaggtgaaaaggccggcgg ccacgaaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 206)	GFGYTYDGTIFKLKDMD VVYPYGGWKCVKVKRP AATKKAGQAKKKK (SEQ ID NO: 207)
RT-Q102R, I206R, V433R variant RNA-guided nickase (H232A) fusion polypeptide	atgaaacggacagccgacggaagcgagttcgagtcaccaa agaagaagcggaaagtcaccctaaatatagaagatgagtat cggctacatgagacctcaaaagagccagatgtttctctaggg tccacatggctgtctgatttttcctcaggcctgggcggaaacc gggggcatgggactggcagttcgccaagctcctctgatcat acctctgaaagcaacctctacccccgtgtccataaaacaata ccccatgtcacaagaagccagactggggatcaagccccac atacagagactgttggaccagggaatactggtaccctgcca gtccccctggaacacgcccctgctacccgttaagaaaccag ggactaatgattataggcctgtccaggatctgagagaagtca acaagcgggtggaagacatccaccccaccgtgcccaaccc ttacaacctcttgagcgggctcccaccgtcccaccagtggta cactgtgcttgatttaaaggatgccttttctgcctgagactcca ccccaccagtcagcctctcttcgcctttgagtggagagatcc agagatgggaatctcaggacaattgacctggaccagactcc cacagggtttcaaaaacagtcccaccctgtttaatgaggcac tgcacagagacctagcagacttccggatccagcacccaga cttgatcctgctacagtacgtggatgacttactgctggccgcc acttctgagctagactgccaacaaggtactcgggccctgtta caaaccctagggaacctcgggtatcgggcctcggccaaga aagcccaaatttgccagaaacaggtcaagtatctggggtatc ttctaaaagagggtcagagatggctgactgaggccagaaaa gagactgtgatggggcagcctactccgaagacccctcgac aactaagggagttcctagggaaggcaggcttctgtcgcctct tcatccctgggtttgcagaaatggcagccccctgtaccctct caccaaaccggggactctgtttaattggggcccagaccaac	*MKRTADGSEFESPKKKRK* *V*TLNIEDEYRLHETSKEP DVSLGSTWLSDFPQAWA ETGGMGLAVRQAPLIIPL KATSTPVSIKQYPMSQEA RLGIKPHIQRLLDQGILVP CQSPWNTPLLPVKKPGT NDYRPVQDLREVNKRVE DIHPTVPNPYNLLSGLPPS HQWYTVLDLKDAFFCLR LHPTSQPLFAFEWRDPEM GISGQLTWTRLPQGFKNS PTLFNEALHRDLADFRIQ HPDLILLQYVDDLLLAAT SELDCQQGTRALLQTLG NLGYRASAKKAQICQKQ VKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQ LREFLGKAGFCRLFIPGFA EMAAPLYPLTKPGTLFN WGPDQQKAYQEIKQALL TAPALGLPDLTKPFELFV DEKQGYAKGVLTQKLGP WRRPVAYLSKKLDPVAA GWPPCLRMVAAIAVLTK DAGKLTMGQPLVILAPHA

	aaaaggcctatcaagaaatcaagcaagctcttctaactgccc cagccctggggttgccagatttgactaagccctttgaactcttt gtcgacgagaagcagggctacgccaaaggtgtcctaacgc aaaaactgggaccttggcgtcggccggtggcctacctgtcc aaaaagctagacccagtagcagctgggtggccccttgcct acggatggtagcagccattgccgtactgacaaaggatgcag gcaagctaaccatgggacagccactagtcattctggccccc catgcagtagaggcactagtcaaacaacccccgaccgct ggctttccaacgcccggatgactcactatcaggccttgctttt ggacacggaccgggtccagttcggaccggtggtagccctg aacccggctacgctgctcccactgcctgaggaagggctgc aacacaactgccttgatatcctggccgaagcccacggaacc cgacccgacctaacggaccagccgctcccagacgccgac cacacctggtacacggatggaagcagtctcttacaagaggg acagcgtaaggcgggagctgcggtgaccaccgagaccga ggtaatctgggctaaagccctgccagccgggacatccgctc agcgggctgaactgatagcactcacccaggccctaaagatg gcagaaggtaagaagctaaatgtttatactgatagccgttatg cttttgctactgcccatatccatggagaaatatacagaaggcg tgggtggctcacatcagaaggcaaagagatcaaaaataaag acgagatcttggccctactaaaagccctctttctgcccaaaag acttagcataatccattgtccaggacatcaaaagggacacag cgccgaggctagaggcaaccggatggctgaccaagcggc ccgaaaggcagccatcacagagactccagacacctctacc ctcctcatagaaaattcatcaccctctggaggatctagcgga ggatcctctggcagcgagacaccaggaacaagcgagtcag caacaccagagagcagtggcggcagcagcggcggcagc agcagacaccagagactcgtgtttctcgtggacggtagaaa caacatcctgcatcctaccagaaagctggatatggtcagaag atggctgagacaaacaaacgccaaggtgatcagattcatcg gaaaaatcatggtggtccaggtgttcaagatcctgtctcagc agaaagacgaccctacatactggatcggcgtggacccagg cgacaccatcggcatctgcctggtcagaatcacagccaaca acagagttaagcggctgctgagcatggaattccacaccaga tctaaggaaatcaagcggctgctggaaacccggagaatgta ccggtgcacacggagatactttagaagaaagcgggtgaag aagaaagggagaattcccaagcacagacctgctagatttga gaacagaaggcggagcgacggctggctgacccctagcgt gcgccacttccttgagagccacaagtccgtgatccggaaga tcgccgagtgtgtgccagatgtgaacatcagcctggagtac gccaagttcgacatccaaaagatgatgaaccccgatatcga gggcgagcagtaccagagaggcaaactgttcggatatcag aatcggcgcgcctacgtgctgcatagagataagtacacctgt cagctgtgcgccaagaaaaacgtgcccttccagctgcac**<u>G</u>** **<u>CC</u>**gtggtgcctagaagccagaacggcaaggaccacccc gataacctgatcaccctgtgtcctgagtgccacgtgaaagtg cacacagacaccaagctgctgaaaaaactgcagcggaaga ttaagagcaagattgctgatttcaccagcaccagcagactga atattatcatgagcttcctgaagcactggctggaaaacgagtt cggcaaggatcgggtgtggttctgcatcggagccgactctta	VEALVKQPPDRWLSNAR MTHYQALLLDTDRVQFG PVVALNPATLLPLPEEGL QHNCLDILAEAHGTRPDL TDQPLPDADHTWYTDGS SLLQEGQRKAGAAVTTE TEVIWAKALPAGTSAQRA ELIALTQALKMAEGKKL NVYTDSRYAFATAHIHGEI YRRRGWLTSEGKEIKNK DEILALLKALFLPKRLSII HCPGHQKGHSAEARGNR MADQAARKAAITETPDT STLLIENSSPSGGSSGGSS GSETPGTSESATPESSGGS SGGSSRHQRLVFLVDGRN NILHPTRKLDMVRRWLR QTNAKVIRFIGKIMVVQV FKILSQQKDDPTYWIGVD PGDTIGICLVRITANNRVK RLLSMEFHTRSKEIKRLL ETRRMYRCTRRYFRRKR VKKKGRIPKHRPARFENR RRSDGWLTPSVRHFLESH KSVIRKIAECVPDVNISLE YAKFDIQKMMNPDIEGE QYQRGKLFGYQNRRAYV LHRDKYTCQLCAKKNVP FQLH**<u>A</u>**VVPRSQNGKDHP DNLITLCPECHVKVHTDT KLLKKLQRKIKSKIADFT STSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVI RKELNLDKTHCNDAYCI AINKLVRFDTDLEHESVV DKQVIRFEQKRQHNRRM KSLERNRLYYVVKDKKR VRVATNRNGNPDAFHEV YDTKILPLLKQGYRLKV NPGVKGYRKKFSERGFK EGATIRIGNRVDVLKGTS GFGYTYDGTIFKLKDMD VVYPYGGWKCVKV*KRPA ATKKAGQAKKKK* (SEQ ID NO: 209)

	tgtgatccggaaggaactgaacctggacaagacacactgca acgacgcttactgcatcgccatcaacaagctggtgagatttg acaccgacctggaacacgaaagcgtggtggataagcaagt gatccggttcgagcagaaaagacagcacaaccggagaatg aaatctctggagagaaacaggctgtactacgtggtcaagga caagaagcgcgtgcgggtggccaccaacagaaatggcaat cctgatgccttccacgaggtgtacgacacaaagatcctgcct ctgctgaagcagggctatcggctgaaagtgaaccctggcgt taagggctacaggaagaagttcagcgagcggggatttaag gaaggcgccacaatcagaatcggaaatcgggtggacgtgc tgaagggcacctccggcttcggctacacctacgacggcacc atcttcaagctcaaggacatggacgtggtgtacccctacggc ggctggaagtgcgtgaaggtgaaaaggccggcggccacg aaaaaggccggccaggcaaaaaagaaaaag (SEQ ID NO: 208)	

Editing template RNAs were designed and cloned into plasmids as described in **Example 7**, with the exception that scaffold 1 was replaced by scaffold 3. The editing template RNA sequences are shown in **Table 23** below. As in **Example 7**, the editing template RNA 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.

Table 23. Full-length Editing Template RNA Sequences

Editing template RNA	Target	Scaffold ID	Sequence
10	AAVS1-T6	Scaffold 3	GGGAGAGGCUACAAAG**GGGACATGACCAGGAA ACTGTCATGTCCTCATGGCAATGGGCTATGGTG ATGCACCTAGGGGTGTCGTTCCAGCTCCTAGCT CTGCGCCAGAAATGGCAAGCCATAGCCCAGCC CAGGGACGTATTACTGCTCCTGCCCTATAGGAG** CUUCUCCUUGACuaagaaUGUAGCCUCUCCCGUCUC UCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 210)
11	EMX1-T7	Scaffold 3	CAAGGGACUAUUCAGG**GGGACATGACCAGGAA ACTGTCATGTCCTCATGGCAATGGGCTATGGTG ATGCACCTAGGGGTGTCGTTCCAGCTCCTAGCT CTGCGCCAGAAATGGCAAGCCATAGCCCAGCC CAGGGACGTATTACTGCTCCTGCCCTATAGGAG** CAUCCCUGCUUCuagggaGAAUAGUCCCUUGCUCU CUCUCUUGACGCGGUUCUAUCUAGUUACGCGUU AAACCAACUAGAAA (SEQ ID NO: 211)

12	VEGFA-T6	Scaffold 3	GACCAAGGAAGUGUCA**GGGACATGACCAGGAA ACTGTCATGTCCTCATGGCAATGGGCTATGGTG ATGCACCTAGGGGTGTCGTTCCAGCTCCTAGCT CTGCGCCAGAAATGGCAAGCCATAGCCCAGCC CAGGGACGTATTACTGCTCCTGCCCTATAGGAG** CUUCACUCAUGCugaacuCACUUCCUUGGUCCUCUC UCUCUUGACGCGGUUCUAUCUAGUUACGCGUUA AACCAACUAGAAA (SEQ ID NO: 212)

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

Editing template RNAs were combined with nickase-reverse transcriptase fusion polypeptides, transfected into HEK293T cells, and the genomic DNA was recovered approximately 72 hours post transfection as described in **Example 7**. Samples were prepared for NGS and analyzed as described in **Example 7**. The percentage of NGS reads comprising indels and the 6-nucleotide edits are shown in **Table 24**. The data presented therein are the average of three technical replicates unless indicated otherwise.

Table 24. Editing Efficiencies with Exemplary RNA-guided Nuclease-Reverse Transcriptase Fusion Polypeptides

RT-Nickase fusion polypeptide	RT terminal Fusion	Editing Template RNAs	Target	Average % Indel	Average % 6-nt substitution installation
Q102R, I206R, V433R variant nickase (H232A)-RT fusion polypeptide	C-terminal	10	AAVS1-T6	0.27	0.02
Q102R, I206R, V433R variant nickase (H232A)-RT fusion polypeptide	C-terminal	11	EMX1-T7	0.53	0.05
Q102R, I206R, V433R variant nickase (H232A)-RT fusion polypeptide	C-terminal	12	VEGFA-T6	0.27	0.02
RT-Q102R, I206R, V433R variant nickase (H232A) fusion polypeptide	N-Terminal	10	AAVS1-T6	0.26	0.01
RT-Q102R, I206R, V433R variant nickase (H232A) fusion polypeptide	N-Terminal	11	EMX1-T7	0.52	0.01
RT-Q102R, I206R, V433R variant nickase	N-Terminal	12	VEGFA-T6	0.24	0.01

(H232A) fusion polypeptide					
Guide-only control	--	10	AAVS1-T6	0.27	0.00
Guide-only control	--	11	EMX1-T7	0.50	0.00
Guide-only control	--	12	VEGFA-T6	0.23	0.00

As shown in **Table 24**, both the Q102R, I206R, V433R variant RNA-guided nickase (H232A)-reverse transcriptase and the reverse transcriptase (N-terminal)-Q102R, I206R, V433R variant RNA-guided nickase (H232A) fusion polypeptides introduced substitutions encoded by the tested editing template RNAs at the AAVS1, EMX1, and VEGFA target loci. As in **Example 7**, editing rates for introducing the encoded substitutions were above background, while indels were observed to be approximately equal to the background rates observed with the guide-only control samples.

This Example thus shows that the tested RNA-guided nickase variant – RT fusion polypeptides, in either orientation, incorporated substitutions encoded by editing template RNAs into human genes. It also shows that the scaffold portion of the editing template RNA can be substituted with other validated scaffolds, such as those that retain activity in the nuclease context. Examples are provided in **Table 5B** above.

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.

WHAT IS CLAIMED IS:

1. A gene editing system comprising:

(a) a fusion polypeptide comprising an RNA-guided nuclease polypeptide and a reverse transcriptase (RT) polypeptide, or a first nucleic acid encoding the fusion polypeptide; wherein the RNA-guided nuclease polypeptide comprises an amino acid sequence at least 90% identical to SEQ ID NO: 1; 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;

wherein the gRNA comprises (i) a scaffold sequence recognizable by the RNA-guided nuclease polypeptide and (ii) a spacer sequence specific to a target sequence within a genomic site of interest, wherein the target sequence is adjacent to a protospacer adjacent motif (PAM) specific to the RNA-guided 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 further comprises one or more nuclear localization signals (NLSs) upstream or downstream to the RNA-guided nuclease polypeptide, the RT polypeptide, or both.

3. The gene editing system of claim 2, wherein the fusion polypeptide comprises a peptide linker located between the RNA-guided 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.

4. The gene editing system of claim 2, wherein the fusion polypeptide, from N-terminus to C-terminus, comprises:

(a) a first NLS, the RNA-guided nuclease polypeptide, the RT polypeptide, and a second NLS; optionally wherein the fusion polypeptide further comprises a peptide linker between the RNA-guided nuclease polypeptide and the RT polypeptide; or

(b) a first NLS, the RT polypeptide, the RNA-guided nuclease polypeptide, and a second NLS; optionally wherein the fusion polypeptide further comprises a peptide linker

between the RT polypeptide and the RNA-guided nuclease polypeptide.

5. The gene editing system of any one of claims 1-4, wherein the RNA-guided nuclease polypeptide is a variant of SEQ ID NO: 1 comprising at least one mutation relative to SEQ ID NO: 1.

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

(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; relative to SEQ ID NO: 1; or

(iii) a combination of (i) and (ii).

7. The gene editing system of claim 6, wherein the variant comprises (i) and wherein the one or more mutations are at positions H231, H232, H255, D67, E176, and/or D329 of SEQ ID NO: 1, optionally at positions H231, H232, and/or H255 of SEQ ID NO: 1.

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

(a) the mutation at H231 is an amino acid substitution of H231A, H231G, H231L, or H231S;

(b) the mutation at H232 is an amino acid substitution of H232A, H232G, H232L, or H232S; and

(c) the mutation at H255 is an amino acid substitution of H255A, H255G, H255L, or H255S,

optionally wherein the mutation at H232 is amino acid substitution of H232A or H232L.

9. The gene editing system of claim 7, wherein the variant comprises a nickase mutation as position H232 of SEQ ID NO: 1, optionally wherein the nickase mutation is H232A.

10. The gene editing system of any one of claims 5-9, wherein the RNA-guided 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, wherein the variant comprises the one or more arginine and/or lysine substitutions of (ii), optionally the one or more arginine substitutions and wherein the one or more arginine and/or lysine substitutions are located in the BH domain, in the REC domain, in the PLL domain, in the WED domain, in the PID domain, or a combination thereof.

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 E99, Q102, E105, T106, T113, E191, I206, N239, V256, I272, A277, E401, E432, and V433 in SEQ ID NO: 1.

12. The gene editing system of any one of claims 5-11, wherein the RNA-guided 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 RNA-guided 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) Q102, I206, and V433;

b) Q102, V256, and V433;

c) Q102, E191, I272, and V433;

d) Q102, V256, and V433;

e) Q102, E191, I206, and V433;

f) Q102, I272, and V433;

g) E105, V256, and V433;

h) E105, E191, V256, and V433;

i) E105, E191, I272, E401, and 433; or

j) E105, E191, N239, I272, and V433.

14. The gene editing system of claim 13, wherein the RNA-guided nuclease polypeptide comprises the following arginine substitutions relative to SEQ ID NO: 1:

a) Q102R, I206R, and V433R;

b) Q102R, V256R, and V433R;

c) Q102R, E191R, I272R, and V433R;

d) Q102R, V256R, and V433R;

e) Q102R, E191R, I206R, and V433R;

f) Q102R, I272R, and V433R;

g) E105R, V256R, and V433R;

h) E105R, E191R, V256R, and V433R;

i) E105R, E191R, I272R, E401R, and 433R; or

j) E105R, E191R, N239R, I272R, and V433R.

optionally wherein the RNA-guided nuclease polypeptide comprises the arginine substitutions of (a).

15. 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 H231, H232, and/or H255 relative to SEQ ID NO: 1; and (b) the one or more arginine and/or lysine substitutions, which optionally are at positions Q102, I206, and V433.

16. The gene editing system of claim 15, wherein the variant comprises:

(a) the nickase mutation of H232A; and

(b) the arginine substitutions of Q102R, I206R, and V433R.

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

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

19. The gene editing system of claim 5, wherein the RNA-guided nuclease variant comprises the at least one mutation in the RNA-guided nuclease polypeptides listed in **Table 1** or **Table 11**.

20. The gene editing system of any one of claims 1-19, wherein the RNA-guided nuclease polypeptide in the fusion polypeptide does not include its native N-terminus methionine residue.

21. The gene editing system of claim 1, wherein the RNA-guided nuclease polypeptide is listed in **Table 1** or **Table 11**.

22. The gene editing system of any one of claims 1-21, 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: 56 or SEQ ID NO: 57.

23. The gene editing system of claim 1, wherein the fusion polypeptide is set forth in any one of **Table 12, Table 17, Table 20,** and **Table 22**.

24. The gene editing system of any one of claims 1-23, wherein the spacer sequence in the gRNA of (b) is 14-30-nucleotide in length; optionally 14-20-nucleotide in length.

25. The gene editing system of any one of claims 1-24, wherein the PAM is 5'-RRT-3' or 5'-NRT-3', in which R represents A or G and N represents any nucleotide.

26. The gene editing system of any one of claims 1-25, wherein the scaffold sequence comprises a nucleotide sequence at least 70% identical to SEQ ID NO: 2; optionally wherein the scaffold sequence comprises a nucleotide sequence at least 75% identical to SEQ ID NO: 2.

27. The gene editing system of claim 26, wherein, relative to SEQ ID NO: 2, the scaffold sequence comprises:

(a) one or more mutations within nucleotides 11-27, optionally wherein the one or more mutations comprise deletions and nucleotide substitutions;

(b) one or more mutations within nucleotides 96-128, optionally wherein the one or more mutations comprise deletions and nucleotide substitutions;

(c) a deletion within nucleotides 155-160, optionally wherein the deletion is at position 160;

(d) a deletion within nucleotides 180-182; or

(e) a combination of any of (a)-(d).

28. The gene editing system of claim 27, wherein the scaffold sequence comprises

the nucleotide sequence of any one of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, and SEQ ID NOs: 146-148 and 150-152.

29. The gene editing system of claim 27, wherein

(a) the RNA-guided nuclease polypeptide comprises the amino acid sequence of SEQ ID NO: 1; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 2;

(b) the RNA-guided nuclease polypeptide is a variant of SEQ ID NO: 1 comprising mutations at positions Q102, I206, and V433, which optionally are arginine substitutions Q102R, I206R, and V433R; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 146, SEQ ID NO 147, SEQ ID NO 148, SEQ ID NO 150, SEQ ID NO 151, or SEQ ID NO 152; or

(c) the RNA-guided nuclease polypeptide is a variant of SEQ ID NO: 1 comprising mutations at positions Q102, I206, V433, and H232, which optionally are Q102R, I206R, V433R, and H232A; and the scaffold comprises the nucleotide sequence of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 146, SEQ ID NO 147, SEQ ID NO 148, SEQ ID NO 150, SEQ ID NO 151, or SEQ ID NO 152.

30. The gene editing system of any one of claims 1-29, wherein the PBS in the RT donor RNA is 5-50-nucleotide in length; optionally 5-20-nucleotide in length.

31. The gene editing system of any one of claims 1-30, wherein the PBS binds a PBS-targeting site that is adjacent to or overlaps with the target sequence.

32. The gene editing system of claim 31, 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.

33. The gene editing system of any one of claims 1-32, wherein the template sequence in the RT donor RNA is 5-100-nucleotide in length; optionally 15-25-nucleotide in length.

34. The gene editing system of any one of claims 1-33, 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.

35. The gene editing system of claim 34, 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.

36. The gene editing system of any one of claims 1-35, wherein the RNA molecule further comprises a 3' end extension.

37. The gene editing system of any one of claims 1-36, 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.

38. The gene editing system of any one of claims 1-37, 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.

39. The gene editing system of any one of claims 1-38, wherein the system comprises the fusion polypeptide.

40. The gene editing system of any one of claims 1-40, wherein the system comprises the first nucleic acid encoding the fusion polypeptide.

41. The gene editing system of claim 40, wherein the first nucleic acid is located on a vector, which optionally is a viral vector; optionally wherein the vector further comprises the second nucleic acid encoding the RNA molecule.

42. The gene editing system of claim 40, wherein the first nucleic acid is a messenger RNA (mRNA); optionally wherein the gene editing system comprises the RNA molecule.

43. The gene editing system of any one of claims 1-42, wherein the system comprises the RNA molecule.

44. The gene editing system of any one of claims 1-42, wherein the system comprises the second nucleic acid encoding the RNA molecule.

45. The gene editing system of claim 44, wherein the second nucleic acid is located on a vector, which optionally is a viral vector.

46. The gene editing system of any one of claims 1-45, wherein the system comprises one or more lipid excipients associated with one or more of the fusion polypeptide or the first nucleic acid encoding the fusion polypeptide, and the RNA molecule or the second nucleic acid encoding the RNA molecule; optionally wherein the one or more lipid excipients form nanoparticles (LNPs); optionally wherein the system comprises (i) the first nucleic acid, which is a messenger RNA encoding the fusion polypeptide, and (ii) the RNA molecule.

47. The gene editing system of any one of claims 1-45, wherein the system comprises one or more viral vectors, optionally one or more adeno-associated viral (AAV) vectors encoding the fusion polypeptide and/or the RNA molecule; optionally wherein the system comprises an AAV encoding both the fusion polypeptide and the RNA molecule.

48. A pharmaceutical composition comprising the gene editing system of any one of claims 1-47.

49. A kit comprising the (a) fusion polypeptide or the first nucleic acid encoding the fusion polypeptide, and (b) the RNA molecule or the second nucleic acid encoding the RNA molecule of the gene editing system set forth in any one of claims 1-47.

50. A gene editing method, comprising delivering the gene editing system of any one of claims 1-47 to a host cell to edit a genomic site targeted by the gRNA of the gene editing system.

51. The gene editing method of claim 50, wherein the host cell is cultured *in vitro*.

52. The gene editing method of claim 51, wherein the host cell is located in a subject who needs the gene editing.

53. A fusion polypeptide, comprising an RNA-guided nuclease polypeptide set forth in any one of claims 1-19 and a reverse transcriptase polypeptide set forth in claim 1 or claim 20.

54. The fusion polypeptide of claim 53, which comprises the amino acid sequence set forth in **Table 12, Table 17, Table 20,** or **Table 22**.

55. A nucleic acid encoding the fusion polypeptide of claim 53 or claim 54.

56. The nucleic acid of claim 55, which comprises the nucleotide sequence set forth in **Table 12, Table 17, Table 20,** or **Table 22**.

57. The nucleic acid of claim 56, which is a vector, optionally an expression vector.