RNA-GUIDED NUCLEASE POLYPEPTIDES AND GENE EDITING SYSTEMS COMPRISING SUCH

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of the filing date of U.S. Provisional Application No. 63/570,061, 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 063586-540001WO_Seq-Listing_ST26.xml and is 209,467 bytes in size.

BACKGROUND

RNA-guided nucleases (*e.g.*, CRISPR-Cas nucleases) are found in archaea and bacteria with roles that include defending the host species against foreign genetic elements as part of adaptive immune systems.

Gene editing systems can comprise a nuclease and one or more RNA components that direct the nuclease to a target genomic site for gene editing. It is of interest to develop efficient RNA-guided nucleases to improve gene editing efficiency.

SUMMARY OF THE INVENTION

The present disclosure provides RNA-guided nucleases derived from RNA-guided Nuclease BT that exhibit advantageous enzymatic activities (*e.g.*, high indel activities and/or binding activities to the scaffold of a cognate guide RNA). Accordingly, the RNA-guided nucleases provided herein would be expected to show superior effectiveness when used in gene editing.

Accordingly, one aspect of the present disclosure provides an RNA-guided nuclease polypeptide, comprising a RuvC nuclease domain and an HNH nuclease domain. The RNA-guided nuclease polypeptide may comprise an amino acid sequence at least 90% identical to SEQ ID NO: 1, which refers to the reference RNA-guided Nuclease BT. In some embodiments, the RNA-guided nuclease polypeptide is a variant of Nuclease BT comprising at least one mutation relative to Nuclease BT (SEQ ID NO: 1). In some instances, the one or more mutation(s) may comprise one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) relative to the reference RNA-guided nuclease. Alternatively or in addition, the

one or more mutation(s) may comprise one or more nickase mutations in the HNH nuclease domain or in the RuvC nuclease domain of SEQ ID NO: 1.

The RNA-guided nuclease polypeptide may comprise a PLMP domain, a bridge helix (BH) domain, a phosphate lock loop (PLL) domain, a wedge (WED) domain, and a PAM-interacting (PID) domain. In some instances, the one or more arginine and/or lysine substitutions (*e.g.*, arginine substitutions) are located in the BH 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 E99, Q102, E105, T106, T113, E191, I206, N239, V256, I272, A277, E401, E432, and V433 SEQ ID NO: 1.

In some specific examples, the engineered 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 engineered RNA-guided nuclease polypeptide comprises arginine and/or lysine substitutions at Q102, I206, and V433 of SEQ ID NO: 1.

Any of the engineered RNA-guided nuclease polypeptides disclosed herein may contain up to 20 arginine and/or lysine substitutions (*e.g.*, arginine substitutions), for example, up to 15 arginine and/or lysine substitutions (*e.g.*, arginine substitutions). In some examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of:

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 examples, the engineered RNA-guided nuclease polypeptide contains the arginine substitutions of (a) Q102R, I206R, and V433R.

Alternatively or in addition, the RNA-guided nuclease polypeptide disclosed herein may be a nickase variant, which may comprise one or more nickase mutations with the RuvC nuclease domain or with the HNH nuclease domain. In some embodiments, the one or more nickase mutations may be at position H231, H232, H255, D67, E176, and/or D329 of SEQ ID NO: 1. In some examples, the nickase variant may comprise a mutation at position H232. In some specific examples, the mutation at position H232 is an amino acid substitution of H232A or H232L.

In some instances, the variant RNA-guided nuclease polypeptide disclosed herein may comprise (a) one or more nickase mutations in the HNH nuclease domain at positions H231, H232, and/or H255 relative to SEQ ID NO: 1 (*e.g.*, at position H232); and (b) one or more arginine and/or lysine substitutions at positions Q102, I206, and V433 relative to SEQ ID NO: 1.

In some specific examples, the variant RNA-guided nuclease polypeptide disclosed herein may comprise (a) the nickase mutations H232A; and (b) the arginine substitutions at positions Q102R, I206R, and V433R. Any of the RNA-guided nuclease polypeptides disclosed herein, including any of the variants of Nuclease BT, may comprise an amino acid sequence at least 95% identical to SEQ ID NO: 1. In some examples, the RNA-guided nuclease may comprise an amino acid sequence at least 98% identical to SEQ ID NO: 1.

In some instances, the RNA-guided nuclease polypeptide is listed in **Table 1, Table 4A**, or **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.

Any of the RNA-guided nuclease polypeptides disclosed herein may be a fusion polypeptide, which may further comprise one or more functional fragments. In some embodiments, the one or more functional fragments may comprise one or more nuclear localization signals (NLSs), one or more peptide linkers, or a combination thereof. The one or

more NLS may be located at the N-terminus, at the C-terminus, or both.

Also provided herein is a nucleic acid, comprising a nucleotide sequence encoding any of the RNA-guided nuclease polypeptides disclosed herein. In some instances, the nucleic acid is an expression vector (e.g., a viral vector), in which the nucleotide sequence encoding the RNA-guided nuclease polypeptide is in operable linkage to a promoter. Alternatively, the nucleic acid may be a messenger RNA (mRNA). Further, the present disclosure provides a host cell comprising the nucleic acid coding for the RNA-guided nuclease polypeptide as disclosed herein.

Still in another aspect, the present disclosure features a gene editing system, comprising: (a) any of the RNA-guided nuclease polypeptides disclosed herein or a first nucleic acid encoding the RNA-guided nuclease polypeptide; and (b) a guide RNA (gRNA) or a second nucleic acid encoding the gRNA, which comprises a scaffold sequence recognizable by the RNA-guided nuclease polypeptide and a spacer sequence specific to a target sequence in a genomic site of interest. The target sequence is adjacent to a protospacer adjacent motif (PAM).

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: 150).

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: 88-90 and 92-94. See **Table 5B** below.

In some instances, the RNA-guided nuclease polypeptide comprised in the gene editing system disclosed herein is the reference RNA-guided nuclease of SEQ ID NO: 1 (or a

counterpart thereof lacking the N-terminus M residue; SEQ ID NO: 110) and the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2. In other instances, the RNA-guided nuclease polypeptide in the gene editing system 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. Still in other instances, the RNA-guided nuclease polypeptide in the gene editing system 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 another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 88. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 89. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 90. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 91. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 92. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 93. In another example, the RNA-guided nuclease polypeptide in the gene editing system 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: 94.

In some embodiments, the target sequence is adjacent to (*e.g.*, 5' 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. A PAM can also be referred to as a target adjacent motif (TAM). The spacer sequence can be about 14-30 nt in length. In some specific examples, the spacer sequence is about 14-20 nt in length, 15-20 nt in length, or 16-22 nt in length, for example, 16 nt or 20 nt in length.

In some embodiments, the gene editing system disclosed herein may further comprise one or more lipid excipients associated with the RNA-guided nuclease polypeptide or an encoding nucleic acid and/or the guide RNA or its encoding nucleic acid of the gene editing system; optionally wherein the one or more lipid excipients form lipid nanoparticles, which are associated with or encapsulate the RNA-guided nuclease polypeptide or an encoding nucleic acid and/or the guide RNA or its encoding nucleic acid of the gene editing system. In some instances, such a gene editing system may comprise a messenger RNA molecule encoding the RNA-guided nuclease and the gRNA as disclosed herein. Alternatively, the gene editing system disclosed herein may comprise the first nucleic acid encoding the RNA-guided nuclease polypeptide. In some instances, the first nucleic acid may be located in a viral vector (*e.g.*, an adeno-associated viral (AAV) vector). In some examples, the viral vector may further comprise the second nucleic acid encoding the gRNA.

Further, provided herein is a gene editing method, comprising delivering any of the gene editing systems as disclosed herein to a host cell to edit a genomic site targeted by the gRNA of the gene editing system. In some instances, the host cell is cultured *in vitro*. In other instances, the host cell is located in a subject.

Still further, provided herein is a guide RNA that comprises a spacer sequence and a scaffold sequence having at least 70% (*e.g.*, at least 75%) sequence identity to SEQ ID NO: 2. The scaffold sequence is recognizable by any of the RNA-guided nuclease polypeptides disclosed herein. In some instances, the scaffold sequence is a variant of SEQ ID NO: 2 comprising at least one mutation (*e.g.*, deletion, insertion, or nucleotide substitution relative to SEQ ID NO: 2), *e.g.*, any of the mutations to the reference scaffold sequence of SEQ ID NO: 2 as disclosed herein. For example, the variant 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: 150).

In one example, the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2. In another example, the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 25. In yet another example, the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 26. In other examples, the scaffold sequence may comprise the nucleotide sequence of any one of SEQ ID NOs: 88-90 and 92-94.

In some embodiments, the spacer sequence is about 14-25 nt in length. In some specific examples, the spacer sequence is about 14-20 nt in length, 15-20 nt in length, or 16-22 nt in length. In one example, the spacer sequence is 16-nt in length. In another example, the spacer sequence is 20-nt in length.

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

Provided herein are RNA-guided nuclease polypeptides derived from the reference RNA-guided Nuclease BT of SEQ ID NO: 1. In some embodiments, the RNA-guided nuclease polypeptides provided herein may be variants of RNA-guided Nuclease BT, comprising at least one mutation relative to RNA-guided Nuclease BT. For example, a variant RNA-guided nuclease polypeptide may comprise one or more mutations (*e.g.*, arginine substitutions, lysine substitutions, or a combination thereof) relative to RNA-guided Nuclease BT. "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 residue.

Alternatively or in addition, a variant RNA-guided nuclease polypeptide may comprise one or more nickase mutations in either the RuvC nuclease domain or the HNH nuclease domain. Such nickase mutations may reduce or eliminate the nuclease activity of either the RuvC or the HNH nuclease domain, leading to a variant exhibiting nickase activity. As used herein, the term "nickase" refers to an enzyme that cuts one strand of a double-stranded DNA at a specific recognition nucleotide sequence (e.g., the target sequence disclosed herein). A nickase may interact with one strand of the DNA duplex to produce DNA molecules that are cut at one strand (*a.k.a.*, nicked). In some embodiments, a nickase is a variant of an RNA-guided nuclease that comprises a deactivated HNH domain. In some embodiments, a nickase is a variant of an RNA-guided nuclease that comprises a deactivated 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 90% sequence identity). The variant RNA-guided nuclease polypeptides provided herein are expected to possess advantageous features relative to the reference RNA-guided nuclease, for example, increased binding to a cognate guide RNA, higher nuclease activity, etc. As such, the variant RNA-guided nuclease polypeptides disclosed herein would be expected to exhibit better activities in

gene editing relative to the reference RNA-guided nuclease, *e.g.*, higher indel activity.

Any of the RNA-guided nuclease polypeptides provided herein may be fusion polypeptides comprising an RNA-guided nuclease moiety (*e.g.*, SEQ ID NO: 1 or a variant thereof) and one or more additional functional fragments such as those described herein. Such additional functional fragments may be heterologous to the nuclease moiety. In addition to the advantageous features noted above, the fusion polypeptides possess additional functions attributable to the functional fragment. In such fusion polypeptides, the RNA-guided nuclease polypeptide therein may have the native N-terminal M residue removed, *e.g.*, when a functional fragment is linked to the N-terminus of the nuclease polypeptide.

Accordingly, the present disclosure provides RNA-guided nuclease polypeptides derived from the reference RNA-guided nuclease of SEQ ID NO: 1 (Nuclease BT), gene editing systems comprising such, and gene editing methods using such.

I. RNA-Guided Nuclease Polypeptides

As used herein, the term "RNA-guided nuclease" refers to an effector that is capable of binding a nucleic acid, as mediated by an RNA guide, and introducing a single-stranded break or double-stranded break. In some embodiments, an RNA-guided nuclease is complexed with an RNA guide to form a ribonucleoprotein (RNP). 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.

The reference RNA-guided 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' or 5'-NRT-3' PAM 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, the reference RNA-guided nuclease of SEQ ID NO: 1 disclosed herein is smaller. Its cognate scaffold comprises a distinctive structure compared to the Cas9 nuclease scaffold. The distinctive scaffold is expected to allow for decreased size of some domains (*e.g.*, the REC domain) and thus contribute to the smaller RNA-guided nuclease size. Additionally, the cognate scaffold can be miniaturized. These features would be beneficial for delivery. Arginine and/or lysine substitutions (*e.g.*, arginine substitutions) can be introduced into the RNA-guided nuclease of SEQ ID NO: 1 to increase indel activity. Additionally, since the reference RNA-guided nuclease 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 the RNA-guided nuclease of SEQ ID NO: 1 are different than those of Cas9, allowing for additional gene targets to be edited. Also, unlike Cas9, the RNA-guided nuclease 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 the RNA-guided nuclease to its cognate scaffold.

(i) Variants of RNA-Guided Nuclease Polypeptide

The variant RNA-guided nuclease polypeptides provided herein are derived from the reference RNA-guided 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 the reference RNA-guided nuclease (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 the reference RNA-guided nuclease. As used herein, the term "activity" refers to a biological activity. In some embodiments, activity includes enzymatic activity, *e.g.*, catalytic ability of an effector. For example, activity can include nuclease activity. In some embodiments, activity includes binding activity, *e.g.*, binding of an effector (*e.g.*, 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).

In some embodiments, the variant RNA-guided nuclease polypeptides 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 the reference RNA-guided nuclease of SEQ ID NO: 1, *e.g.*, one or more amino acid residue substitutions, one or more deletions, one or more insertions, fusion, or a combination thereof. In some instances, the alterations may be introduced into the PLMP domain, 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;I) 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. 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 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. For example, the variant RNA-guided nuclease polypeptide may contain one or more arginine and/or lysine substitutions at one or more of the following positions in SEQ ID NO: 1: E99, Q102, E105, T106, T113, E191, I206, N239, I272, V256, A277, E401, E432, and V433. In some examples, the variant RNA-guided nuclease polypeptide may contain a combination of arginine and/or lysine substitutions relative to SEQ ID NO: 1. Non-limiting examples are provided below: 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 some specific examples, the mutations such as arginine and/or lysine substitutions are at positions Q102, I206, and V433 of SEQ ID NO: 1.

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 engineered 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, a variant CRISPR nuclease may be a dead nuclease (*e.g.*, for use in base editing) having mutations at one or more of these positions: 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.

Alternatively or in addition, the variant RNA-guided nuclease polypeptide provided herein may further comprise one or more nickase 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. Such 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 yet other examples, the one or more mutations may be within both the RuvC domain and the HNH domain to reduce or diminish the nuclease enzymatic activity.

In some embodiments, the variant RNA-guided nuclease polypeptide provided herein may comprise both arginine/lysine substitutions, e.g., at one or more positions provided herein, and nickase mutations. 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.

Any of the RNA-guided nuclease polypeptides provided herein, e.g., those containing one or more arginine and/or lysine substitutions and/or one or more nickase mutations, may share a sequence identity at least 90% (*e.g.*, 95%, 97%, 98%, 99%, 99.5%, or greater) with SEQ ID NO: 1. For instance, the RNA-guided nuclease polypeptide is listed in **Table 1, Table 4A**, or **Table 11**. In 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, 43, 210, 213, or 216. In one example, a variant RNA-guided nuclease polypeptide may comprise (*e.g.*, consist of) SEQ ID NO: 33 or SEQ ID NO: 141. In another example, a variant RNA-guided nuclease polypeptide may comprise (*e.g.*, consist of) SEQ ID NO: 35 or SEQ ID NO: 142.

In some embodiments, any of the RNA-guided nuclease polypeptides disclosed herein (*e.g.*, Nuclease BT or a variant thereof as provided herein) may be a fusion polypeptide comprising a 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 (*i.e.*, Nuclease BT). 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 a nuclear localization signal (NLS), a nuclear export signal (NES), or a combination thereof. In some examples, the fusion polypeptide may comprise an NLS, which may be located at either the N-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 examples, the fusion polypeptide may further comprise one or more peptide linkers, which may be used to connect the NLS and the RNA-guided nuclease moiety in the fusion polypeptide.

(ii) Preparation of RNA-guided Nuclease Polypeptides

The RNA-guided nuclease polypeptides as disclosed herein may be prepared by conventional methods or the methods disclosed herein. For example, the RNA-guided nuclease polypeptides 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 polypeptides thus prepared may be complexed with a gRNA.

The RNA-guided nuclease polypeptides can be also prepared by an *in vitro* coupled transcription-translation system and optionally complexes with gRNA. Bacteria that can be used for preparation of the RNA-guided nuclease polypeptides are not particularly limited as long as they can produce the RNA-guided nuclease polypeptides. Some nonlimiting examples of the bacteria include *E. coli* cells described herein.

Unless otherwise noted, all compositions and complexes and polypeptides provided herein are made in reference to the active level of that composition or complex or polypeptide, and are exclusive of impurities, for example, residual solvents or by-products, which may be

present in commercially available sources. Enzymatic component weights are based on total active protein. All percentages and ratios are calculated by weight unless otherwise indicated. All percentages and ratios are calculated based on the total composition unless otherwise indicated. In the exemplified composition, the enzymatic levels are expressed by pure enzyme by weight of the total composition and unless otherwise specified, the ingredients are expressed by weight of the total compositions.

(i) Vectors

The present disclosure provides vectors for expressing the RNA-guided nuclease polypeptides. In some embodiments, a vector disclosed herein includes a nucleotide sequence encoding the RNA-guided nuclease polypeptides. 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 polypeptides 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 polypeptides 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.

(ii) Methods of Expression

The present disclosure includes a method for protein expression, comprising translating the RNA-guided nuclease polypeptides described herein.

In some embodiments, a host cell described herein is used to express the RNA-guided nuclease polypeptides. 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 RNA-guided nuclease polypeptides. After expression, the host cells can be collected and RNA-guided nuclease polypeptides 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 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 (MA), 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 RNA-guided nuclease polypeptides (and optionally the gRNA in the gene editing system disclosed herein). Such a method may comprise providing a polyribonucleotide encoding the RNA-guided nuclease polypeptide to a host cell in a subject (*e.g.*, a human subject) wherein the polyribonucleotide encodes the RNA-guided nuclease polypeptide expressing the RNA-guided nuclease polypeptide from the cell.

II. Gene Editing System

In some aspects, the present disclosure provides gene editing systems with enhanced gene editing efficiencies. The gene editing system comprises any of the RNA-guided nuclease polypeptides disclosed herein (e.g., a variant of Nuclease BT as disclosed herein) or a nucleic encoding the RNA-guided nuclease and one or more guide RNAs (gRNAs) or nucleic acid(s) encoding the gRNAs.

(a) RNA-Guided Nuclease

In some embodiments, the gene editing system disclosed herein comprises an RNA-guided nuclease polypeptide as provided herein, *e.g.*, a variant RNA-guided nuclease polypeptide comprising one or more arginine and/or lysine substitutions (e.g., arginine substitutions) and/or one or more nickase mutations. See above disclosures. Such a protein

component may form a complex with the gRNA(s) in the same gene editing system. Alternatively, the gene editing system comprises a nucleic acid encoding the RNA-guided nuclease polypeptide. In some instances, the nucleic acid can be an expression vector (e.g., a viral vector) for producing the encoded nuclease polypeptide in host cells. In some instances, the expression vector may further comprise a coding sequence for producing one or more gRNAs of the gene editing system. In other examples, the nucleic acid can be a messenger RNA (mRNA) encoding the RNA-guided nuclease polypeptide.

(b) 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 an 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.

(i) 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 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 the target sequence disclosed herein, and the PAM sequence as described herein is present in the PAM-strand. The PAM motif can be located downstream to the target sequence (3' 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, about 14 nucleotides to about 25 nucleotides, about 20 nucleotides to about 25 nucleotides, or 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 about 14-20, 15-20, or 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).

(ii) Scaffold Sequence

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

GGGACAUGACCAGAUGCGAAAGUAACUGUCAUGUCCUCAUGCCAAUGGA CUAUGGUGAUGCACCUAGGGGUGUCGUUCCAGCUCCUAGCUCUGCGAAG GAGCAUUCGUUAUCGGAUGCGUUUGUACUUCAAGCCAUAGCCCAGGCCA GGGACGUAUUACUGCUCCUGCCCUAUAGGAGCAUA (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: 150). 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.

In some instances, the RNA guides for use in the gene editing systems provided herein may comprise one or more modifications such as those disclosed below. In specific examples, the RNA guides provided herein may comprise one or more 2'-O-methylated nucleotide residues and/or one or more phosphorothioate linkages. Such modifications may be located at the 5' end, the 3' end, or both of an RNA guide.

(c) Modification of Nucleic Acids

Any of the RNA components in a gene editing system as disclosed herein, e.g., the gRNA, 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.

The gRNA or any of the nucleic acid sequences encoding components of the composition may include any useful modification, such as to the sugar, the nucleobase, or the internucleoside linkage (*e.g.*, to a linking phosphate/to a phosphodiester linkage/to the phosphodiester backbone). One or more atoms of a pyrimidine nucleobase may be replaced or substituted with optionally substituted amino, optionally substituted thiol, optionally substituted alkyl (*e.g.*, methyl or ethyl), or halo (*e.g.*, chloro or fluoro). One or more atoms of a purine nucleobase may be replaced or substituted with optionally substituted amino, optionally substituted thiol, optionally substituted alkyl (*e.g.*, methyl or ethyl), or halo (*e.g.*, chloro or fluoro). In certain embodiments, modifications (*e.g.*, one or more modifications) are present in each of the sugar and the internucleoside linkage. 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). An abasic site can also be referred to as an apurinic/apyrimidinic site. 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, the modification may include a chemical or cellular induced modification. For example, some nonlimiting examples of intracellular RNA modifications are described by Lewis and Pan in "RNA modifications and structures cooperate to guide RNA-protein interactions" from Nat Reviews Mol Cell Biol, 2017, 18:202-210.

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

50% to 100%, from 70% to 80%, from 70% to 90%, from 70% to 95%, from 70% to 100%, from 80% to 90%, from 80% to 95%, from 80% to 100%, from 90% to 95%, from 90% to 100%, and from 95% to 100%).

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

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

The modified nucleotides, which may be incorporated into the sequence, can be modified on the internucleoside linkage (*e.g.*, phosphate backbone). Herein, in the context of the polynucleotide backbone, the phrases "phosphate" and "phosphodiester" are used interchangeably. Backbone phosphate groups can be modified by replacing one or more of the oxygen atoms with a different substituent. Further, the modified nucleosides and nucleotides can include the wholesale replacement of an unmodified phosphate moiety with another internucleoside linkage as described herein. Examples of modified phosphate groups include, but are not limited to, phosphorothioate, phosphoroselenates, boranophosphates, boranophosphate esters, hydrogen phosphonates, phosphoramidates, phosphorodiamidates, alkyl or aryl phosphonates, and phosphotriesters. Phosphorodithioates have both non-linking

oxygens replaced by sulfur. The phosphate linker can also be modified by the replacement of a linking oxygen with nitrogen (bridged phosphoramidates), sulfur (bridged phosphorothioates), and carbon (bridged methylene-phosphonates).

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

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

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

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

In some embodiments, the sequence includes one or more post-transcriptional modifications (*e.g.*, capping, cleavage, polyadenylation, splicing, poly-A sequence, methylation, acylation, phosphorylation, methylation of lysine and arginine residues, acetylation, and nitrosylation of thiol groups and tyrosine residues, etc.). The one or more post-transcriptional modifications can be any post-transcriptional modification, such as any of the more than one hundred different nucleoside modifications that have been identified in RNA (Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197) In some embodiments, the first 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.

III. Gene 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.

The gene editing systems and compositions disclosed herein are applicable for editing and introducing edits into a variety of target sequences. In some embodiments, the target sequence is a DNA molecule, such as a DNA locus (referred to herein as a target sequence or an on-target sequence). The target sequence is adjacent to the PAM motif of 5'-RRT-3' or 5'-NRT-3', in which R represents A or G and N represents any nucleotide. In some instances, the PAM motif is 3' to the target sequence. In some embodiments, the target nucleic acid is a genomic site in a cell. In some instances, the target nucleic acid where the genetic edit would occur can be in a protein-coding region. Alternatively, the target nucleic acid may be in a regulatory region, such as a promoter, enhancer, a 5' or 3' untranslated region. In other instances, the target nucleic acid can be in a non-coding gene, such as transposon, miRNA, tRNA, ribosomal RNA, ribozyme, or lincRNA.

A. Gene Edits

Any of the gene editing systems disclosed herein may be used to edit a target gene of

interest, *e.g.*, a gene involved in a disease (*e.g.*, a genetic disease). In some embodiments, the target gene can be one that is involved in an immune response in a subject. For example, the target gene can be an immune checkpoint gene or a tumor necrosis factor receptor superfamily member. The gene edit may occur in an exon (*e.g.*, in a coding region). Alternatively, the gene editing may occur in an intron or in a regulatory element (*e.g.*, promoter, enhancer, inhibitory element, etc.). In some instances, the gene edit may result in reducing or eliminating the expression of the target gene. In other instances, the gene edit may result in enhancing expression of the target gene (*e.g.*, disrupting an inhibitory factor).

In some aspects, provided herein are methods for introducing at least one edit into a target nucleic acid (e.g., a genomic site of interest such as in any of the target genes disclosed herein) using the gene editing system described herein.

As used herein, the term "edit" refers to one or more modifications introduced into a nucleotide sequence in a target nucleic acid such as in a genomic site of interest. The edit may occur within a target sequence as defined herein. Alternatively, the edit may occur outside the target sequence (e.g., adjacent to the target sequence). The edit can be one or more substitutions, one or more insertions, one or more deletions, or a combination thereof.

Deletion refers to a loss of a nucleotide or nucleotides in a nucleic acid sequence, relative to a reference sequence. No particular process is implied in how to make a sequence comprising a deletion. For instance, a sequence comprising a deletion can be synthesized directly from individual nucleotides. In other embodiments, a deletion is made by providing and then altering a reference sequence. The nucleic acid sequence can be in a genome of an organism. The nucleic acid sequence can be in a cell. The nucleic acid sequence can be a DNA sequence. The deletion can be a frameshift mutation or a non-frameshift mutation. A deletion described herein refers to an insertion of up to several kilobases.

Insertion refers to a gain of a nucleotide or nucleotides in a nucleic acid sequence, relative to a reference sequence. No particular process is implied in how to make a sequence comprising an insertion. For instance, a sequence comprising an insertion can be synthesized directly from individual nucleotides. In other embodiments, an insertion is made by providing and then altering a reference sequence. The nucleic acid sequence can be in a genome of an organism. The nucleic acid sequence can be in a cell. The nucleic acid sequence can be a DNA sequence. The insertion can be a frameshift mutation or a non-frameshift mutation. An insertion described herein refers to an insertion of up to several kilobases.

In some embodiments, the gene editing methods disclosed herein may introduce edits, including a substitution, an insertion, a deletion, or a combination thereof, into the target

nucleic acid.

In some examples, the edits can include at least one substitution, at least one insertion, and/or at least one deletion. In some embodiments, the edit comprises at least one substitution, insertion, or deletion. In some embodiments, the substitution, insertion, or deletion is at least 1-500 nucleotides (*e.g.*, 1-10 nucleotides, 10-30 nucleotides, 30-50 nucleotides, 50-100 nucleotides, 100-200 nucleotides, 200-300 nucleotides, 300-400 nucleotides, or 400-500 nucleotides).

In some examples, the edit may occur within about 500 nucleotides of a 5'-RRT-3' or 5'-NRT-3' PAM sequence, in which R is either A or G and N is any nucleotide. In some embodiments, the edit occurs adjacent to the PAM sequence, *e.g.*, within about 1-500 nucleotides upstream or downstream of the PAM sequence. In some embodiments, the edit may occur within about 1-10 nucleotides, 10-30 nucleotides, 30-50 nucleotides, 50-100 nucleotides, 100-200 nucleotides, 200-300 nucleotides, 300-400 nucleotides, or 400-500 nucleotides upstream of the PAM sequence. Alternatively, or in addition, the edit may occur within about 1-10 nucleotides, 10-30 nucleotides, 30-50 nucleotides, 50-100 nucleotides, 100-200 nucleotides, 200-300 nucleotides, 300-400 nucleotides, or 400-500 nucleotides downstream of the PAM sequence.

In some embodiments, the edit starts at the PAM sequence. In some embodiments, the edit may start within about 1-30 nucleotide downstream of the PAM. Alternatively, the edits may start within about 1-30 nucleotide upstream of the PAM.

In some embodiments, the edit may end within about 1-300 nucleotides upstream of the PAM sequence, for example, ends within about 1-10 nucleotides, about 10-30 nucleotides, about 30-50 nucleotides, about 50-100 nucleotides, about 100-200 nucleotides, or about 200-300 nucleotides upstream of the PAM sequence. Alternatively, the edit may end within about 1-300 nucleotides downstream of the PAM sequence, for example, ends within about 1-10 nucleotides, about 10-30 nucleotides, about 30-50 nucleotides, about 50-100 nucleotides, about 100-200 nucleotides, or about 200-300 nucleotides downstream of the PAM sequence.

In some embodiments, the edit may end at the PAM sequence. In some embodiments, the edit ends within about 1-30 nucleotide downstream of the PAM. In other embodiments, the edit may end within about 1-30 nucleotide upstream of the PAM.

B. Gene Editing in Cells

In some aspects, provided herein are methods for editing a genomic site of interest (*e.g.*, a target gene as disclosed herein) in cells using any of the gene editing systems disclosed herein. To perform this method, the gene editing system can be delivered to or introduced into

a population of cells. In some instances, cells comprising the desired genetic editing may be collected and optionally cultured and expanded *in vitro*.

The cell described herein can be a variety of cells. In some embodiments, the cell is an isolated 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 a single-cellular organism.

In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a bacterial cell or derived from a bacterial cell. In some embodiments, the cell is an archaeal cell or derived from an archaeal cell.

In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a plant cell or derived from a plant cell. In some embodiments, the cell is a fungal cell or derived from a fungal cell. In some embodiments, the cell is an animal cell or derived from an animal cell. In some embodiments, the cell is an invertebrate cell or derived from an invertebrate cell. In some embodiments, the cell is a vertebrate cell or derived from a vertebrate cell. In some embodiments, the cell is a mammalian cell or derived from a mammalian cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is a zebra fish cell. In some embodiments, the cell is a primate cell. In some embodiments, the cell is a rodent cell. In some embodiments, the cell is synthetically made, sometimes termed an artificial cell.

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, HEK293T, 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 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 mesenchymal stem cell. In some embodiments, the cell is an embryonic stem cell. In some embodiments, the cell is a hematopoietic stem cell. In some embodiments, the cell is a differentiated cell. For example, in some embodiments, the differentiated cell is a muscle cell (*e.g.*, a myocyte), a fat cell (*e.g.*, an adipocyte), a bone cell (*e.g.*, an osteoblast, osteocyte, osteoclast), a blood cell (*e.g.*, a monocyte, a lymphocyte, a neutrophil, an eosinophil, a basophil, a macrophage, a erythrocyte, or a platelet), a nerve cell (*e.g.*, a neuron), an epithelial cell, an immune cell (*e.g.*, a

lymphocyte, a neutrophil, a monocyte, or a macrophage), a liver cell (*e.g.*, a hepatocyte), a fibroblast, or a sex cell. In some embodiments, the cell is a terminally differentiated cell. For example, in some embodiments, the terminally differentiated cell is a neuronal cell, an adipocyte, a cardiomyocyte, a skeletal muscle cell, an epidermal cell, or a gut cell. In some embodiments, the cell is a glial cell. In some embodiments, the cell is a pancreatic islet cell, including an alpha cell, beta cell, delta cell, or enterochromaffin cell. In some embodiments, the cell is an immune cell. In some embodiments, the immune cell is a T cell. In some embodiments, the immune cell is a B cell. In some embodiments, the immune cell is a Natural Killer (NK) cell. In some embodiments, the immune cell is a Tumor Infiltrating Lymphocyte (TIL). In some embodiments, the cell is a mammalian cell, *e.g.*, a human cell or primate 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.

In some embodiments, the cell is a primary cell. For example, cultures of primary cells can be passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, 15 times or more. In some embodiments, the primary cells are harvest from an individual by any known method. For example, leukocytes may be harvested by apheresis, leukocytapheresis, density gradient separation, etc. Cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. can be harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution can generally be a balanced salt solution, (*e.g.*, normal saline, phosphate-buffered saline (PBS), Hank's balanced salt solution, etc.), conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration. Buffers can include HEPES, phosphate buffers, lactate buffers, etc. Cells may be used immediately, or they may be stored (*e.g.*, by freezing). Frozen cells can be thawed and can be capable of being reused. Cells can be frozen in a DMSO, serum, medium buffer (*e.g.*, 10% DMSO, 50% serum, 40% buffered medium), and/or some other such common solution used to preserve cells at freezing temperatures.

In embodiments wherein a gene editing system disclosed herein is introduced into a plurality of cells, at least about 0.5% of the cells comprise the desired edit. In some embodiments, at least about 1% of the cells comprise the desired edit. In some embodiments, at least about 2% of the cells comprise the desired edit. In some embodiments, at least about 3% of the cells comprise the desired edit. In some embodiments, at least about 4% of the cells comprise the desired edit. In some embodiments, at least about 5% of the cells comprise the

desired edit. In some embodiments, at least about 10% of the cells comprise the desired edit. In some embodiments, at least about 20% of the cells comprise the desired edit. In some embodiments, at least about 30% of the cells comprise the desired edit. In some embodiments, at least about 40% of the cells comprise the desired edit. In some embodiments, at least about 50% of the cells comprise the desired edit.

The cells carrying the desired genetic edit, *e.g.*, produced by the method disclosed herein using any of the gene editing systems also disclosed herein, are also within the scope of the present disclosure. In some instances, the cells modified by the RNA-guided nuclease polypeptide disclosed herein may be useful as an expression system to manufacture biomolecules. For example, the modified cells may be useful to produce biomolecules such as proteins (*e.g.*, cytokines, antibodies, antibody-based molecules), peptides, lipids, carbohydrates, nucleic acids, amino acids, and vitamins. In other embodiments, the modified cell may be useful in the production of a viral vector such as a lentivirus, adenovirus, adeno-associated virus, and oncolytic virus vector. In some embodiments, the modified cell may be useful in cytotoxicity studies. In some embodiments, the modified cell may be useful as a disease model. In some embodiments, the modified cell may be useful in vaccine production. In some embodiments, the modified cell may be useful in therapeutics. For example, in some embodiments, the modified cell may be useful in cellular therapies such as transfusions and transplantations.

In some embodiments, the cells modified by the RNA-guided nuclease polypeptide as disclosed herein may be useful to establish a new cell line comprising a modified genomic sequence. In some embodiments, a modified cell of the disclosure is a modified stem cell (*e.g.*, a modified totipotent/omnipotent stem cell, a modified pluripotent stem cell, a modified multipotent stem cell, a modified oligopotent stem cell, or a modified unipotent stem cell) that differentiates into one or more cell lineages comprising the deletion of the modified stem cell. The disclosure further provides organisms (such as animals, plants, or fungi) comprising or produced from a modified cell of the disclosure.

C. Delivery of Gene Editing Systems to Cells

In some embodiments, any of the gene editing systems or components thereof as disclosed herein may be formulated, for example, including a carrier, such as a carrier and/or a polymeric carrier, *e.g.*, a liposome or lipid nanoparticle, and delivered by known methods to a cell (*e.g.*, a prokaryotic, eukaryotic, plant, mammalian, etc.). 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, 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 embodiments, the method comprises delivering one or more nucleic acids (*e.g.*, nucleic acids encoding the RNA-guided nuclease polypeptide and/or the gRNAs, and/or a pre-formed ribonucleoprotein to a cell. Exemplary intracellular delivery methods, include, but are not limited to: viruses 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, magnetofection or magnet assisted transfection, particle bombardment; and hybrid methods, such as nucleofection. 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 embodiments, a composition of the present invention is further delivered with an agent (*e.g.*, compound, molecule, or biomolecule) that affects DNA repair or DNA repair machinery. In some embodiments, a composition of the present invention is further delivered with an agent (*e.g.*, compound, molecule, or biomolecule) that affects the cell cycle.

In some embodiments, a first composition comprising an RNA-guided nuclease polypeptide is delivered to a cell. In some embodiments, a second composition comprising a gRNA is delivered to the cell. In some embodiments, the first composition is contacted with a cell before the second composition is contacted with the cell. In some embodiments, the first composition is contacted with a cell at the same time as the second composition is contacted with the cell. In some embodiments, the first composition is contacted with a cell after the second composition is contacted with the cell. In some embodiments, the first composition is delivered by a first delivery method and the second composition is delivered by a second delivery method. In some embodiments, the first delivery method is the same as the second delivery method. For example, in some embodiments, the first composition and the second composition are delivered via viral delivery. In some embodiments, the first delivery method is different than the second delivery method. For example, the first composition is delivered by viral delivery and the second composition is delivered by lipid nanoparticle-mediated transfer, or the first composition is delivered by lipid nanoparticle-mediated transfer and the second

composition is delivered by viral delivery. In some embodiments, the first delivery method is the same as the second delivery method. For example, both the first and second compositions are delivered by a viral vector (*e.g.*, an adeno-associated viral vector (AAV) or an adenoviral vector (Adv)).

In some examples, the nucleic acid encoding an RNA-guided nuclease polypeptide and the nucleic acid encoding a gRNA can be carried by a single vector, for example, a viral vector (*e.g.*, an AAV or Adv) 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 components as disclosed herein (*e.g.*, a mRNA encoding the RNA-guided nuclease polypeptide and a gRNA) may be delivered to host cells via lipid nanoparticles (LNPs).

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 may be benefit from the gene edit introduced by the gene editing system or carried by the modified cells. For example, the disease may be a genetic disease and the gene edit fixes the gene mutation associated with the genetic disease. Alternatively, the disease may be associated with abnormal expression of a gene and the gene edit rescues such abnormal expression.

In some embodiments, provided herein is a method for treating a disease comprising administering to a subject (*e.g.*, a human patient) in need of the treatment any of the gene editing system disclosed herein. The gene editing system may be delivered to a specific tissue or specific type of cells where the gene edit is needed. A composition comprising the gene editing system may further comprise LNPs encompassing one or more of the components, or one or more vectors (*e.g.*, viral vectors such as AAV 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 and the gRNA as disclosed herein. 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 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 or the modified cells. 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 comprising a predetermined number of cells. The number of cells is generally equal to the dosage of the cells 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 or systems that can be used, for example, to carry out a method described herein. In some embodiments, the kits or systems include an RNA-guided nuclease polypeptide and optionally a gRNA. In some embodiments, the kits or systems include a polynucleotide that encodes the RNA-guided nuclease polypeptide and optionally the gRNA. The gRNA of the kits can be designed to target a sequence of interest. The RNA-guided nuclease polypeptide and the gRNA can be packaged within the same vial or other vessel within a kit or system or can be packaged in separate vials or other vessels, the contents of which can be mixed prior to use. The kits or systems can additionally include, optionally, a buffer and/or instructions for use of the RNA-guided nuclease polypeptide and the gRNA.

In some embodiments, the kit comprises a first composition comprising an RNA-guided nuclease polypeptide as disclosed herein. In some embodiments, the kit comprises a second composition comprising a gRNA as also disclosed herein. In some embodiments, the first composition and the second composition are packaged within the same vial. In some embodiments, the first composition and the second composition are packaged within different vials.

In some embodiments, the kit may be useful for research purposes. For example, in some embodiments, the kit may be useful to study gene function.

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 (from N-terminus to C-terminus): SV40 NLS – Nuclease – Linker – nucleoplasmin NLS. When fused to an N-terminal sequences (e.g.: NLS) the native N-terminal methionine is removed from the reference nuclease. The reference (Nuclease BT), NLS, and linker sequences used in exemplary nuclease polypeptides are shown in **Table 1** below.

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	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: 110)*

Component	Nucleotide Sequence (Codon Optimized for Human Cells)	Amino Acid sequence
	ggtcaaggacaagaagcgcgtgcgggtggccacca acagaaatggcaatcctgatgccttccacgaggtgta cgacacaaagatcctgcctctgctgaagcagggctat cggctgaaagtgaaccctggcgttaagggctacagg aagaagttcagcgaggtgggatttaaggaaggcgcc acaatcagaatcggaaatcgggtggacgtgctgaag gcacctccggcttcggctacacctacgacggcacc atcttcaagctcaaggacatggacgtggtgtacccta cggcggctggaagtgcgtgaaggtg (SEQ ID NO: 3)	
SV40 NLS	aaacggacagccgacggaagcgagttcgagtcacc aaagaagaagcggaaagtc (SEQ ID NO: 4)	KRTADGSEFESPKKKRKV (SEQ ID NO: 5)
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	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
	aagacacactgcaacgacgcttactgcatcgccatca acaagctggtgagatttgacaccgacctggaacacg aaagcgtggtggataagcaagtgatccggttcgagc agaaaagacagcacaaccggagaatgaaatctctgg agagaaacaggctgtactacgtggtcaaggacaaga agcgcgtgcgggtggccaccaacagaaatggcaat cctgatgccttccacgaggtgtacgacacaaagatcc tgcctctgctgaagcagggctatcggctgaaagtgaa ccctggcgttaagggctacaggaagaagttcagcga ggtgggatttaaggaaggcgccacaatcagaatcgg aaatcgggtggacgtgctgaagggcacctccggctt cggctacacctacgacggcaccatcttcaagctcaag 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	GATAGCATAGGTACACATTA (SEQ ID NO: 16)	GAUAGCAUAGGUACACAUUAgggacaugaccagaugcgaaaguaacugucauguccucaugccaauggacuaugugaugcaccuaggggugucguuccagcuccuagcucugcgaaggagcauucguuaucggaugcguuuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 17)
EMX1-T7	GAT	TTAGCAAGGGACTATTCAGG (SEQ ID NO: 18)	UUAGCAAGGGACUAUUCAGGgggacaugaccagaugcgaaaguaacugucauguccucaugccaauggacuaugugaugcaccuaggggugucguuccagcuccuagcucugcgaaggagcauucguuaucggaugcguuuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 19)
VEGFA-T3	AAT	AGCCGATTCCAGGTGTGGAC (SEQ ID NO: 20)	AGCCGAUUCCAGGUGUGGACgggacaugaccagaugcgaaguaacugucauguccucaugccaauggacuauggugaugcaccuaggggugucguuccagcuccuagcucugcgaaggagcauucguuaucggaugcguuuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauaggagcaua (SEQ ID NO: 21)
VEGFA-T6	AGT	TGGGGACCAAGGAAGTGTCA (SEQ ID NO: 22)	UGGGGACCAAGGAAGUGUCAgggacaugaccagaugcgaaaguaacugucauguccucaugccaauggacuaugugaugcaccuaggggugucguuccagcuccuagcucugcgaaggagcauucguuaucggaugcguuuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauaggagcaua (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 37°C for approximately 5 minutes. D10 media was then added to each well and mixed to resuspend cells. The resuspended cells were centrifuged for

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

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

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: E99R, Q102R, E105R, T106R, T113R, E191R, I206R, N239R, V256R, I272R, A277R E401R, E432R, and V433R.

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	152 *With N-terminus M residue (SEQ ID NO: 153)*
E105R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY	154 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 155)*
Q102R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	156 *With N-terminus M residue (SEQ ID NO: 157)*
Q102R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	158 *With N-terminus M residue (SEQ ID NO: 159)*
Q102R, E191R, I206R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	160 *With N-terminus M residue (SEQ ID NO: 161)*
Q102R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY	162 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 163)*
E105R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	164 *With N-terminus M residue (SEQ ID NO: 165)*
E105R, E191R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	166 *With N-terminus M residue (SEQ ID NO: 167)*
E105R, E191R, I272R, E401R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHRVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	168 *With N-terminus M residue (SEQ ID NO: 169)*
E105R, E191R, N239R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY	170 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQRGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 171)*
E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	172 *With N-terminus M residue (SEQ ID NO: 173)*
E105R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	174 *With N-terminus M residue (SEQ ID NO: 175)*
E105R, T106R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRRRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	176 *With N-terminus M residue (SEQ ID NO: 177)*
E105R, E191R, I272R, E432R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY	178 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSRRGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 179)*
Q102R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	180 *With N-terminus M residue (SEQ ID NO: 181)*
E105R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	182 *With N-terminus M residue (SEQ ID NO: 183)*
I206R, I272R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	184 *With N-terminus M residue (SEQ ID NO: 185)*
E105R, E191R, I272R, A277R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY	186 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIRDFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 187)*
Q102R, E105R, E191R, I206R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	188 *With N-terminus M residue (SEQ ID NO: 189)*
Q102R, E105R, E191R, I206R, V256R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	190 *With N-terminus M residue (SEQ ID NO: 191)*
I206R, V256R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	192 *With N-terminus M residue (SEQ ID NO: 193)*
Q102R, E105R, E191R, I206R,	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY	76194 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
I272R, V433R	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 195)*
E105R, T113R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLRTRRMYRCRRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	196 *With N-terminus M residue (SEQ ID NO: 197)*
I272R, V256R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKQLLETRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSEVGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	198 *With N-terminus M residue (SEQ ID NO: 199)*
Q102R, E105R, E191R, I206R, V256R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNRRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	200 *With N-terminus M residue (SEQ ID NO: 201)*
Q102R, E105R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFGYQNIRAY	202 *With N-terminus M residue*

Variants	Amino Acid Sequences	SEQ ID NO
	VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	*(SEQ ID NO: 203)*
Q102R, E105R, E191R, V256R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHRKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	204 *With N-terminus M residue (SEQ ID NO: 205)*
Q102R, E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKEIKRLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	206 *With N-terminus M residue (SEQ ID NO: 207)*
E99R, E105R, E191R, I272R, V433R	RHQRLVFLVDGRNNILHPTRKLDMVRRWLRQTNAKVIRFIG KIMVVQVFKILSQQKDDPTYWIGVDPGDTIGICLVRITANNR VKRLLSMEFHTRSKRIKQLLRTRRMYRCTRRYFRRKRVKKK GRIPKHRPARFENRRRSDGWLTPSVRHFLESHKSVIRKIAECV PDVNISLEYAKFDIQKMMNPDIRGEQYQRGKLFGYQNIRAY VLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDNLITL CPECHVKVHTDTKLLKKLQRKRKSKIADFTSTSRLNIIMSFLK HWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNR LYYVVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGY RLKVNPGVKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGGWKCVKV	208 *With N-terminus M residue (SEQ ID NO: 209)*
Nuclease BT	See Table 1 above	1 or 110

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
I206R, V256R	1.03	1.60	2.12	1.58
Q102R, E105R, E191R, I206R, I272R, V433R	0.72	2.56	1.44	1.57

Variants	AAVS1 Average Indel Percentage	EMX1 Average Indel Percentage	VEGFA Average Indel Percentage	Average Indel Percentage (All Targets)
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	152
NLS-tagged variant	MKRTADGSEFESPKKKRKVRHQRLVFLVDGRNNILHPTRKLDMV RRWLRQTNAKVIRFIGKIMVVQVFKILSQQKDDPTYWIGVDPGD TIGICLVRITANNRVKRLLSMEFHTRSKEIKRLLETRRMYRCTR RYFRRKRVKKKGRIPKHRPARFENRRRSDGWLTPSVRHFLESHK SVIRKIAECVPDVNISLEYAKFDIQKMMNPDIEGEQYQRGKLFG YQNRRAYVLHRDKYTCQLCAKKNVPFQLHHVVPRSQNGKDHPDN LITLCPECHVKVHTDTKLLKKLQRKIKSKIADFTSTSRLNIIMS FLKHWLENEFGKDRVWFCIGADSYVIRKELNLDKTHCNDAYCIA INKLVRFDTDLEHESVVDKQVIRFEQKRQHNRRMKSLERNRLYY VVKDKKRVRVATNRNGNPDAFHEVYDTKILPLLKQGYRLKVNPG VKGYRKKFSERGFKEGATIRIGNRVDVLKGTSGFGYTYDGTIFK LKDMDVVYPYGGWKCVKVSGGSSGGSSGSETPGTSESATPESSG GSSGGSSKRPAATKKAGQAKKKK	109

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: 84)
16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCA (SEQ ID NO: 24)
20-nt AAVS1-T6 spacer	GAGCGGGAGAGGCUACAAAG (SEQ ID NO: 85)
16-nt AAVS1-T6 spacer	GGGAGAGGCUACAAAG (SEQ ID NO:86)
16-nt EMX1-T7 spacer	CAAGGGACUAUUCAGG (SEQ ID NO: 87)
Reference Scaffold	SEQ ID NO: 2
Scaffold 1	gggacaugaccaggaaacugucauguccucaugccaauggacuauggugaugcacc uaggggugucguuccagcuccuagcucugcgaaggagcauucggaaacggaugcgu uuguacuucaagccauagcccaggccagggacguauuacugcuccugcccuauagga gc (SEQ ID NO: 25)
Scaffold 1a	gggacaugaccaggaaacugucauguccucaugccaauggacuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 88)
Scaffold 2	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 26)
Scaffold 3	gggacaugaccaggaaacugucauguccucauggcaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccagc ccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 89)
Scaffold 4	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 90)
Scaffold 5	gggacaugaccaggaaacugucauguccucaugccaaugggcuauggugaugcacc uaggggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccag gccagggacgua (SEQ ID NO: 91)
Scaffold 6	gggacaugacggaaacgucauguccucaugccaaugggcuauggugaugcaccuag gggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccaggcca gggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 92)
Scaffold 7	gggacaugacggaaacgucauguccucauggcaaugggcuauggugaugcaccuag gggugucguuccagcuccuagcucugcgccagaaauggcaagccauagcccagccca gggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 93)
Scaffold 8	gggacaugacggaaacgucauguccucauggcaaugggcuauggugaugcaccuag cgguucguuccagcugcuagcucugcgccagaaauggcaagccauagcccagccca gggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 94)

5

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: 95)
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: 96)
Scaffold 4 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 97)
Scaffold 5 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugaccaggaaacugucauguccucaug ccaaugggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgcca gaaauggcaagccauagcccaggccagggacgua (SEQ ID NO: 98)
Scaffold 6 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucaugcca augggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgccagaa auggcaagccauagcccaggccagggacguauuacugcuccugcccuauaggagc (SEQ ID NO: 99)
Scaffold 7 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucauggca augggcuauggugaugcaccuaggggugucguuccagcuccuagcucugcgccagaa auggcaagccauagcccagcccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 100)
Scaffold 8 with 16-nt VEGFA-T6 spacer	GACCAAGGAAGUGUCAgggacaugacggaaacgucauguccucauggca augggcuauggugaugcaccuagcgguucguuccagcugcuagcucugcgccagaa auggcaagccauagcccagcccagggacguauuacgcuccugcccuauaggagc (SEQ ID NO: 101)

* *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: 151), 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 RNA-Guided Nickase Variants 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 aaacggtaaggaccacccggacaaccttatcacgctgt gcccggaatgccatgttaaggttcataccgataccaaact cttgaagaaactgcagcggaaaattaaaagcaaaatcgc tgattttacctccacctctcgtctgaacatcattatgagcttt ctgaagcattggctggagaacgaattcggcaaagaccg tgtgtggttctgcataggcgcggatagctatgtgatccgc	SEQ ID NO: 1 or SEQ ID NO: 110 (with N-terminal M residue removed) See **Table 1** above.

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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 ggttgaccccaagcgtgcgccactttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg gcaaactgtttggctaccaaaacattcgtgcgtacgttctg caccgcgataagtatacctgtcagctgtgcgcgaaaaaa aatgttccgtttcagttg**GCG**cacgtggtcccgcgttc gcaaaacggtaaggaccacccggacaaccttatcacgc tgtgcccggaatgccatgttaaggttcataccgataccaa actcttgaagaaactgcagcggaaaattaaaagcaaaat cgctgattttacctccacctctcgtctgaacatcattatgag ctttctgaagcattggctggagaacgaattcggcaaaga ccgtgtgtggttctgcataggcgcggatagctatgtgatc cgcaaggagctgaatctcgataagacccattgtaatgat gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QL**A**HVVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 31) *With no N-terminus M residue (SEQ ID NO: 140)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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)	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLH**A**VVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 33) *With no N-terminus M residue (SEQ ID NO: 141)*
H232L (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccactttttggagagccata 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)	YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLH**L**VVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 35) *With no N-terminus M residue (SEQ ID NO: 142)*
H255A (putative HNH-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccactttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttggagtatgctaaattcgacatccagaaaa tgatgaacccggatattgaaggtgaacaatatcaacgtg	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLHHVVPRSQNGKDHPDNLITLC PEC**A**VKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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)	LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 37) *With no N-terminus M residue (SEQ ID NO: 143)*
D67A (putative RuvC-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt g**GCG**ccgggcgatacgatcggcatttgcctggttcgt atcaccgcaaataaccgcgtcaagcgcttgctgtccatg gaattccacaccagaagcaaagagatcaaacaactgtta gagacacgccgtatgtatcgttgtacccgtcgttatttccg ccggaagcgcgtgaagaagaagggcagaattccgaaa caccgtccggcacgtttcgaaaaccgtcgccgctctgat ggttggttgaccccaagcgtgcgccacttttttggagagc cataaatcggtcattcgtaagatcgccgaatgtgttccgg acgttaatatcagcttggagtatgctaaattcgacatccag aaaatgatgaacccggatattgaaggtgaacaatatcaa cgtggcaaactgtttggctaccaaaacattcgtgcgtacg ttctgcaccgcgataagtatacctgtcagctgtgcgcgaa aaaaaatgttccgtttcagttgcaccacgtggtcccgcgtt cgcaaaacggtaaggaccacccggacaaccttatcacg ctgtgcccggaatgccatgttaaggttcataccgatacca aactcttgaagaaactgcagcggaaaattaaaagcaaaa tcgctgattttacctccacctctcgtctgaacatcattatga gctttctgaagcattggctggagaacgaattcggcaaag accgtgtgtggttctgcataggcgcggatagctatgtgat ccgcaaggagctgaatctcgataagacccattgtaatgat	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGV**A**P GDTIGICLVRITANNRVVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLHHVVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 39) *With no N-terminus M residue (SEQ ID NO: 144)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	gcgtactgcattgcgattaataaattggttcgtttcgatact gatctggagcatgaatctgttgtggataagcaagtgatcc gtttcgagcaaaagcgccagcataatagacgtatgaaat ccctggaacgtaaccgcctgtattacgtagtaaaagaca agaagcgtgtgcgcgtggcgacgaaccgcaacggtaa ccccgacgcctttcacgaggtttacgacaccaagatcct gccgctgctgaaacagggttaccgtttgaaggtaaatcc aggtgtgaagggttaccgtaagaagttcagcgaagttgg tttcaaagagggtgctaccatccgcatcggcaaccgtgt ggacgtgctgaagggtacttcaggttttggttacacgtac gacggtaccatttttaagcttaaggacatggatgtggtgta tccgtatggcggttggaaatgcgttaaggtg (SEQ ID NO: 40)	
E176A (putative RuvC-knockout)	atgagacaccaaaggctagtatttttagttgacggccgta acaacattctgcacccgacccgtaaattagacatggttcg acgttggctgcgtcagaccaatgcaaaagtcatccggtt catcggcaaaatcatggttgttcaggtctttaagattctga gccagcagaaagacgacccgacgtactggattggggt ggatccgggcgatacgatcggcatttgcctggttcgtatc accgcaaataaccgcgtcaagcgcttgctgtccatggaa ttccacaccagaagcaaagagatcaaacaactgttagag acacgccgtatgtatcgttgtacccgtcgttatttccgccg gaagcgcgtgaagaagaagggcagaattccgaaacac cgtccggcacgtttcgaaaaccgtcgccgctctgatggtt ggttgaccccaagcgtgcgccacttttggagagccata aatcggtcattcgtaagatcgccgaatgtgttccggacgt taatatcagcttg**GCG**tatgctaaattcgacatccagaa aatgatgaacccggatattgaaggtgaacaatatcaacgt ggcaaactgtttggctaccaaaacattcgtgcgtacgttct gcaccgcgataagtatacctgtcagctgtgcgcgaaaaa aaatgttccgtttcagttgcaccacgtggtcccgcgttcgc aaaacggtaaggaccacccggacaaccttatcacgctgt gcccggaatgccatgttaaggttcataccgataccaaact cttgaagaaactgcagcggaaaattaaaagcaaaatcgc tgattttacctccacctctcgtctgaacatcattatgagcttt ctgaagcattggctggagaacgaattcggcaaagaccg tgtgtggttctgcataggcgcggatagctatgtgatccgc aaggagctgaatctcgataagacccattgtaatgatgcgt actgcattgcgattaataaattggttcgtttcgatactgatct ggagcatgaatctgttgtggataagcaagtgatccgtttc gagcaaaagcgccagcataatagacgtatgaaatccct ggaacgtaaccgcctgtattacgtagtaaaagacaagaa gcgtgtgcgcgtggcgacgaaccgcaacggtaacccc gacgcctttcacgaggtttacgacaccaagatcctgccg ctgctgaaacagggttaccgtttgaaggtaaatccaggtg tgaagggttaccgtaagaagttcagcgaagttggtttcaa agagggtgctaccatccgcatcggcaaccgtgtggacg tgctgaagggtacttcaggttttggttacacgtacgacggt	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISL**A**YAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLHHVVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 41) *With no N-terminus M residue (SEQ ID NO: 145)*

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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)	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIKQLLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQNI RAYVLHRDKYTCQLCAKKNVPF QLHHVVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCN**A**AYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSEV GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 43) *With no N-terminus M residue (SEQ ID NO: 146)*
H232A + Q102R, I206R, and V433R	atgagacaccagagactcgtgtttctcgtggacggtaga aacaacatcctgcatcctaccagaaagctggatatggtc agaagatggctgagacaaacaaacgccaaggtgatca gattcatcggaaaaatcatggtggtccaggtgttcaagat cctgtctcagcagaaagacgaccctacatactggatcgg cgtggacccaggcgacaccatcggcatctgcctggtca	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGVDP GDTIGICLVRITANNRVKRLLSME FHTRSKEIK**R**LLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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: 212)	RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQN**R** RAYVLHRDKYTCQLCAKKNVPF QLH**A**VVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSE**R** GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 210) *With no N-terminus M residue (SEQ ID NO: 211)*
D67A + Q102R, I206R, and V433R	atgagacaccagagactcgtgtttctcgtggacggtaga aacaacatcctgcatcctaccagaaagctggatatggtc agaagatggctgagacaaacaaacgccaaggtgatca gattcatcggaaaaatcatggtggtccaggtgttcaagat cctgtctcagcagaaagacgaccctacatactggatcgg cgtg**GCC**ccaggcgacaccatcggcatctgcctggt cagaatcacagccaacaacagagttaagcggctgctga gcatggaattccacaccagatctaaggaaatcaag**CG** **G**ctgctggaaacccggagaatgtaccggtgcacacgg agatactttagaagaaagcgggtgaagaagaaaggga gaattcccaagcacagacctgctagatttgagaacagaa ggcggagcgacggctggctgacccctagcgtgcgcca cttccttgagagccacaagtccgtgatccggaagatcgc cgagtgtgtgccagatgtgaacatcagcctggagtacgc caagttcgacatccaaaagatgatgaaccccgatatcga	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGV**A**P GDTIGICLVRITANNRVKRLLSME FHTRSKEIK**R**LLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQN**R** RAYVLHRDKYTCQLCAKKNVPF QLHHVVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	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: 215)	LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSE**R** GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 213) *With no N-terminus M residue (SEQ ID NO: 214)*
D67A and H232A + Q102R, I206R, and V433R (dead Nuclease BT)	atgagacaccagagactcgtgtttctcgtggacggtaga aacaacatcctgcatcctaccagaaagctggatatggtc agaagatggctgagacaaacaaacgccaaggtgatca gattcatcggaaaaatcatggtggtccaggtgttcaagat cctgtctcagcagaaagacgaccctacatactggatcgg cgtg**GCC**ccaggcgacaccatcggcatctgcctggt cagaatcacagccaacaacagagttaagcggctgctga gcatggaattccacaccagatctaaggaaatcaag**CG** **G**ctgctggaaacccggagaatgtaccggtgcacacgg agatactttagaagaaagcgggtgaagaagaaaggga gaattcccaagcacagacctgctagatttgagaacagaa ggcggagcgacggctggctgacccctagcgtgcgcca cttccttgagagccacaagtccgtgatccggaagatcgc cgagtgtgtgccagatgtgaacatcagcctggagtacgc caagttcgacatccaaaagatgatgaaccccgatatcga gggcgagcagtaccagagaggcaaactgttcggatatc agaat**CGG**cgcgcctacgtgctgcatagagataagta cacctgtcagctgtgcgccaagaaaaacgtgcccttcca gctgcac**GCC**gtggtgcctagaagccagaacggca aggaccaccccgataacctgatcaccctgtgtcctgagt gccacgtgaaagtgcacacagacaccaagctgctgaaa aaactgcagcggaagattaagagcaagattgctgatttc accagcaccagcagactgaatattatcatgagcttcctga agcactggctggaaaacgagttcggcaaggatcgggtg	MRHQRLVFLVDGRNNILHPTRKL DMVRRWLRQTNAKVIRFIGKIM VVQVFKILSQQKDDPTYWIGV**A**P GDTIGICLVRITANNRVKRLLSME FHTRSKEIK**R**LLETRRMYRCTRR YFRRKRVKKKGRIPKHRPARFEN RRRSDGWLTPSVRHFLESHKSVI RKIAECVPDVNISLEYAKFDIQK MMNPDIEGEQYQRGKLFGYQN**R** RAYVLHRDKYTCQLCAKKNVPF QLH**A**VVPRSQNGKDHPDNLITLC PECHVKVHTDTKLLKKLQRKIKS KIADFTSTSRLNIIMSFLKHWLEN EFGKDRVWFCIGADSYVIRKELN LDKTHCNDAYCIAINKLVRFDTD LEHESVVDKQVIRFEQKRQHNRR MKSLERNRLYYVVKDKKRVRV ATNRNGNPDAFHEVYDTKILPLL KQGYRLKVNPGVKGYRKKFSE**R** GFKEGATIRIGNRVDVLKGTSGF GYTYDGTIFKLKDMDVVYPYGG WKCVKV (SEQ ID NO: 216)

RNA-Guided Nuclease/Nickase	Nucleotide sequence	Amino acid sequence
	tggttctgcatcggagccgactcttatgtgatccggaagg aactgaacctggacaagacacactgcaacgacgcttact gcatcgccatcaacaagctggtgagatttgacaccgacc tggaacacgaaagcgtggtggataagcaagtgatccgg ttcgagcagaaaagacagcacaaccggagaatgaaatc tctggagagaaacaggctgtactacgtggtcaaggaca agaagcgcgtgcgggtggccaccaacagaaatggcaa tcctgatgccttccacgaggtgtacgacacaaagatcctg cctctgctgaagcagggctatcggctgaaagtgaaccct ggcgttaagggctacaggaagaagttcagcgag**CG G**ggatttaaggaaggcgccacaatcagaatcggaaatc gggtggacgtgctgaagggcacctccggcttcggctac acctacgacggcaccatcttcaagctcaaggacatggac gtggtgtacccctacggcggctggaagtgcgtgaaggt g (SEQ ID NO: 218)	*With no N-terminus M residue (SEQ ID NO: 217)*

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 ggugucguuccagcuccuagcucugcgaaggagcauucguuaucggau 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.

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. An RNA-guided nuclease polypeptide, comprising a RuvC nuclease domain and an HNH nuclease domain, wherein the RNA-guided nuclease polypeptide comprises an amino acid sequence at least 90% identical to SEQ ID NO: 1 and comprises at least one mutation relative to SEQ ID NO: 1.

2. The RNA-guided nuclease polypeptide of claim 1, wherein the at least one mutation comprises:

(a) one or more arginine and/or lysine substitutions, optionally one or more arginine substitutions, relative to SEQ ID NO: 1;

(b) one or more nickase mutations in the HNH nuclease domain or in the RuvC nuclease domain; or

(c) a combination of (a) and (b).

3. The RNA-guided nuclease polypeptide of claim 1 or claim 2, wherein the at least one mutation comprises (a) and is located in a bridge helix (BH) domain, in a phosphate lock loop (PLL) domain, in a wedge (WED) domain, in a PAM-interacting (PID) domain, or a combination thereof.

4. The RNA-guided nuclease polypeptide of claim 2, wherein the at least one mutation comprises (a) and 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 SEQ ID NO: 1.

5. The RNA-guided nuclease polypeptide of claim 4, wherein the RNA-guided nuclease polypeptide comprises arginine and/or lysine substitutions, optionally 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.

6. The RNA-guided nuclease of any one of claims 2-5, wherein the RNA-guided nuclease polypeptide contains up to 20 arginine and/or lysine substitutions, optionally up to 20 arginine substitutions, relative to the reference RNA-guided nuclease; 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.

7. The RNA-guided nuclease of claim 6, wherein the RNA-guided nuclease polypeptide comprises the following combination of arginine substitutions:

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.

8. The RNA-guided nuclease of claim 7, wherein the RNA-guided nuclease polypeptide comprises the arginine substitutions of Q102R, I206R, and V433R.

9. The RNA-guided nuclease of any one of claims 2-8, wherein the RNA-guided nuclease polypeptide comprises or further comprises the one or more nickase mutations, which are at positions H231, H232, H255, D67, E176, and/or D329 of SEQ ID NO: 1.

10. The RNA-guided nuclease of claim 9, wherein the nickase mutation comprises a mutation at position H232, optionally wherein the mutation is amino acid substitution of H232A or H232L.

11. The RNA-guided nuclease polypeptide of any one of claims 1-10, wherein the RNA-guided nuclease polypeptide comprises an amino acid sequence at least 95% identical to SEQ ID NO: 1.

12. The RNA-guided nuclease polypeptide of claim 11, wherein the RNA-guided nuclease polypeptide comprises an amino acid sequence at least 98% identical to SEQ ID NO: 1.

13. The RNA-guided nuclease polypeptide of claim 1, wherein the RNA-guided nuclease polypeptide is listed in **Table 1, Table 4A**, or **Table 11**.

14. The RNA-guided nuclease polypeptide of any one of claims 1-13, wherein the RNA-guided nuclease polypeptide is in a fusion polypeptide, which further comprises one or more functional fragments.

15. The RNA-guided nuclease polypeptide of claim 14, wherein the one or more functional fragments comprise a nuclear localization signal (NLS), a peptide linker, or a combination thereof.

16. The RNA-guided nuclease polypeptide of claim 14 or claim 15, wherein the RNA-guided nuclease polypeptide in the fusion polypeptide does not includes its native N-terminal methionine residue.

17. A nucleic acid, comprising a nucleotide sequence encoding the RNA-guided nuclease polypeptide of any one of claims 1-16.

18. The nucleic acid of claim 17, wherein the nucleic acid is an expression vector, in which the nucleotide sequence encoding the RNA-guided nuclease is in operable linkage to a promoter.

19. A host cell comprising the nucleic acid of claim 17 or claim 18.

20. A gene editing system, comprising:

(a) an RNA-guided nuclease polypeptide or a first nucleic acid encoding the RNA-guided nuclease, wherein the RNA-guided nuclease polypeptide is set forth in any one of claims 1-16, and

(b) a guide RNA (gRNA) or a second nucleic acid encoding the gRNA, wherein the gRNA comprises (i) a spacer sequence, which is specific to a target sequence in a genomic site of interest, the target sequence being adjacent to a protospacer adjacent motif (PAM), and (ii) a scaffold sequence recognizable by the RNA-guided nuclease polypeptide.

21. The gene editing system of claim 20, wherein the gene editing system comprises the RNA-guided nuclease polypeptide.

22. The gene editing system of claim 20 or claim 21, 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.

23. The gene editing system of claim 22, 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).

24. The gene editing system of claim 22, wherein the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, or any one of SEQ ID NOs: 88-90 and 92-94.

25. The gene editing system of any one of claims 20-24, wherein the target sequence is upstream to the PAM of 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 20-25, wherein the spacer sequence is about 15-25 nt long.

27. The gene editing system of claim 26, wherein the spacer sequence is about 16-20 nt long.

28. The gene editing system of any one of claims 20-27, which further comprises one or more lipid excipients associated with the RNA-guided nuclease polypeptide or the first nucleic acid encoding such and/or the guide RNA or the second nucleic acid encoding such of the gene editing system; optionally wherein the one or more lipid excipients form lipid nanoparticles (LNPs), which are associated with or encapsulate the RNA-guided nuclease polypeptide or the first nucleic acid encoding such and/or the guide RNA or the second nucleic acid encoding such of the gene editing system; optionally wherein the gene editing system comprises the first nucleic acid encoding the RNA-guided nuclease, the first nucleic acid being a messenger RNA, and the gRNA.

29. The gene editing system of any one of claims 20-27, which comprise the first nucleic acid encoding the RNA-guided nuclease polypeptide, optionally wherein the first nucleic acid is located in a vector.

30. The gene editing system of claim 29, wherein the first nucleic acid is located in a vector, which is a viral vector, optionally wherein the viral vector is an adeno-associated viral (AAV) vector.

31. The gene editing system of claim 30, wherein the vector further comprises the second nucleic acid encoding the gRNA.

32. A gene editing method, comprising delivering the gene editing system of any one of claims 20-31 to a host cell to edit a genomic site targeted by the gRNA of the gene editing system.

33. A guide RNA, comprising a spacer sequence and a scaffold sequence, wherein the scaffold sequence comprises a nucleotide sequence at least 70%, optionally at least 75%, identical to SEQ ID NO: 2, and wherein the scaffold sequence is recognizable by the RNA-guided nuclease polypeptide set forth in any one of claims 1-16.

34. The guide RNA of claim 33, 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).

35. The guide RNA of claim 34, wherein the scaffold sequence comprises the nucleotide sequence of SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 26, or any one of SEQ ID NOs: 88-90 and 92-94.

36. The guide RNA of claim 34 or claim 35, wherein the spacer sequence is about 15-25 nt long.

37. The guide RNA of claim 36, wherein the spacer sequence is about 16-20 nt long.