(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 January 2025 (30.01.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/022367 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/IB2024/057272

(22) International Filing Date:
27 July 2024 (27.07.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/516,127 27 July 2023 (27.07.2023) US

(71) Applicant: **LIFE EDIT THERAPEUTICS, INC.** [US/US]; 300 Morris Street, Suite 300, Durham, North Carolina 27701 (US).

(72) Inventors: **CHONG, Shaorong**; 645 Pond Edge Ct., Chapel Hill, North Carolina 27516 (US). **COYLE, Michael**; 407 N. Elliott Road, Chapel Hill, North Carolina 27514 (US). **CRAWLEY, Alexandra Briner**; 1220 Scott Pl., Cary, North Carolina 27511 (US). **FAHAM, Salem**; 211 Zante Currant Rd., Durham, North Carolina 27703 (US). **KELSO, Drew**; 113 Oakbeech Ct., Holly Springs, North Carolina 27540 (US).

(74) Agent: **NGUYEN, Thu** et al.; Womble Bond Dickinson (US) LLP, P.O. Box 7037, Atlanta, Georgia 30357-0037 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

(54) Title: RNA-GUIDED NUCLEASES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF AND METHODS OF USE

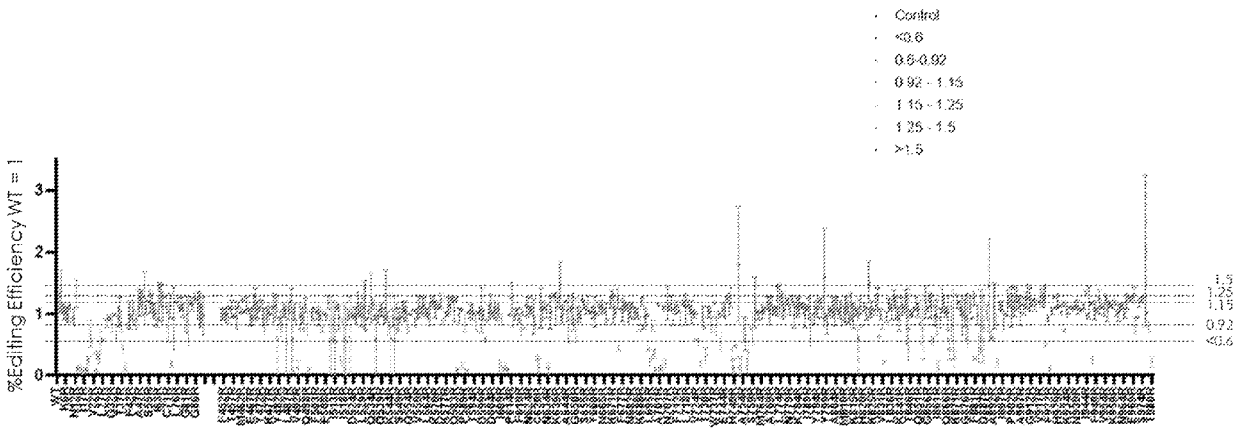


FIG. 1

(57) Abstract: Compositions and methods for binding to a target sequence of interest are provided. The compositions find use in cleaving or modifying a target sequence of interest, visualization of a target sequence of interest, and modifying the expression of a sequence of interest. Compositions comprise RNA-guided nuclease (RGN) polypeptides, CRISPR RNAs, trans-activating CRISPR RNAs, guide RNAs, and nucleic acid molecules encoding the same. Vectors and host cells comprising the nucleic acid molecules are also provided. Further provided are RGN systems for binding a target sequence of interest, wherein the RGN system comprises an RNA-guided nuclease polypeptide and one or more guide RNAs.

WO 2025/022367 A2

[Continued on next page]

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
— *with sequence listing part of description (Rule 5.2(a))*
— *in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE*

RNA-GUIDED NUCLEASES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF AND METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application No. 63/516,127, filed July 27, 2023, which is incorporated by referenced herein in its entirety.

REFERENCE TO A SEQUENCE LISTING SUBMITTED ELECTRONICALLY AS AN XML FILE

The instant application contains a Sequence Listing which has been submitted in xml format and is hereby incorporated by reference in its entirety. Said xml copy, created on July 24, 2024, is named L103438_1440WO_SL, and is 482,503 bytes in size.

FIELD OF THE INVENTION

The present invention relates to the field of molecular biology and gene editing.

BACKGROUND OF THE INVENTION

Targeted genome editing or modification is rapidly becoming an important tool for basic and applied research. Initial methods involved engineering nucleases such as meganucleases, zinc finger fusion proteins or TALENs, requiring the generation of chimeric nucleases with engineered, programmable, sequence-specific DNA-binding domains specific for each particular target sequence. RNA-guided nucleases, such as the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) proteins of the CRISPR-Cas bacterial system, allow for the targeting of specific sequences by complexing the nucleases with guide RNA that specifically hybridizes with a particular target sequence. Producing target-specific guide RNAs is less costly and more efficient than generating chimeric nucleases for each target sequence. Such RNA-guided nucleases can be used to edit genomes optionally through the introduction of a sequence-specific, double-stranded break that is repaired via error-prone non-homologous end-joining (NHEJ) to introduce a mutation at a specific genomic location. Alternatively, heterologous DNA may be introduced into the genomic site via homology-directed repair. RNA-guided nucleases (RGNs) can also be used for base editing when fused with a deaminase or prime editing when fused with reverse transcriptase.

Prime editing is a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site using an RNA-guided DNA binding protein (e.g., RGN) working in association with a reverse transcriptase (described in, e.g., US 11,447,770B1; WO2021072328; WO2021226558; WO2020156575; WO2021042047; US11193123; each incorporated by reference in its entirety herein). The prime editing system uses an RGN that is a nickase and a polymerase, and the system is programmed with a prime editing guide RNA that comprises a primer binding site (PBS) and a DNA synthesis template that serves as the template for the replacement strand comprising the edit. The prime

editor nicks the non-target strand upstream of the sequence to be edited and upstream of the PAM, creating a 3' flap on the non-target strand. The PBS is complementary to the 3' flap of the non-target strand and hybridrization of the PBS and 3' flap of the non-target strand allows for the polymerization of the replacement strand containing the edit using the DNA synthesis template.

BRIEF SUMMARY OF THE INVENTION

Compositions and methods for binding a target sequence of interest are provided. The compositions find use in cleaving or modifying a target sequence of interest, detection of a target sequence of interest, and modifying the expression of a sequence of interest. Compositions comprise: RNA-guided nuclease (RGN) polypeptide variants that have increased editing efficiency as compared to their counterpart wild-type RGN polypeptide; base editors; polymerase editors (PEs); CRISPR RNAs (crRNAs); trans-activating CRISPR RNAs (tracrRNAs); guide RNAs (gRNAs); nucleic acid molecules encoding the same; vectors and host cells comprising the nucleic acid molecules; and pharmaceutical compositions comprising the same. Also provided are RGN systems and ribonucleoprotein complexes for binding a target sequence of interest, wherein the RGN system comprises an RNA-guided nuclease polypeptide and one or more guide RNAs. Polymerase editor (PE) systems comprising one or more polymerase editing guide RNAs (PEgRNAs), a polymerase, and an RGN polypeptide are also provided. Thus, methods disclosed herein are drawn to binding a target sequence of interest in a target polynucleotide, and in some embodiments, cleaving or modifying the target polynucleotide of interest. The target polynucleotide of interest can be modified, for example, as a result of non-homologous end joining, homology-directed repair with an introduced donor sequence, or base editing, or polymerase editing.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 shows percent gene editing efficiency for 581 constructs in an initial screen. Each construct encoding a variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. One guide RNA was tested in duplicate (n=2) in the initial screen. Tracking of Indels by DEcomposition (TIDE) analysis was performed to determine % gene editing efficiency 2 days post-transfection. The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1.

FIG. 2 shows that 123 variant LPG10145 nucleases having an increase in gene editing activity ≥ 15% were obtained from the initial screen. The 123 variant LPG10145 nucleases were subjected to next generation sequencing (NGS) for confirmation of editing activity, which narrowed the 123 hits to 71 variants. The 71 variants were further tested with more guide RNAs. The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1.

FIG. 3 shows gene editing activity (% Indel) for variant LPG10145 nucleases tested with 3 additional guide RNAs (for a total of 4 guide RNAs tested: guides A, B, C, D). Each construct encoding a variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). Percent Indel (insertions/deletions) was determined by NGS 2 days post-transfection. The numbers on the x-axis for each guide from left to right are: 778, 969, 856, 822, 974, 55, 643, 653, 780, 954, 472, 968, 52, 86, 745, 973, 795, 541, 911, 533, 975, 774, and 872, and represent the amino acid position in LPG10145 nuclease of SEQ ID NO: 1 that is mutated.

FIG. 4 shows confirmation of gene editing activity by NGS of variant LPG10145 nucleases having an increase in gene editing activity from the initial screen. Twenty R variants increased gene editing > 20%. Each construct encoding a variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Three additional guide RNAs were tested (for a total of 4 guide RNAs tested). All guides were normalized and averaged. The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1.

FIG. 5 shows ranking of the top single variant LPG10145 nucleases with increased activity by statistical analysis. Left graph: Variant E778R at top to variant K872R, $p < 0.001$; variant K843R to variant K871R, $p < 0.01$. The right graph shows the statistical ranking of variants according to their position.

FIG. 6 shows a structural alignment of LPG10145 with a guide RNA. Many mutations, but not all, are at the interface with DNA/RNA. Set2 residues are labeled (unless disordered). The LPG10145 homology model is based on the structure of the RGN from *S. thermophilus* (6M0W).

FIG. 7 shows that 6 variant LPG10145 nucleases were selected based on statistical analysis and structural modeling for a combinatorial library. The single variants (1x variant), double variant combinations (2x variants), triple variant combinations (3x variants), quadruple variant combinations (4x variants), quintuple variant combinations (5x variant), and the sextuple combination (6x variant) are shown. Sixty-three constructs encoding the 63 variant/combinatorial variant were generated and were tested with multiple different guide RNAs. Each construct encoding a variant/combinatorial variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). Percent Indel (insertions/deletions) was determined by NGS 2 days post-transfection.

FIG. 8 shows increased gene editing by combinatorial variant LPG10145 nucleases with multiple different guide RNAs. Each construct encoding a combinatorial variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). A total of seven guide RNAs (shown as A, B1, B2, C1, C2, D1, and D2) were tested. Percent Indel (insertions/deletions) was determined by NGS 2 days post-transfection. Six of the seven guide RNAs tested showed significant increase in editing with the variants. Higher editing guide RNAs see less of an effect.

FIG. 9 shows combinatorial editing efficiency by combinatorial variant LPG10145 nucleases. The highest gene editing was obtained with 3x, 4x, and 5x variants. E778R and E969R were the common variants in the high editing populations. Each construct encoding a variant/combinatorial variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1.

FIG. 10 shows the top combinatorial variants with the highest significant editing. Fifteen combinatorial variant LPG10145 nucleases demonstrate the highest editing.

DETAILED DESCRIPTION

Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended embodiments. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

I. Overview

RNA-guided nucleases (RGNs) allow for the targeted manipulation of specific site(s) within a genome and are useful in the context of gene targeting for therapeutic and research applications. In a variety of organisms, including mammals, RNA-guided nucleases have been used for genome engineering by stimulating non-homologous end joining and homologous recombination, for example. The compositions and methods described herein are useful for creating single- or double-stranded breaks in polynucleotides, modifying polynucleotides, detecting a particular site within a polynucleotide, or modifying the expression of a particular gene.

The engineered variant LPG10145 RGNs, or active variants or fragments thereof, are directed to the target sequence by a guide RNA (gRNA) as part of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA-guided nuclease system. The RGNs are considered “RNA-guided” because guide RNAs form a complex with the RNA-guided nucleases to direct the RNA-guided nuclease to bind to a target sequence and in some embodiments, introduce a single-stranded or double-stranded break at the target sequence. The engineered variant LPG10145 RGNs, or active variants or fragments thereof, disclosed herein can increase gene editing efficiency as compared to the corresponding wild-type LPG10145 RGN. In some embodiments, the increase in gene editing efficiency is at least 15%. After the target sequence has been cleaved, the break can be repaired such that the DNA sequence of the target sequence is modified during the repair process. Thus, provided herein are methods for using the RNA-guided nucleases to modify

a target sequence in the DNA of host cells. For example, RNA-guided nucleases can be used to modify a target sequence at a genomic locus of eukaryotic cells or prokaryotic cells. In some embodiments, the variant LPG10145 RGNs can alter gene expression by modifying a target sequence.

Also provided herein are base editors, polymerase editors, base editing systems, polymerase editor systems, and methods of using the same for editing a target DNA molecule, wherein the editing systems comprise a DNA polymerase and an engineered variant LPG10145 RGN polypeptide as described herein, or an active variant or fragment thereof.

II. RNA-guided nucleases

Provided herein are RNA-guided nucleases. The term RNA-guided nuclease (RGN) refers to a polypeptide that binds to a particular target nucleotide sequence in a sequence-specific manner and is directed to the target nucleotide sequence by a guide RNA molecule that is complexed with the polypeptide and hybridizes with the target sequence. Although an RNA-guided nuclease can be capable of cleaving the target sequence upon binding, the term RNA-guided nuclease also encompasses nuclease-dead RNA-guided nucleases that are capable of binding to, but not cleaving, a target sequence. Cleavage of a target sequence by an RNA-guided nuclease can result in a single- or double-stranded break. RNA-guided nucleases only capable of cleaving a single strand of a double-stranded nucleic acid molecule are referred to herein as nickases.

The RNA-guided nucleases disclosed herein include engineered variants of the LPG10145 RNA-guided nuclease described in International Publ. No. WO 2023/139557, filed January 23, 2023, which is incorporated herein in its entirety. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. When referring to a variant LPG10145 RGN that "comprises an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs", it is intended to mean that the variant LPG10145 RGN comprises an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at the recited positions differ from the corresponding amino acid residues in SEQ ID NO: 1. When referring to an amino acid position, the number of the amino acid position is counted from the amino-terminus of a given polypeptide. The first amino acid residue at the amino-terminus of a polypeptide is denoted as position 1. In some embodiments, the first amino acid residue is a Methionine. The position numbers then increase numerically with each amino acid residue from the amino terminus of the polypeptide to the carboxy terminus of the polypeptide, with the last amino acid position being the last amino acid residue at the carboxy terminus of the polypeptide.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. When referring to a "variant LPG10145 RGN comprising an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue" or "a variant LPG10145 RGN comprising an amino acid sequence having at least x% (e.g., 85%) sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue", it is intended that the positively charged amino acid residue in the variant LPG10145 RGN is a different positively charged amino acid residue if the amino acid residue at a particular position of SEQ ID NO: 1 already comprises a positively charged amino acid residue.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that the amino acid residue at amino acid position 778 and/or 969 is an R, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at

positions 778 and 856 are positively charged amino acid residues, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1,

wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some of these embodiments, the active fragment or variant of a variant LPG10145 RGN is capable of cleaving a single- or double-stranded target sequence. When referring to a variant LPG10145 RGN that "comprises an amino acid sequence having at least x% (e.g., 85%) sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs", it is intended to mean that the variant LPG10145 RGN comprises an amino acid sequence that has at least x% (e.g., 85%) sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, and the amino acid residues at the recited positions differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID

NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some of these embodiments, the active fragment or variant of a variant LPG10145 RGN is capable of cleaving a single- or double-stranded target sequence. The binding and/or cleaving of a target sequence by an active variant LPG10145 RGN can be dependent upon the active variant LPG10145 RGN recognizing a protospacer adjacent motif (PAM) adjacent and 3' to the target sequence, and the PAM comprises a consensus sequence of NNGG. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1,

wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R; (ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R; (hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to: (a) the amino acid sequence set forth as SEQ

ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R; (c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and (o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence

set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1,

wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID

NO: 1. In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, an active fragment of a variant LPG10145 RGN comprises at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050 or more contiguous amino acid residues of an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16.

An RGN polypeptide of the disclosure can comprise one or more of the following domains: a linker domain (linker domain 1, linker domain 2), a wedge (WED) domain, a RuvC nuclease domain, an HNH nuclease domain, a Rec domain, or a PAM-interacting (PI) domain. The RuvC domain can include a RuvCI, a RuvCII, a RuvCIII domain, or a combination thereof. A Rec or recognition lobe mediates nucleic acid binding through multiple Rec domains (e.g., Rec1-3) by sensing nucleic acids, regulating the HNH conformational transition, and locking the catalytic HNH domain at the cleavage site. A wedge domain is responsible for the recognition of guide RNA scaffolds. A PAM-interacting domain is the domain of an RGN polypeptide that binds to a PAM site. Non-limiting examples of domains within LPG10145 RGN (SEQ ID NO: 1) or a variant LPG10145 RGN (SEQ ID NOs: 2-16, 182-196, and 271-285) include: RuvC-I from amino acid residues 1 to 42; BH from amino acid residues 43 to 79; REC1 from amino acid residues 80 to 236; REC2 from amino acid residues 237 to 476; RuvC-II from amino acid residues 477 to 524; L1 from amino acid residues 525 to 560; HNH from amino acid residues 561 to 676; L2 from amino acid residues 677 to 690; RuvC-III from amino acid residues 691 to 828; WED from amino acid residues 829 to 976; and PI from amino acid residues 977 to 1130. The general domains of RGN polypeptides can be determined via structural comparison to RGN polypeptides with defined domains.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a RuvCI domain comprising amino acid residues 1 to 42 or that differs from amino acid residues 1 to 42 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%,

94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a RuvCI domain comprising amino acid residues 1 to 42 or that differs from amino acid residues 1 to 42 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a BH domain comprising amino acid residues 43 to 79 or that differs from amino acid residues 43 to 79 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a BH domain comprising amino acid residues 43 to 79 or that differs from amino acid residues 43 to 79 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a REC1 domain comprising amino acid residues 80 to 236 or that differs from amino acid residues 80 to 236 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a REC1 domain comprising amino acid residues 80 to 236 or that differs from amino acid residues 80 to 236 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a REC2 domain comprising amino acid residues 237 to 476 or that differs from amino acid residues 237 to 476 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a REC2 domain comprising amino acid residues 237 to 476 or that differs from amino acid residues 237 to 476 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a RuvCII domain comprising amino acid residues 477 to 524 or that differs from amino acid residues 477 to 524 by 1,

2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a RuvCII domain comprising amino acid residues 477 to 524 or that differs from amino acid residues 477 to 524 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an L1 domain comprising amino acid residues 525 to 560 or that differs from amino acid residues 525 to 560 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise an L1 domain comprising amino acid residues 525 to 560 or that differs from amino acid residues 525 to 560 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an HNH domain comprising amino acid residues 561 to 676 or that differs from amino acid residues 561 to 676 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise an HNH domain comprising amino acid residues 561 to 676 or that differs from amino acid residues 561 to 676 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an L2 domain comprising amino acid residues 677 to 690 or that differs from amino acid residues 677 to 690 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise an L2 domain comprising amino acid residues 677 to 690 or that differs from

amino acid residues 677 to 690 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a RuvCIII domain comprising amino acid residues 691 to 828 or that differs from amino acid residues 691 to 828 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a RuvCIII domain comprising amino acid residues 691 to 828 or that differs from amino acid residues 691 to 828 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a WED domain comprising amino acid residues 829 to 976 or that differs from amino acid residues 829 to 976 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a WED domain comprising amino acid residues 829 to 976 or that differs from amino acid residues 829 to 976 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

An active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise a PI domain comprising amino acid residues 977 to 1130 or that differs from amino acid residues 977 to 1130 by 1, 2, or 3 amino acid residues, wherein the amino acid positions are in reference to SEQ ID NO: 1 or to a variant LPG10145 RGN, or an active variant or fragment thereof, disclosed herein (e.g., any one of SEQ ID NOs: 2-16, 182-196, and 271-285). For example, an active variant or fragment of a variant LPG10145 RGN disclosed herein can comprise an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 and can comprise a PI domain comprising amino acid residues 977 to 1130 or that differs from amino acid residues 977 to 1130 of the RGN sequence (i.e. SEQ ID NOs: 2-16, 182-196, and 271-285) by 1, 2, or 3 amino acid residues.

In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or

more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has nuclease activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the nuclease activity of a reference LPG10145 RGN. In some embodiments, a reference LPG10145 RGN has the amino acid sequence set forth as SEQ ID NO: 1. In some embodiments, a reference LPG10145 RGN is a variant of SEQ ID NO: 1 that lacks the corresponding mutations of the active variant or fragment thereof that has nuclease activity, e.g., as described above. In some embodiments, a reference LPG10145 RGN is a non-identical variant LPG10145 RGN. In some embodiments of this aspect, the active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, an active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments of the aspect wherein an active variant or fragment of a variant LPG10145 RGN disclosed herein can have nuclease activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the nuclease activity of a reference LPG10145 RGN, the active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, an active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, an active variant of a variant LPG10145 RGN comprises an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments of the aspect wherein an active variant or fragment of a variant LPG10145 RGN disclosed herein can have nuclease activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the nuclease activity of a reference LPG10145 RGN, the active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1,

wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16.

In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an

amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments of the aspect wherein an active variant or fragment of a variant LPG10145 RGN disclosed herein can have nuclease activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the nuclease activity of a reference LPG10145 RGN, the active variant of a variant LPG10145 RGN can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16.

RNA-guided nucleases provided herein can in embodiments comprise at least one nuclease domain (*e.g.*, DNase, RNase domain) and at least one RNA recognition and/or RNA binding domain to interact with guide RNAs. In some embodiments, the RGN comprises only one active nuclease domain and thus functions as a nickase. The RGN nuclease domain that is active in an RGN nickase can be a RuvC domain or an HNH domain. An RGN nickase can comprise an inactivated HNH nuclease domain or can lack an HNH domain. Further domains that can be found in RNA-guided nucleases provided herein include, but are not limited to: DNA binding domains, helicase domains, protein-protein interaction domains, and dimerization domains. In specific embodiments, the RNA-guided nucleases provided herein can comprise at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to one or more of a

DNA binding domain, helicase domain, protein-protein interaction domain, and dimerization domain. In some embodiments, variant LPG10145 RGNs of the disclosure do not comprise amino acid residues in the PAM recognition domain, the REC1 domain, the REC2 domain, and/or the Wing domain that differ from the corresponding residues in the wild-type LPG10145 RGN (e.g., SEQ ID NO: 1).

A target nucleotide sequence is bound by an RNA-guided nuclease provided herein and hybridizes with the guide RNA associated with the RNA-guided nuclease. The target sequence can then be subsequently cleaved by the RNA-guided nuclease if the polypeptide possesses nuclease activity. The terms "cleave" or "cleavage" refer to the hydrolysis of at least one phosphodiester bond within the backbone of a target nucleotide sequence that can result in either single-stranded or double-stranded breaks within the target sequence. The presently disclosed RGNs can cleave nucleotides within a polynucleotide, functioning as an endonuclease or can be an exonuclease, removing successive nucleotides from the end (the 5' and/or the 3' end) of a polynucleotide. In some embodiments, the disclosed RGNs can cleave nucleotides of a target sequence within any position of a polynucleotide and thus function as both an endonuclease and exonuclease. The cleavage of a target polynucleotide by the presently disclosed RGNs can result in staggered breaks or blunt ends. A staggered cut in a polynucleotide leads to two sticky ends or overhanging ends, and is formed when the nuclease cuts each strand of a polynucleotide such that the cuts are not directly opposite each other. For each sticky end of the cut polynucleotide, one strand (i.e. the overhanging strand) is longer than the other (typically by at least a few nucleotides), such that the longer strand has bases which are left unpaired. The longer strand of an overhanging end of a cleaved polynucleotide can have one unpaired nucleotide, two unpaired nucleotides, 3 unpaired nucleotides, 4 unpaired nucleotides, 5 unpaired nucleotides, or more unpaired nucleotides. In some embodiments, the longer strand of an overhanging end of a cleaved polynucleotide can have one unpaired nucleotide. The overhanging end of a cleaved polynucleotide can be a 3′ overhang or a 5′ overhang. In some embodiments, the overhanging end of a cleaved polynucleotide is a 3′ overhang. In some embodiments, the overhanging end of a cleaved polynucleotide is a 5′ overhang. In some embodiments, a variant LPG10145 RGN, or an active variant or fragment thereof, of the disclosure cleaves a target polynucleotide to form a staggered cut, wherein the staggered cut creates a 3′ overhang with one unpaired nucleotide. By contrast, a blunt cut generates two blunt ends, such that each blunt end of the cut polynucleotide has both strands that are of equal length – i.e. there are no unpaired bases on either strand of a blunt end.

The presently disclosed RNA-guided nucleases are variants of wild-type polypeptides. The wild-type RGN can be modified to alter nuclease activity or alter PAM specificity, for example. In some embodiments, the RNA-guided nuclease is not naturally-occurring.

In certain embodiments, the RNA-guided nuclease functions as a nickase, only cleaving a single strand of the target nucleotide sequence. Such RNA-guided nucleases have a single functioning nuclease domain. In particular embodiments, the nickase is capable of cleaving the positive strand or negative strand.

In some of these embodiments, additional nuclease domains have been mutated such that the nuclease activity is reduced or eliminated.

In other embodiments, the RNA-guided nuclease lacks nuclease activity altogether and is referred to herein as nuclease-dead or nuclease inactive. Any method known in the art for introducing mutations into an amino acid sequence, such as PCR-mediated mutagenesis and site-directed mutagenesis, can be used for generating nickases or nuclease-dead RGNs. See, e.g., U.S. Publ. No. 2014/0068797 and U.S. Pat. No. 9,790,490; each of which is incorporated by reference in its entirety. A non-limiting example of a nickase is the LPG10145 D16A nickase, which is set forth herein as SEQ ID NO: 124. Nickases which comprise a mutation in the RuvC domain and have a functional HNH domain are useful in base editing wherein the nickase is fused to a base editing polypeptide such as a deaminase. Another non-limiting example of a nickase is the LPG10145 H611A nickase, which is set forth herein as SEQ ID NO: 125. Nickases which comprise a mutation in the HNH domain and have a functional RuvC domain are useful in polymerase (i.e. prime) editing wherein the nickase is fused to a polymerase (i.e. prime) editing polypeptide such as a reverse transcriptase. A non-limiting example of a nuclease dead RGN is the LPG10145 D16A H611A sequence set forth as SEQ ID NO: 126. Variant LPG10145 RGNs of the disclosure can thus additionally comprise mutations in one or more nuclease domains that confer nickase function to the variant LPG10145 RGNs. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an alanine (or another non-conserved amino acid residue) at a position corresponding to 634 of SEQ ID NO: 1.

In some embodiments, variant LPG10145 RGNs of the disclosure that comprise the H611A mutation in the HNH domain comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% sequence identity to any one of SEQ ID NOs: 182-196. In some embodiments, variant LPG10145 RGNs of the disclosure that comprise the D16A mutation comprises the amino acid sequence set forth as any one of SEQ ID NOs: 271-285. In some embodiments, variant LPG10145 RGNs of the disclosure that comprise the D16A mutation comprises an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% sequence identity to any one of SEQ ID NOs: 271-285. In some embodiments, variant LPG10145 RGNs of the disclosure that comprise the D16A mutation comprises the amino acid sequence set forth as any one of SEQ ID NOs: 271-285.

RNA-guided nucleases that lack nuclease activity can be used to deliver a fused polypeptide, polynucleotide, or small molecule payload to a particular genomic location. In some of these embodiments,

the RGN polypeptide or guide RNA can be fused to a detectable label to allow for detection of a particular sequence. As a non-limiting example, a nuclease-dead RGN can be fused to a detectable label (*e.g.*, fluorescent protein) and targeted to a particular sequence associated with a disease to allow for detection of the disease-associated sequence.

Alternatively, nuclease-dead RGNs can be targeted to particular genomic locations to alter the expression of a desired sequence. In some embodiments, the binding of a nuclease-dead RNA-guided nuclease to a target sequence results in the reduction in expression of the target sequence or a gene under transcriptional control by the target sequence by interfering with the binding of RNA polymerase or transcription factors within the targeted genomic region. In some embodiments, the RGN (*e.g.*, a nuclease-dead RGN) or its complexed guide RNA further comprises an expression modulator that, upon binding to a target sequence, serves to either repress or activate the expression of the target sequence or a gene under transcriptional control by the target sequence. In some of these embodiments, the expression modulator modulates the expression of the target sequence or regulated gene through epigenetic mechanisms.

In other embodiments, the nuclease-dead RGNs or an RGN with nickase activity can be targeted to particular genomic locations to modify the sequence of a target polynucleotide through fusion to a base-editing polypeptide, for example a deaminase polypeptide or active variant or fragment thereof, that directly chemically modifies (e.g., deaminates) a nucleobase, resulting in conversion from one nucleobase to another. The base-editing polypeptide can be fused to the RGN at its N-terminal or C-terminal end. Additionally, the base-editing polypeptide may be fused to the RGN via a peptide linker. A non-limiting example of a deaminase polypeptide that is useful for such compositions and methods includes a cytosine deaminase or an adenine deaminase (such as the adenine deaminase base editor described in Gaudelli *et al.* (2017) *Nature* 551:464-471, U.S. Publ. Nos. 2017/0121693 and 2018/0073012, and International Publ. No. WO 2018/027078, or any of the deaminases disclosed in International Publ. No. WO 2020/139783, International Publ. No. WO 2022/056254, International Publ. No. WO 2022/204093, and International Publ. No. WO 2024/095245, each of which is herein incorporated by reference in its entirety). In some embodiments, the deaminase polypeptide that is useful for such compositions and methods is a cytosine deaminase or an adenine deaminase comprising an amino acid sequence selected from any one of SEQ ID NOs: 42-113, and 257. In one embodiment, the deaminase polypeptide that is useful for such compositions and methods is a cytosine deaminase or an adenine deaminase having a sequence that is at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to any one of the amino acid sequences set forth as SEQ ID NOs: 42-113, and 257. In some embodiments, the deaminase polypeptide that is useful for such the presently disclosed compositions and methods is a deaminase disclosed in Table 17 of International Publ. No. WO 2020/139783, which is incorporated herein by reference in its entirety. Further, it is known in the art that certain fusion proteins between an RGN and a base-editing enzyme (e.g., cytosine deaminase) may also comprise at least one uracil stabilizing polypeptide that increases the mutation rate of a cytidine,

deoxycytidine, or cytosine to a thymidine, deoxythymidine, or thymine in a nucleic acid molecule by a deaminase. Non-limiting examples of uracil stabilizing polypeptides include those disclosed in International Publ. No. WO 2022/015969, which is herein incorporated by reference in its entirety, including USP2 (SEQ ID NO: 40), and a uracil glycosylase inhibitor (UGI) domain (SEQ ID NO: 35), which may increase base editing efficiency. Therefore, a fusion protein may comprise an RGN described herein or variant thereof, a deaminase, and optionally at least one uracil stabilizing polypeptide, such as UGI or USP2. In certain embodiments, the RGN that is fused to the base-editing polypeptide is a nickase that cleaves the DNA strand that is not acted upon by the base-editing polypeptide (e.g., deaminase).

A variant LPG10145 RGN, or active variant or fragment thereof, of the disclosure can comprise a protospacer adjacent motif (PAM)-interacting (PI) domain that contributes to recognition of a PAM site in a target polynucleotide. In some embodiments, a variant LPG10145 RGN, or active variant or fragment thereof, recognizes and binds a consensus nucleotide sequence set forth as NNGG adjacent to the target polynucleotide. The PI domain can comprise 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150 or more amino acid residues. In some embodiments, the PI domain of an RGN, or an active variant or fragment thereof, of the disclosure is located within the carboxy (C)-terminal region of the RGN. The C-terminal region comprising the PI domain of an RGN, or an active variant or fragment thereof, of the disclosure can include the C-terminal 151 amino acid residues, the C-terminal 150 amino acid residues, the C-terminal 140 amino acid residues, the C-terminal 135 amino acid residues, the C-terminal 132 amino acid residues, the C-terminal 130 amino acid residues, the C-terminal 125 amino acid residues, the C-terminal 120 amino acid residues, the C-terminal 110 amino acid residues, the C-terminal 100 amino acid residues, the C-terminal 90 amino acid residues, the C-terminal 80 amino acid residues, the C-terminal 70 amino acid residues, the C-terminal 60 amino acid residues, the C-terminal 50 amino acid residues, the C-terminal 40 amino acid residues, the C-terminal 30 amino acid residues, the C-terminal 20 amino acid residues, or the C-terminal 10 amino acid residues of the RGN. In some embodiments, the PI domain of an RGN, or an active variant or fragment thereof, of the disclosure is within or includes amino acid residues 977-1130 of the RGN. In some embodiments, the PI domain of a variant LPG10145 RGN polypeptide, or an active variant or fragment thereof, of the disclosure has the amino acid sequence set forth as SEQ ID NO: 253.

III. Fusion Proteins Comprising variant LPG10145 RGN polypeptides

The present disclosure provides fusion proteins comprising the presently disclosed engineered variant LPG10145 RGN polypeptides, or active variants or fragments thereof, operably fused to at least one heterologous polypeptide, as well as polynucleotides encoding the fusion proteins. When used to refer to the joining of two protein coding regions (either by fusion or insertion), by "operably linked" or "operably fused" is intended that the coding regions are in the same reading frame, even if one is inserted into another.

In some embodiments, polypeptides that are "operably fused" or "operably linked" means that the structure and/or biological activity of each individual peptide is also present in the fusion.

As used herein, "heterologous", in reference to a polypeptide that is heterologous to another polypeptide (e.g., a variant LPG10145 RGN), is a polypeptide that is not operably fused to the presently described variant LPG10145 RGN in nature. The heterologous polypeptide can originate from a foreign species or from the same species. The heterologous polypeptide can be in its native form or is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. The heterologous polypeptide can be any polypeptide, including but not limited to a localization signal, a cell-penetrating domain, a polymerase (e.g., reverse transcriptase), a base editing polypeptide, an effector domain (e.g., a cleavage domain, a deaminase domain, or an expression modulator domain), a detectable label (e.g., fluorescent protein), or a purification tag, or active variant or fragment thereof. The heterologous polypeptide can be operably fused to the amino (N)-terminus, to the carboxy (C)-terminus, or to an internal location of a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, described herein. In some embodiments, the heterologous polypeptide is operably fused to a variant LPG10145 RGN polypeptide by a peptide linker as described herein.

In some embodiments, a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, of the disclosure may be fused to a polymerase (i.e. prime) editing polypeptide (e.g., DNA polymerase or reverse transcriptase) to generate a polymerase editor (PE). The polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) can be operably fused to the amino (N)-terminus, to the carboxy (C)-terminus, or to an internal location of a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, described herein. In some embodiments, the polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) is operably fused to a variant LPG10145 RGN polypeptide by a peptide linker as described herein. Polymerase (i.e.p) editing is a versatile and precise genome editing method that directly writes new genetic information into a specified DNA site using a nucleic acid programmable DNA binding protein working in association with a polymerase. The polymerase (i.e. prime) editing system uses an RGN that is a nickase, and the system is programmed with a polymerase (i.e. prime) editing (PE) guide RNA ("PEgRNA"). The PEgRNA is a guide RNA that both specifies the target sequence and provides the template for polymerization of the replacement strand containing the edit by way of an extension engineered onto the guide RNA (e.g., at the 5' or 3' end, or at an internal portion of the guide RNA). The RGN nickase/polymerase (i.e. prime) editing polypeptide fusion is guided to the target sequence by the PEgRNA and nicks the target strand upstream of sequence to be edited and upstream of the PAM, creating a 3′ flap on the target strand. The PEgRNA includes a primer binding site (PBS) that is complementary to the 3′ flap of the target strand. In some embodiments, a PBS is at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides in length. In certain embodiments, the pegRNA comprises a PBS that is at least 5 (e.g., at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 28, 19, or 20) nucleotides in length. In some

embodiments, the pegRNA may comprise a PBS that is at least 8 nucleotides in length. Hybridrization of the PBS and 3′ flap of the target strand allows polymerization of the replacement strand containing the edit using the extension of the PEgRNA as template. The extension of the PEgRNA can be formed from RNA or DNA. In the case of an RNA extension, the polymerase of the polymerase (i.e. prime) editor can be an RNA-dependent DNA polymerase (such as a reverse transcriptase). In the case of a DNA extension, the polymerase of the polymerase (i.e. prime) editor may be a DNA-dependent DNA polymerase.

The replacement strand containing the desired edit (e.g., a single nucleobase substitution) shares the same sequence as the target strand of the target sequence to be edited (with the exception that it includes the desired edit). Through DNA repair and/or replication machinery, the target strand of the target sequence is replaced by the newly synthesized replacement strand containing the desired edit. In some cases, polymerase (i.e. prime) editing may be thought of as a "search-and-replace" genome editing technology since the polymerase (i.e. prime) editors not only search and locate the desired target sequence to be edited, but at the same time, encode a replacement strand containing a desired edit which is installed in place of the corresponding target strand of the target sequence. Thus, in some embodiments, a guide RNA of the disclosure comprises an extension comprising an edit template for polymerase (i.e. prime) editing. In some embodiments, a polymerase (i.e. prime) editing polypeptide that can be fused to an RGN includes a DNA polymerase. In certain embodiments, the DNA polymerase is a reverse transcriptase. In certain embodiments, the RGN is a nickase. A fusion protein comprising a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, disclosed herein and a reverse transcriptase can comprise an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to any one of SEQ ID NOs: 135, and 162-181. In some embodiments, a fusion protein comprising a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, disclosed herein and a reverse transcriptase comprises the amino acid sequence set forth as any one of SEQ ID NOs: 135, and 162-181.

A variant LPG10145 RGN polypeptide, or active variant or fragment thereof, of the disclosure may be operably fused to a base-editing polypeptide as described herein. The base-editing polypeptide (e.g., deaminase) can be operably fused to the amino (N)-terminus, to the carboxy (C)-terminus, or to an internal location of a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, described herein. The base-editing polypeptide (e.g., deaminase) may be operably fused to the RGN via a peptide linker. In some embodiments, the base editing polypeptide in a fusion protein of the disclosure is a cytosine deaminase or an adenine deaminase comprising an amino acid sequence set forth as any one of SEQ ID NOs: 42-113, and 257. In some embodiments, the base editing polypeptide in a fusion protein of the disclosure is a cytosine deaminase or an adenine deaminase having an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set

forth as any one of SEQ ID NOs: 42-113, and 257. In some embodiments, the base editing polypeptide in a fusion protein of the disclosure is a deaminase described in Gaudelli *et al.* (2017) *Nature* 551:464-471, U.S. Publ. Nos. 2017/0121693 and 2018/0073012, and International Publ. No. WO 2018/027078, or any of the deaminases disclosed in International Publ. No. WO 2020/139783, International Publ. No. WO 2022/056254, International Publ. No. WO 2022/204093, and International Publ. No. WO 2024/095245, each of which is herein incorporated by reference in its entirety. The fusion protein comprising an engineered variant LPG10145 RGN polypeptide and a base-editing polypeptide (e.g., cytosine or adenine deaminase) may further comprise at least one uracil stabilizing polypeptide that increases the mutation rate of a nucleobase (e.g., cytidine, deoxycytidine, or cytosine to a thymidine, deoxythymidine, or thymine) in a nucleic acid molecule by a deaminase. In some embodiments, a USP comprises USP2 (SEQ ID NO: 40) or a uracil glycosylase inhibitor (UGI) domain (SEQ ID NO: 35), which may increase base editing efficiency. In some embodiments, the variant LPG10145 RGN polypeptide that is fused to the base-editing polypeptide is a nickase that cleaves the DNA strand that is not acted upon by the base-editing polypeptide (e.g., deaminase). In some embodiments, the variant LPG10145 RGN polypeptide that is fused to the base-editing polypeptide is a nuclease-dead RGN polypeptide. A fusion protein comprising a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, disclosed herein and a deaminase can comprise an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to any one of SEQ ID NOs: 258-270. In some embodiments, a fusion protein comprising a variant LPG10145 RGN polypeptide, or active variant or fragment thereof, disclosed herein and a deaminase comprises the amino acid sequence set forth as any one of SEQ ID NOs: 258-270.

The presently disclosed variant LPG10145 RNA-guided nucleases, or active variants or fragments thereof, can comprise at least one nuclear localization signal (NLS) to enhance transport of the RGN to the nucleus of a cell. Nuclear localization signals are known in the art and generally comprise a stretch of basic amino acids (see, *e.g.*, Lange *et al.*, *J. Biol. Chem.* (2007) 282:5101-5105). In some embodiments, the RGN comprises 2, 3, 4, 5, 6 or more nuclear localization signals. The nuclear localization signal(s) can be a heterologous NLS. Non-limiting examples of nuclear localization signals useful for the presently disclosed RGNs are the nuclear localization signals of SV40 Large T-antigen, nucleoplasmin, and c-Myc (*see, e.g.*, Ray et al. (2015) *Bioconjug Chem* 26(6):1004-7). In some embodiments, the NLS has an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more sequence identity to any one of SEQ ID NOs: 36, 37, 234, or 235. In particular embodiments, the RGN comprises the NLS sequence set forth as SEQ ID NO: 36, 37, 234, or 235. The RGN can comprise one or more NLS sequences operably fused at its N-terminus, C- terminus, or both the N-terminus and C-terminus. For example, the RGN can comprise two NLS sequences at the N-terminal region and four NLS sequences at the C-terminal region.

Other localization signal sequences known in the art that localize polypeptides to particular subcellular location(s) can also be used to target the presently disclosed variant LPG10145 RGNs, or active variants or fragments thereof, including, but not limited to, plastid localization sequences, mitochondrial localization sequences, and dual-targeting signal sequences that target to both the plastid and mitochondria (see, *e.g.*, Nassoury and Morse (2005) *Biochim Biophys Acta* 1743:5-19; Kunze and Berger (2015) *Front Physiol* dx.doi.org/10.3389/fphys.2015.00259; Herrmann and Neupert (2003) *IUBMB Life* 55:219-225; Soll (2002) *Curr Opin Plant Biol* 5:529-535; Carrie and Small (2013) *Biochim Biophys Acta* 1833:253-259; Carrie *et al.* (2009) *FEBS J* 276:1187-1195; Silva-Filho (2003) *Curr Opin Plant Biol* 6:589-595; Peeters and Small (2001) *Biochim Biophys Acta* 1541:54-63; Murcha *et al.* (2014) *J Exp Bot* 65:6301-6335; Mackenzie (2005) *Trends Cell Biol* 15:548-554; Glaser *et al.* (1998) *Plant Mol Biol* 38:311-338).

In certain embodiments, the presently disclosed variant LPG10145 RNA-guided nucleases, or active variants or fragments thereof, comprise at least one cell-penetrating domain that facilitates cellular uptake of the RGN. Cell-penetrating domains are known in the art and generally comprise stretches of positively charged amino acid residues (*i.e.*, polycationic cell-penetrating domains), alternating polar amino acid residues and non-polar amino acid residues (*i.e.*, amphipathic cell-penetrating domains), or hydrophobic amino acid residues (*i.e.*, hydrophobic cell-penetrating domains) (*see, e.g.*, Milletti F. (2012) *Drug Discov Today* 17:850-860). A non-limiting example of a cell-penetrating domain is the trans-activating transcriptional activator (TAT) from the human immunodeficiency virus 1.

The nuclear localization signal, plastid localization signal, mitochondrial localization signal, dual-targeting localization signal, and/or cell-penetrating domain can be located at the amino-terminus (N-terminus), the carboxyl-terminus (C-terminus), or in an internal location of the RNA-guided nuclease.

The presently disclosed variant LPG10145 RGNs, or active variants or fragments thereof, can be fused to an effector domain, such as a cleavage domain, a deaminase domain, or an expression modulator domain, either directly or indirectly via a linker peptide. Such a domain can be located at the N-terminus, the C-terminus, or an internal location of the RNA-guided nuclease. In some of these embodiments, the RGN component of the fusion protein is a nuclease-dead RGN or a nickase.

In some embodiments, the RGN fusion protein comprises a cleavage domain, which is any domain that is capable of cleaving a polynucleotide (*i.e.*, RNA, DNA, or RNA/DNA hybrid) and includes, but is not limited to, restriction endonucleases and homing endonucleases, such as Type IIS endonucleases (*e.g.*, *Fok*I) (*see, e.g.*, Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Linn *et al.* (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993).

In other embodiments, the RGN fusion protein comprises a deaminase domain that deaminates a nucleobase, resulting in conversion from one nucleobase to another, and includes, but is not limited to, a cytosine deaminase or an adenine deaminase (see, *e.g.*, Gaudelli *et al.* (2017) *Nature* 551:464-471, U.S. Publ. Nos. 2017/0121693 and 2018/0073012, U.S. Patent No. 9,840,699, and International Publ. No. WO/2018/027078, or any of the deaminases disclosed in International Publ. No. WO 2020/139783,

International Publ. No. WO 2022/056254, and International Publ. No. WO 2022/204093, each of which is herein incorporated by reference in its entirety).

In some embodiments, the effector domain of the RGN fusion protein can be an expression modulator domain, which is a domain that either serves to upregulate or downregulate transcription. The expression modulator domain can be an epigenetic modification domain, a transcriptional repressor domain or a transcriptional activation domain.

In some of these embodiments, the expression modulator of the RGN fusion protein comprises an epigenetic modification domain that covalently modifies DNA or histone proteins to alter histone structure and/or chromosomal structure without altering the DNA sequence, leading to changes in gene expression (*i.e.*, upregulation or downregulation). Non-limiting examples of epigenetic modifications include acetylation or methylation of lysine residues, arginine methylation, serine and threonine phosphorylation, and lysine ubiquitination and sumoylation of histone proteins, and methylation and hydroxymethylation of cytosine residues in DNA. Non-limiting examples of epigenetic modification domains include histone acetyltransferase domains, histone deacetylase domains, histone methyltransferase domains, histone demethylase domains, DNA methyltransferase domains, and DNA demethylase domains.

In other embodiments, the expression modulator of the fusion protein comprises a transcriptional repressor domain, which interacts with transcriptional control elements and/or transcriptional regulatory proteins, such as RNA polymerases and transcription factors, to reduce or terminate transcription of at least one gene. Transcriptional repressor domains are known in the art and include, but are not limited to, Sp1-like repressors, IκB, and Krüppel associated box (KRAB) domains.

In yet other embodiments, the expression modulator of the fusion protein comprises a transcriptional activation domain, which interacts with transcriptional control elements and/or transcriptional regulatory proteins, such as RNA polymerases and transcription factors, to increase or activate transcription of at least one gene. Transcriptional activation domains are known in the art and include, but are not limited to, a herpes simplex virus VP16 activation domain and an NFAT activation domain.

The presently disclosed variant LPG10145 RGN polypeptides, or active variants or fragments thereof, can comprise a detectable label or a purification tag. The detectable label or purification tag can be located at the N-terminus, the C-terminus, or an internal location of the RNA-guided nuclease, either directly or indirectly via a linker peptide. In some of these embodiments, the RGN component of the fusion protein is a nuclease-dead RGN. In other embodiments, the RGN component of the fusion protein is an RGN with nickase activity.

A detectable label is a molecule that can be visualized or otherwise observed. The detectable label may be fused to the RGN as a fusion protein (*e.g.*, fluorescent protein) or may be a small molecule conjugated to the RGN polypeptide that can be detected visually or by other means. Detectable labels that can be fused to the presently disclosed RGNs as a fusion protein include any detectable protein domain, including but not limited to, a fluorescent protein or a protein domain that can be detected with a specific

antibody. Non-limiting examples of fluorescent proteins include green fluorescent proteins (*e.g.*, GFP, EGFP, ZsGreen1) and yellow fluorescent proteins (*e.g.*, YFP, EYFP, ZsYellow1). Non-limiting examples of small molecule detectable labels include radioactive labels, such as ^{3}H and ^{35}S.

Variant LPG10145 RGN polypeptides, or active variants or fragments thereof, of the disclosure can also comprise a purification tag, which is any molecule that can be utilized to isolate a protein or fused protein from a mixture (e.g., biological sample, culture medium). Non-limiting examples of purification tags include biotin, myc, maltose binding protein (MBP), glutathione-S-transferase (GST), and 3X FLAG tag.

Variant LPG10145 RNA-guided nucleases, or active variants or fragments thereof, of the disclosure that are fused to a heterologous polypeptide or domain can be separated or joined by a linker. The term "linker," as used herein, refers to a chemical group or a molecule linking two molecules or moieties, *e.g.*, a binding domain and a cleavage domain of a nuclease. In some embodiments, a linker joins a gRNA binding domain of an RNA guided nuclease and a base-editing polypeptide, such as a deaminase. In some embodiments, a linker joins a nuclease-dead RGN and a deaminase. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety.

In some embodiments, the linker is an amino acid or a plurality of amino acids (*e.g.*, a peptide or protein). A peptide linker of the disclosure can connect one polypeptide to another in a fusion protein. For example, a peptide linker can connect a variant LPG10145 RGN polypeptide and a heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain). A fusion protein comprising 3 polypeptides (e.g., a variant LPG10145 RGN polypeptide, a polymerase, and a detectable label) can comprise at least one peptide linker. In some embodiments, a peptide linker comprises at least one NLS. In some embodiments, a peptide linker comprises 2 NLSs. The peptide linker can be operably fused at the N-terminus, the C-terminus, or both the N-terminus and C-terminus of the variant LPG10145 RGN polypeptide or the heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain). In some embodiments, a peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1. In certain embodiments, a peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x or y is 0, 1, 2, 3, or 4, and y is 0; and wherein one of m or n is 0, and the other is 1. In other embodiments, the peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 1. In some embodiments, a peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241). In some embodiments, the peptide linker is 4-100 amino acids in length, for example, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a peptide linker has a length of at

least 13 amino acids, including but not limited to about 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, or more amino acids. In some embodiments, the peptide linker has an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more sequence identity to any one of SEQ ID NOs: 236-241. In some embodiments, the peptide linker has the amino acid sequence of any one of SEQ ID NOs: 236-241.

In some embodiments, the heterologous polypeptide comprises a polymerase (e.g., reverse transcriptase), and the peptide linker between the variant LPG10145 RGN polypeptide and the polymerase (e.g., reverse transcriptase) comprises at least 2 amino acids, at least 3 amino acids, at least 4 amino acids, at least 5 amino acids, at least 6 amino acids, at least 7 amino acids, at least 8 amino acids, at least 9 amino acids, at least 10 amino acids, at least 11 amino acids, at least 12 amino acids, or at least 13 amino acids. In some embodiments, the peptide linker between the variant LPG10145 RGN polypeptide and the polymerase (e.g., reverse transcriptase) comprises at least 13 amino acids, including but not limited to about 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, or more amino acids. In some embodiments, the fusion protein comprises a variant LPG10145 RGN polypeptide connected to a polymerase (e.g., reverse transcriptase) by a peptide linker having an amino acid sequence having at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more sequence identity to any one of SEQ ID NOs: 236-241. In some embodiments, the fusion protein comprises a variant LPG10145 RGN polypeptide connected to a polymerase (e.g., reverse transcriptase) by a peptide linker having the amino acid sequence set forth as any one of SEQ ID NOs: 236-241.

A fusion protein of the disclosure in some embodiments comprises a presently disclosed variant LPG10145 RGN polypeptide and a presently disclosed heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain). In some embodiments, a fusion protein of the disclosure comprises from amino terminus to carboxy terminus: a variant LPG10145 RGN polypeptide and a heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain). In some embodiments, a fusion protein of the disclosure comprises from amino terminus to carboxy terminus: a heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain) and a variant LPG10145 RGN polypeptide.

The fusion protein can in some embodiments comprises a heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain) inserted within a variant LPG10145 RGN polypeptide. In some embodiments, the heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain) is inserted between surface amino acid residues. The heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain) can be inserted within or

between a linker domain 2, a wedge (WED) domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting (PI) domain of the RGN as described in more detail elsewhere herein.

The heterologous polypeptide (e.g., a polymerase, a base editing polypeptide, an effector domain) in some embodiments is inserted within a variant LPG10145 RGN polypeptide immediately after the amino acid position selected from the group consisting of: i) amino acid position corresponding to position 347 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; ii) amino acid position corresponding to position 524 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; iii) amino acid position corresponding to position 640 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; iv) amino acid position corresponding to position 666 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; v) amino acid position corresponding to position 680 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; vi) amino acid position corresponding to position 740 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; vii) amino acid position corresponding to position 785 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; viii) amino acid position corresponding to position 910 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; and ix) amino acid position corresponding to position 1077 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285.

IV. Guide RNA

The present disclosure provides guide RNAs and polynucleotides encoding the same. The term "guide RNA" refers to a nucleotide sequence having sufficient complementarity with a target nucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of an associated RNA-guided nuclease to the target nucleotide sequence. More specifically, when the target nucleotide sequence is double-stranded as is the case with DNA, the target nucleotide sequence comprises a target strand and a non-target strand (which comprises the PAM sequence). In these embodiments, the guide RNA has sufficient

complementarity with the target strand of a double-stranded target sequence (e.g., target DNA sequence) such that the guide RNA hybridizes with the target strand and directs sequence-specific binding of an associated RGN to the target sequence (e.g., target DNA sequence). Therefore, in some embodiments, a guide RNA includes a spacer that is identical to the sequence of the non-target strand except that uracil (U) replaces thymidine (T) in the guide RNA. In embodiments where multiplex gene editing is used and there are multiple guide RNAs, each of the one or more guide RNA has sufficient complementarity with the target strand of a particular target sequence and is capable of hybridizing to the target strand of that target sequence. Thus, "a corresponding target sequence" for a guide RNA refers to the target sequence that the guide RNA has sufficient complementarity with and is capable of hybridizing to.

Thus, an RGN's respective guide RNA is one or more RNA molecules (generally, one or two), that can bind to the RGN and guide the RGN to bind to a particular target nucleotide sequence, and in those embodiments wherein the RGN has nickase or nuclease activity, also cleave the target strand and/or the non-target strand. In general, a guide RNA comprises a CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). Native guide RNAs that comprise both a crRNA and a tracrRNA generally comprise two separate RNA molecules that hybridize to each other through the repeat sequence of the crRNA and the anti-repeat sequence of the tracrRNA. A guide RNA can encompass a polymerase editing (PE) guide RNA.

Native direct repeat sequences within a CRISPR array generally range in length from 28 to 37 base pairs, although the length can vary between about 23 bp to about 55 bp. Spacer sequences within a CRISPR array generally range from about 32 to about 38 bp in length, although the length can be between about 21 bp to about 72 bp. Each CRISPR array generally comprises less than 50 units of the CRISPR repeat-spacer sequence. The CRISPRs are transcribed as part of a long transcript termed the primary CRISPR transcript, which comprises much of the CRISPR array. The primary CRISPR transcript is cleaved by Cas proteins to produce crRNAs or in some cases, to produce pre-crRNAs that are further processed by additional Cas proteins into mature crRNAs. Mature crRNAs comprise a spacer sequence and a CRISPR repeat sequence. In some embodiments in which pre-crRNAs are processed into mature (or processed) crRNAs, maturation involves the removal of about one to about six or more 5', 3', or 5' and 3' nucleotides. For the purposes of genome editing or targeting a particular target nucleotide sequence of interest, these nucleotides that are removed during maturation of the pre-crRNA molecule are not necessary for generating or designing a guide RNA.

A guide RNA of the disclosure can comprise at least one chemical modification. The at least one chemical modification includes: a bridged nucleic acid (BNA) modification; 2'-O-methyl (2'-O-Me) modification; 2'-O-methoxy-ethyl (2'MOE) modification; 2'-fluoro (2'-F) modification; 2'F-4'Cα-OMe modification; 2',4'-di-Cα-OMe modification; 2'-O-methyl 3'phosphorothioate (MS) modification; 2'-O-methyl 3'thiophosphonoacetate (MSP) modification; 2'-O-methyl 3'phosphonoacetate (MP) modification; and phosphorothioate (PS) modification; or a combination thereof. In some embodiments, the BNA comprises a 2′,4′ BNA modification. In some embodiments, the 2′,4′ BNA modification is selected from the

group consisting of: locked nucleic acid (LNA) modification, BNA^{NC}[N-Me] modification, 2′-O,4′-C-ethylene bridged nucleic acid (2′,4′-ENA) modification, and S-constrained ethyl (cEt) modification. In some embodiments, the 2′,4′ BNA is a LNA modification. In some embodiments, the 2′,4′ BNA is a cEt modification. In some embodiments, the at least one chemical modification comprises a BNA modification, 2'-O-Me modification, or PS modification. Chemical modifications of spacers, crRNA repeats, crRNAs, tracrRNAs, and guide RNAs are described in International Application Publication No. WO 2024/042489, which is hereby incorporated by reference in its entirety herein.

The present disclosure provides compositions comprising guide RNAs comprising CRISPR RNAs (crRNAs). A CRISPR RNA (crRNA) comprises a spacer and a CRISPR repeat. The "spacer" is the nucleotide sequence that directly hybridizes with the target nucleotide sequence of interest. The spacer is engineered to be fully or partially complementary with the target sequence of interest. In various embodiments, the spacer can comprise from about 8 nucleotides to about 30 nucleotides, or more. For example, the spacer sequence can be about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In some embodiments, the spacer is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In some embodiments, the spacer is about 10 to about 26 nucleotides in length, or about 12 to about 30 nucleotides in length. In some embodiments, the spacer is 19-27 nucleotides in length. In particular embodiments, the spacer is about 30 nucleotides in length. In some embodiments, the spacer is 30 nucleotides in length. In some embodiments, the spacer is about 20-25 nucleotides in length. In some embodiments, the degree of complementarity between a spacer and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is between 50% and 99% or more, including but not limited to about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a spacer and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more. In some embodiments, the spacer can be identical in sequence to the non-target strand of a target sequence. In some of those embodiments wherein the target sequence is a target DNA sequence, the spacer can be identical in sequence to the non-target strand of the target DNA sequence, with the exception of the thymidines (Ts) in the non-target strand being replaced by uracils (Us) in the spacer. In particular embodiments, the spacer is free of secondary structure, which can be predicted using any suitable polynucleotide folding algorithm known in the art, including but not limited to mFold (see, *e.g.*, Zuker and Stiegler (1981) *Nucleic Acids Res.* 9:133-148) and RNAfold (see, *e.g.*, Gruber et al. (2008) *Cell* 106(1):23-24).

Along with a spacer, a crRNA further comprises a CRISPR RNA (crRNA) repeat. Generally, a CRISPR RNA repeat comprises a nucleotide sequence that forms a structure, either on its own or in concert with a hybridized tracrRNA, that is recognized by the RGN polypeptide. For LPG10145 guide crRNAs, the spacer is 5′ of the crRNA repeat. In various embodiments, the CRISPR RNA repeat can comprise from about 8 nucleotides to about 30 nucleotides, or more. For example, the CRISPR repeat can be about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In particular embodiments, the CRISPR repeat is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In some embodiments, a crRNA repeat comprises a total length of 19 to 40 nucleotides (nt). In some embodiments, a crRNA repeat comprises a total length of at most 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, 26 nt, 27 nt, 28 nt, 29 nt, or 30 nt. In some embodiments, the crRNA repeat is about 19 nt or 21 nt. In some embodiments, the degree of complementarity between a CRISPR repeat and its corresponding tracrRNA, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a CRISPR repeat and its corresponding tracrRNA, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more.

In particular embodiments, the CRISPR repeat comprises the nucleotide sequence of SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof that when comprised within a guide RNA, is capable of directing the sequence-specific binding of an associated variant LPG10145 RGN provided herein to a target sequence of interest. In certain embodiments, an active CRISPR repeat sequence variant of a wild-type sequence comprises a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245. In certain embodiments, an active CRISPR repeat fragment of a wild-type sequence comprises at least 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245. The CRISPR repeat can comprise a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides. In some

embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide. In some embodiments, the CRISPR repeat comprises the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245.

In those embodiments wherein the RGN has the amino acid sequence set forth as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285, the crRNA repeat of the associated gRNA can have the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof.

In certain embodiments, the crRNA is not naturally-occurring. In some of these embodiments, the specific CRISPR repeat is not linked to the engineered spacer in nature and the CRISPR repeat is considered heterologous to the spacer. In certain embodiments, the spacer is an engineered sequence that is not naturally occurring.

The guide RNA can further comprise a trans-activating CRISPR RNA (tracrRNA). A tracrRNA molecule comprises a nucleotide sequence comprising a region that has sufficient complementarity to hybridize to a CRISPR repeat of a crRNA, which is referred to herein as the anti-repeat region. In some embodiments, the tracrRNA molecule further comprises a region with secondary structure (*e.g.*, stem-loop) or forms secondary structure upon hybridizing with its corresponding crRNA. In particular embodiments, the region of the tracrRNA that is fully or partially complementary to a CRISPR repeat sequence is at the 5' end of the molecule and the 3' end of the tracrRNA comprises secondary structure. This region of secondary structure generally comprises several hairpin structures, including the nexus hairpin, which is found adjacent to the anti-repeat sequence. The nexus forms the core of the interactions between the guide RNA and the RGN, and is at the intersection between the guide RNA, the RGN, and the target DNA. The nexus hairpin often has a conserved nucleotide sequence in the base of the hairpin stem, with the motif UNANNC found in many nexus hairpins in tracrRNAs. In some embodiments, a tracrRNA comprises a non-canonical sequence in the base of the hairpin stem of its nexus hairpin, including UNANNA, UNANNG, and CNANNC. There are often terminal hairpins at the 3' end of the tracrRNA that can vary in structure and number, but often comprise a GC-rich Rho-independent transcriptional terminator hairpin followed by a string of U's at the 3' end. See, for example, Briner et al. (2014) *Molecular Cell* 56:333-339, Briner and Barrangou (2016) *Cold Spring Harb Protoc*; doi: 10.1101/pdb.top090902, and U.S. Publication No. 2017/0275648, each of which is herein incorporated by reference in its entirety.

In various embodiments, the anti-repeat region of the tracrRNA that is fully or partially complementary to the CRISPR repeat comprises from about 8 nucleotides to about 30 nucleotides, or more. For example, the region of base pairing between the tracrRNA anti-repeat and the CRISPR repeat can be about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In particular embodiments, the region of base pairing between the

tracrRNA anti-repeat and the CRISPR repeat is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In some embodiments, the degree of complementarity between a CRISPR repeat and its corresponding tracrRNA anti-repeat, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a CRISPR repeat and its corresponding tracrRNA anti-repeat, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more.

In various embodiments, the entire tracrRNA can comprise from about 60 nucleotides to more than about 210 nucleotides. For example, the tracrRNA can be about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, about 125, about 130, about 135, about 140, about 150, about 160, about 170, about 180, about 190, about 200, about 210, or more nucleotides in length. In particular embodiments, the tracrRNA is 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 150, 160, 170, 180, 190, 200, 210 or more nucleotides in length. In particular embodiments, the tracrRNA is about 70 to about 105 nucleotides in length, including about 70, about 71, about 72, about 73, about 74, about 75, about 76, about 77, about 78, about 79, about 80, about 81, about 82, about 83, about 84, about 85, about 86, about 87, about 88, about 89, about 90, about 91, about 92, about 93, about 94, about 95, about 96, about 97, about 98, about 99, about 100, about 101, about 102, about 103, about 104, and about 105 nucleotides in length. In particular embodiments, the tracrRNA is 70 to 105 nucleotides in length, including 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, and 105 nucleotides in length.

In particular embodiments, the tracrRNA comprises the nucleotide sequence of SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof that when comprised within a guide RNA is capable of directing the sequence-specific binding of an associated RNA-guided nuclease provided herein to a target sequence of interest. In certain embodiments, an active tracrRNA sequence variant of a wild-type sequence comprises a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the nucleotide sequences set forth as SEQ ID NO: 34, 246, 247, or 248. In certain embodiments, an active tracrRNA sequence fragment of a wild-type sequence comprises at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248

In those embodiments wherein the RGN has the amino acid sequence set forth as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285, the tracrRNA of the

associated gRNA can have the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof. An active variant or fragment of a variant LPG10145 RGN disclosed herein can bind a tracrRNA comprising the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof.

Two polynucleotide sequences can be considered to be substantially complementary when the two sequences hybridize to each other under stringent conditions. Likewise, an RGN is considered to bind to a particular target sequence within a sequence-specific manner if the guide RNA bound to the RGN binds to the target sequence under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which the two polynucleotide sequences will hybridize to each other to a detectably greater degree than to other sequences (*e.g.*, at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. Typically, stringent conditions will be those in which the salt concentration is less than about 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is at least about 30°C for short sequences (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long sequences (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37°C, and a wash in 1X to 2X SSC (20X SSC = 3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55°C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37°C, and a wash in 0.5X to 1X SSC at 55 to 60°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60 to 65°C. Optionally, wash buffers may comprise about 0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours. The duration of the wash time will be at least a length of time sufficient to reach equilibrium.

The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched sequence. For DNA-DNA hybrids, the Tm can be approximated from the equation of Meinkoth and Wahl (1984) Anal. Biochem. 138:267-284: Tm = 81.5°C + 16.6 (log M) + 0.41 (%GC) - 0.61 (% form) - 500/L; where M is the molarity of monovalent cations, %GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence and its complement at a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4°C lower than the thermal melting point (Tm); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10°C lower than the thermal melting point (Tm); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20°C lower than the thermal melting point (Tm). Using the equation, hybridization and wash compositions, and desired Tm, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash

solutions are inherently described. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, Part I, Chapter 2 (Elsevier, New York); and Ausubel et al., eds. (1995) Current Protocols in Molecular Biology, Chapter 2 (Greene Publishing and Wiley-Interscience, New York). See Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

The term "sequence specific" can also refer to the binding of a RGN polypeptide or RGN polypeptide fusion to a target sequence at a greater affinity than binding to a randomized background sequence.

The guide RNA can be a single guide RNA (sgRNA) or a dual-guide RNA (dgRNA). A single guide RNA comprises the crRNA and tracrRNA on a single molecule of RNA, whereas a dual-guide RNA system comprises a crRNA and a tracrRNA present on two distinct RNA molecules, hybridized to one another through at least a portion of the CRISPR repeat sequence of the crRNA and at least a portion of the tracrRNA, which may be fully or partially complementary to the CRISPR repeat sequence of the crRNA. In some of those embodiments wherein the guide RNA is a single guide RNA, the crRNA and tracrRNA are separated by a linker nucleotide sequence. In general, the linker nucleotide sequence is one that does not include complementary bases in order to avoid the formation of secondary structure within or comprising nucleotides of the linker nucleotide sequence. In some embodiments, the linker nucleotide sequence between the crRNA and tracrRNA is at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, or more nucleotides in length. In particular embodiments, the linker nucleotide sequence of a single guide RNA is at least 4 nucleotides in length. In certain embodiments, the linker nucleotide sequence is the nucleotide sequence AAAG.

In some embodiments, the guide RNA is a single guide RNA (sgRNA) having the backbone sequence (comprising a crRNA repeat, an optional linker nucleotide sequence, and a tracrRNA) of any one of SEQ ID NOs: 249-252, or an active variant or fragment thereof. In certain embodiments, an active sgRNA backbone sequence variant comprises a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the nucleotide sequence set forth as any one of SEQ ID NOs: 249-252. In certain embodiments, an active sgRNA backbone sequence fragment comprises at least 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more contiguous nucleotides of the nucleotide sequence set forth as any one of SEQ ID NOs: 249-252. In those embodiments wherein the RGN has the amino acid sequence set forth as SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285, the sgRNA backbone can have the nucleotide sequence set forth as any one of SEQ ID NOs: 249-252, or an active variant or fragment thereof.

The single guide RNA or dual-guide RNA can be synthesized chemically or via *in vitro* transcription. Assays for determining sequence-specific binding between an RGN and a guide RNA are known in the art and include, but are not limited to, *in vitro* binding assays between an expressed RGN and the guide RNA, which can be tagged with a detectable label (e.g., biotin) and used in a pull-down detection assay in which the guide RNA:RGN complex is captured via the detectable label (e.g., with streptavidin beads). A control guide RNA with an unrelated sequence or structure to the guide RNA can be used as a negative control for non-specific binding of the RGN to RNA.

In certain embodiments, the guide RNA can be introduced into a target cell, organelle, or embryo as an RNA molecule. The guide RNA can be transcribed *in vitro* or chemically synthesized. In other embodiments, a nucleotide sequence encoding the guide RNA is introduced into the cell, organelle, or embryo. In some of these embodiments, the nucleotide sequence encoding the guide RNA is operably linked to a promoter (*e.g.*, an RNA polymerase III promoter). The promoter can be a native promoter or heterologous to the guide RNA-encoding nucleotide sequence.

In various embodiments, the guide RNA can be introduced into a target cell, organelle, or embryo as a ribonucleoprotein complex, as described herein, wherein the guide RNA is bound to an RNA-guided nuclease polypeptide.

The guide RNA directs an associated RNA-guided nuclease to a particular target nucleotide sequence of interest through hybridization of the guide RNA to the target nucleotide sequence. A target nucleotide sequence can comprise DNA, RNA, or a combination of both and can be single-stranded or double-stranded. A target nucleotide sequence can be genomic DNA (*i.e.*, chromosomal DNA), plasmid DNA, or an RNA molecule (*e.g.*, messenger RNA, ribosomal RNA, transfer RNA, micro RNA, small interfering RNA). The target nucleotide sequence can be bound (and in some embodiments, cleaved) by an RNA-guided nuclease *in vitro* or in a cell. The chromosomal sequence targeted by the RGN can be a nuclear, plastid or mitochondrial chromosomal sequence. In some embodiments, the target nucleotide sequence is unique in the target genome. In some embodiments, the target sequence is double-stranded and comprises a target strand and a non-target strand.

The target sequence is adjacent to a protospacer adjacent motif (PAM) and the target strand of the target sequence is the strand that comprises the PAM. The PAM is immediately adjacent to the target sequence and often comprise Ns, which represent any nucleotide. In some embodiments, the PAM comprises about 1 to about 10 Ns, including about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, or about 10 Ns. In some embodiments, a PAM comprises 1 to 10 Ns, including 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 Ns. In general, the PAM can be 5' or 3' of the target sequence on its non-target strand. In some embodiments, the PAM is 3' of the target sequence on its non-target strand for the presently disclosed guide RNAs and RGN systems. Generally, the PAM is a consensus sequence of about 3-4 nucleotides, but in some embodiments, it can be 2, 3, 4, 5, 6, 7, 8, 9, or more nucleotides in length. In some embodiments,

the PAM sequence recognized by the presently disclosed variant LPG10145 RGNs comprises the consensus sequence set forth as NNGG.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In particular embodiments, a variant LPG10145 RGN having an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid

sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof, binds a target nucleotide sequence adjacent to a PAM sequence set forth as NNGG.

In particular embodiments, a variant LPG10145 RGN having an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof, binds a target nucleotide sequence adjacent to a PAM sequence set forth as NNGG.

In some embodiments, the variant LPG10145 RGN binds to a guide sequence comprising a CRISPR repeat set forth as SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof, and a tracrRNA sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof. The RGN systems are described further in Examples 1-6 of the present specification.

It is well-known in the art that PAM sequence specificity for a given nuclease enzyme is affected by enzyme concentration (*see, e.g.*, Karvelis *et al.* (2015) *Genome Biol* 16:253), which may be modified by altering the promoter used to express the RGN, or the amount of ribonucleoprotein complex delivered to the cell, organelle, or embryo.

Upon recognizing its corresponding PAM sequence, the RGN, if active, can cleave the target nucleotide sequence at a specific cleavage site. As used herein, a cleavage site is made up of the two particular nucleotides within a target nucleotide sequence at which the target strand, non-target strand, or both strands of the target nucleotide sequence is cleaved by an RGN. The cleavage site can comprise the 1st and 2nd, 2nd and 3rd, 3rd and 4th, 4th and 5th, 5th and 6th, 7th and 8th, or 8th and 9th nucleotides from the PAM in either the 5' or 3' direction. In some embodiments, the cleavage site may be over 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides from the PAM in either the 5' or 3' direction. As RGNs can cleave a target nucleotide sequence resulting in staggered ends, in some embodiments, the cleavage site is defined based on the distance of the two nucleotides from the PAM on the non-target strand of the target polynucleotide and for the target strand, the distance of the two nucleotides from the complement of the PAM.

V. Polymerases

Compositions of the disclosure, including fusion proteins, polymerase editors (PEs), and PE systems comprising engineered variant LPG10145 RGN polypeptides, can comprise polymerases (e.g., DNA polymerases, reverse transcriptases). As used herein, a "polymerase" is an enzyme that catalyzes the formation of a nucleic acid polymer. A polymerase can be an RNA polymerase (catalyzing an RNA polymer) or a DNA polymerase (catalyzing a DNA polymer). In some embodiments, the polymerase of the

polymerase editor or system is a DNA polymerase. The PE or PE system can comprise a DNA-dependent DNA polymerase (uses DNA as a template) or an RNA-dependent DNA polymerase (uses RNA as a template). In some embodiments, the DNA polymerase of the presently disclosed PEs and PE systems is an RNA-dependent DNA polymerase (i.e., reverse transcriptase).

Reverse transcriptases (RTs) are a class of enzymes that catalyze the transcription of RNA into DNA, a process known as reverse transcription. This enzymatic activity is critical in the life cycles of retroviruses, such as Human Immunodeficiency Virus (HIV), and in the replication of various mobile genetic elements, including retrotransposons. First, the RT uses its RNA-dependent DNA polymerase activity to convert single-stranded RNA (ssRNA) templates into complementary DNA (cDNA). RTs can also possess RNase H activity, which degrades the RNA strand of an RNA-DNA hybrid, providing a template for the synthesis of the second DNA strand. The RT then synthesizes the second DNA strand through its DNA-dependent DNA polymerase activity, resulting in a double-stranded DNA (dsDNA) molecule that can integrate into the host genome. As used herein, a "reverse transcriptase" or "RT" is an enzyme that has polymerase activity to catalyze the formation of a nucleic acid polymer. In some embodiments, the RT synthesizes a nucleic acid polymer using a template nucleic acid molecule. In some embodiments, the polymerase activity is a DNA polymerase activity. In some embodiments, an RT catalyzes the addition of nucleotides to a nicked polynucleotide strand, using a template.

RTs include retroviral RTs such as HIV-1 RT, hepatitis B RT, and Murine Leukemia Virus (MLV)-RT (also known as Moloney Murine Leukemia Virus (MMLV)-RT). MLV-RT serves as a model for understanding the basic mechanisms of reverse transcription. Reverse transcriptases typically exhibit a "right hand" structure with three main domains: a 'finger' domain involved in binding the template-primer and dNTPs; the 'palm' domain containing the active site with highly conserved motifs responsible for catalysis; and the 'thumb', which maintains the enzyme's interaction with the nucleic acid substrate. RTs also have distinct polymerase and RNase H domains, where the polymerase domain is responsible for nucleic acid molecule (e.g., DNA) synthesis, and the RNase H domain degrades the RNA strand in RNA-DNA hybrids to allow second-strand DNA synthesis. RTs are important tools in molecular biology and biotechnology, with uses in RT-PCR, cDNA synthesis from RNA, amplification and quantification of RNA, RNA Sequencing (RNA-Seq), preparing cDNA libraries from RNA samples, and gene cloning and expression studies.

A polymerase (e.g., reverse transcriptase) of the disclosure includes but is not limited to the polymerases (e.g., reverse transcriptases) described herein and variants or fragments thereof, including but not limited to a reverse transcriptase comprising an amino acid sequence having at least 40%, at least 41%, at least 42%, at least 43%, at least 44%, at least 45%, at least 46%, at least 47%, at least 48%, at least 49%, at least 50%, at least 51%, at least 52%, at least 53%, at least 54%, at least 55%, at least 56%, at least 57%, at least 58%, at least 59%, at least 60%, at least 61%, at least 62%, at least 63%, at least 64%, at least 65%, at least 66%, at least 67%, at least 68%, at least 69%, at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%,

at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 99.1%, at least 99.2%, at least 99.3%, at least 99.4%, at least 99.5%, at least 99.6%, at least 99.7%, at least 99.8%, or at least 99.9% sequence identity to the amino acid sequence set forth as SEQ ID NO: 254 or 255.

An RT of the presently disclosed compositions and methods can lack an RNase H domain.

VI. Polymerase Editors (PEs)

As used herein, a "polymerase editor" or "PE" refers to a protein or a plurality of proteins comprising an RGN polypeptide and a polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) that, along with a polymerase editing guide RNA (PEgRNA) that comprises an extension arm comprising a primer binding site (PBS) and a DNA synthesis template comprising a desired edit, is capable of editing a double-stranded polynucleotide through the replacement of a target sequence using the DNA synthesis template as a template for the polymerase. In certain embodiments, the RGN polypeptide and the polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) are operably linked (by fusion or insertion). In other embodiments, the RGN polypeptide and the polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) are not operably linked. In one particular embodiment, the RGN polypeptide and the polymerase editing polypeptide (e.g., DNA polymerase or reverse transcriptase) are two separate polypeptides. In some embodiments, the PE does not require the introduction of a double-stranded break, but rather utilizes an RGN nickase that nicks the non-target strand upstream of the sequence to be edited and upstream of the PAM, creating a 3' flap on the non-target strand. The PBS of the PEgRNA is complementary to the 3' flap of the non-target strand and hybridrization of the PBS and 3' flap of the non-target strand allows for the polymerization of the replacement strand containing the edit using the DNA synthesis template and polymerase. Those polymerase editors that utilize a reverse transcriptase as the polymerase are referred to herein as "RT editors" or "RTEs".

The presently disclosed polymerase editors (PEs) comprise a polymerase (e.g., RT) and an engineered variant LPG10145 RGN polypeptide. The RT includes RTs, or active variants or fragments thereof, as described herein, including but not limited to an RT comprising an amino acid sequence having at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or more sequence identity to the amino acid sequence set forth as SEQ ID NO: 254 or 255.

The engineered variant LPG10145 RGN polypeptides, or active variants or fragments thereof, include those described herein. The binding and/or cleaving activity of an active variant or fragment of a variant LPG10145 RGN disclosed herein can be dependent upon recognizing a protospacer adjacent motif (PAM) adjacent and 3' to the target sequence. In some embodiments, the PAM comprises a consensus

sequence of NNGG. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid

residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid

sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

1. Various Formats of Polymerase Editors

The presently disclosed PEs can be provided *in trans*, wherein the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide are separate polypeptides. In some embodiments, the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide are transcribed together and have a sequence encoding a self-cleaving peptide (e.g., 2A peptide such as P2A) in between, such that translation results in two separate polypeptides. Any self-cleaving peptide known in the art can be used in such embodiments, including but not limited to, 2A peptides, which is a class of 18-22 amino acid long peptides that may function through ribosomal skipping during translation. Non-limiting examples of 2A peptides are T2A, P2A, E2A, and F2A.

In some embodiments, the presently disclosed PEs can comprise a polymerase (e.g., RT) operably fused to a variant LPG10145 RGN polypeptide, wherein the RT and variant LPG10145 RGN polypeptide are fused to each other end-to-end or wherein the polymerase (e.g., RT) is inserted into the variant

LPG10145 RGN polypeptide, such as those inlaid base editors described in International Appl. Publ. No. WO 2024/095245, which is herein incorporated by reference in its entirety. In an end-to-end fusion, the polymerase (e.g., RT) can be fused to the amino terminus of the variant LPG10145 RGN polypeptide or the carboxy terminus of the variant LPG10145 RGN polypeptide. The presently disclosed PEs can comprise from amino terminus to carboxy terminus: the polymerase (e.g., RT) and the variant LPG10145 RGN polypeptide; or the variant LPG10145 RGN polypeptide and the polymerase (e.g., RT).

In those embodiments wherein the polymerase (e.g., RT) is inserted within a variant LPG10145 RGN polypeptide, the polymerase (e.g., RT) is inserted between surface amino acid residues of the variant LPG10145 RGN polypeptide. The polymerase (e.g., RT) can be inserted within or between a linker domain 2, a wedge (WED) domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting (PI) domain. In some embodiments, the RuvC domain is the RuvCIII domain. A Rec or recognition lobe mediates nucleic acid binding through multiple Rec domains (e.g., Rec1-3) by sensing nucleic acids, regulates the HNH conformational transition, and locks the catalytic HNH domain at the cleavage site. A wedge domain is responsible for the recognition of guide RNA scaffolds. A PAM-interacting domain is the domain of an RGN polypeptide that binds to a PAM site. Non-limiting examples of domains within LPG10145 RGN (SEQ ID NO: 1) or a variant LPG10145 RGN (SEQ ID NOs: 2-16, 124-126, 182-196, and 271-285) include: RuvC-I from amino acid residues 1 to 42; BH from amino acid residues 43 to 79; REC1 from amino acid residues 80 to 236; REC2 from amino acid residues 237 to 476; RuvC-II from amino acid residues 477 to 524; L1 from amino acid residues 525 to 560; HNH from amino acid residues 561 to 676; L2 from amino acid residues 677 to 690; RuvC-III from amino acid residues 691 to 828; WED from amino acid residues 829 to 976; and PI from amino acid residues 977 to 1130. The general domains of RGN polypeptides can be determined via structural comparison to RGN polypeptides with defined domains.

In those embodiments wherein the PE comprises a variant LPG10145 RGN polypeptide comprising the amino acid sequence set forth as SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285, or an active variant or fragment thereof, the polymerase (e.g., RT) can be inserted within the variant LPG10145 RGN polypeptide, or an active variant or fragment thereof, immediately after the amino acid position selected from the group consisting of: i) amino acid position corresponding to position 347 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; ii) amino acid position corresponding to position 524 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; iii) amino acid position corresponding to position 640 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; iv) amino acid position corresponding to position 666 of SEQ ID

NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; v) amino acid position corresponding to position 680 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; vi) amino acid position corresponding to position 740 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; vii) amino acid position corresponding to position 785 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; viii) amino acid position corresponding to position 910 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285; and ix) amino acid position corresponding to position 1077 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, or 285.

The polymerase (e.g., RT) may be fused directly to the variant LPG10145 RGN polypeptide or a peptide linker can connect the polymerase (e.g., RT) and the variant LPG10145 RGN polypeptide. In those embodiments wherein the polymerase (e.g., RT) is inserted into the variant LPG10145 RGN polypeptide, there can be peptide linkers on one or both ends of the polymerase (e.g., RT). Any suitable peptide linker can be used to connect the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide, but one suitable peptide linker comprises one or more copies of SGGS (SEQ ID NO: 241). In some embodiments, the peptide linker comprises 1 SGGS (SEQ ID NO: 241) sequence, 2 SGGS (SEQ ID NO: 241) sequences, 3 SGGS (SEQ ID NO: 241) sequences, 4 SGGS (SEQ ID NO: 241) sequences, or more, such that the linker sequence can be 4, 8, 12, or 16 amino acids long. The linker between the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide can be 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, or more amino acids in length. The peptide linker separating the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide can also comprise an NLS, such as but not limited to those disclosed elsewhere herein, including SEQ ID NO: 36, 37, 234, or 235, wherein the NLSs can be connected by peptide linkers (such as SGGS, SEQ ID NO: 241). In some embodiments, the peptide linker separating the polymerase (e.g., RT) and variant LPG10145 RGN polypeptide comprises more than one localization sequence, such as 2, 3, or more localization sequences. In some embodiments, a peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1. In certain embodiments, the peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x or y is 0, 1, 2, 3, or 4, and y is 0; and wherein one of m or n is 0, and the other is 1. In other embodiments, the peptide linker has a

formula of -$(SGGS)_x$-NLS_m-$(SGGS)_y$-NLS_n-$(SGGS)_z$-, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 1.

The presently disclosed polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs can comprise at least one nuclear localization signal (NLS) to enhance transport of the protein to the nucleus of a cell. Nuclear localization signals are known in the art and generally comprise a stretch of basic amino acids (see, *e.g.*, Lange *et al.*, *J. Biol. Chem.* (2007) 282:5101-5105). In some embodiments, the polymerase (e.g., RT), variant LPG10145 RGN polypeptide, fusion protein, or PE comprises 2, 3, 4, 5, 6 or more nuclear localization signals. The nuclear localization signal(s) can be a heterologous NLS. Non-limiting examples of nuclear localization signals useful for the presently disclosed polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs are the nuclear localization signals of SV40 Large T-antigen, nucleoplasmin, and c-Myc (*see, e.g.*, Ray et al. (2015) *Bioconjug Chem* 26(6):1004-7). In some embodiments, the polymerase (e.g., RT), variant LPG10145 RGN polypeptide, fusion protein, or PE comprises the NLS sequence set forth as SEQ ID NO: 36, 37, 234, and/or 235. The polymerase (e.g., RT), variant LPG10145 RGN polypeptide, fusion protein, or PE can comprise one or more NLS sequences at its N-terminus, C- terminus, or both the N-terminus and C-terminus. For example, the polymerase (e.g., RT), variant LPG10145 RGN polypeptide, fusion protein, or PE can comprise two NLS sequences at the N-terminal region and four NLS sequences at the C-terminal region. In some embodiments, a peptide linker can connect the NLS to the polymerase (e.g., RT), variant LPG10145 RGN polypeptide, or fusion protein.

Other localization signal sequences known in the art that localize polypeptides to particular subcellular location(s) can also be used to target the polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs, including, but not limited to, plastid localization sequences, mitochondrial localization sequences, and dual-targeting signal sequences that target to both the plastid and mitochondria (see, *e.g.*, Nassoury and Morse (2005) *Biochim Biophys Acta* 1743:5-19; Kunze and Berger (2015) *Front Physiol* dx.doi.org/10.3389/fphys.2015.00259; Herrmann and Neupert (2003) *IUBMB Life* 55:219-225; Soll (2002) *Curr Opin Plant Biol* 5:529-535; Carrie and Small (2013) *Biochim Biophys Acta* 1833:253-259; Carrie *et al.* (2009) *FEBS J* 276:1187-1195; Silva-Filho (2003) *Curr Opin Plant Biol* 6:589-595; Peeters and Small (2001) *Biochim Biophys Acta* 1541:54-63; Murcha *et al.* (2014) *J Exp Bot* 65:6301-6335; Mackenzie (2005) *Trends Cell Biol* 15:548-554; Glaser *et al.* (1998) *Plant Mol Biol* 38:311-338).

Polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs can comprise at least one cell-penetrating domain that facilitates cellular uptake of the polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs. Cell-penetrating domains are known in the art and generally comprise stretches of positively charged amino acid residues (*i.e.*, polycationic cell-penetrating domains), alternating polar amino acid residues and non-polar amino acid residues (*i.e.*, amphipathic cell-penetrating domains), or hydrophobic amino acid residues (*i.e.*, hydrophobic cell-penetrating domains) (*see,*

e.g., Milletti F. (2012) *Drug Discov Today* 17:850-860). A non-limiting example of a cell-penetrating domain is the trans-activating transcriptional activator (TAT) from the human immunodeficiency virus 1.

The nuclear localization signal, plastid localization signal, mitochondrial localization signal, dual-targeting localization signal, and/or cell-penetrating domain can be located at the amino-terminus (N-terminus), the carboxyl-terminus (C-terminus), and/or in an internal location of the polymerase (e.g., RT), variant LPG10145 RGN polypeptide, fusion protein, or PE.

Polymerases (e.g., RTs), variant LPG10145 RGN polypeptides, fusion proteins, or PEs can also comprise a purification tag, which is any molecule that can be utilized to isolate a protein or fused protein from a mixture (e.g., biological sample, culture medium). Non-limiting examples of purification tags include biotin, myc, maltose binding protein (MBP), glutathione-S-transferase (GST), and 3X FLAG tag.

A PE of the present disclosure can comprise an amino acid sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set forth as any one of SEQ ID NOs: 135, and 162-181. In some embodiments, a PE of the disclosure comprises the amino acid sequence set forth as any one of SEQ ID NOs: 135, and 162-181.

2. PEgRNA

A PE system utilizes a polymerase editing guide RNA ("PEgRNA"). The PEgRNA is a guide RNA that both specifies the target sequence and provides the template for polymerization of the replacement strand containing a desired edit by way of an extension engineered onto the RGN guide RNA or a part thereof, referred to herein as an extension arm. The PEgRNA can be a single guide RNA, wherein the extension arm can be at the 5' or 3' end, or at an internal portion of the guide RNA, or multiple polynucleotides (e.g., a dual guide RNA). In embodiments wherein the PEgRNA is a dual guide RNA, the extension arm can be at the 5' or 3' end, or at an internal portion of the crRNA or tracrRNA molecule. The template for polymerization within an extension arm is referred to herein as the DNA synthesis template. In those embodiments wherein the polymerase of the PE is an RT, the DNA synthesis template can be referred to as the reverse transcriptase template (RTT). The RGN is guided to the target sequence by the PEgRNA and in those embodiments wherein the RGN is a nickase with an inactivated HNH domain and active RuvC domain, the RGN nickase nicks the non-target strand upstream of the sequence to be edited and upstream of the PAM, creating a 3' flap on the non-target strand. The PEgRNA includes a primer binding site (PBS) that is complementary to the 3' flap of the non-target strand. The PBS can be at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 nucleotides in length. In certain embodiments, the PEgRNA comprises a PBS that is at least 5 (e.g., 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20) nucleotides in length. In some embodiments, the PEgRNA may comprise a PBS that is 9, 11, 12, 13, 15, or 18 nucleotides in length. Hybridization of the PBS and 3′ flap of the non-target strand allows polymerization of

the replacement strand containing the edit using the DNA synthesis template in the extension of the PEgRNA. The DNA synthesis template can be at least about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, or more nucleotides in length. In certain embodiments, the PEgRNA comprises a DNA synthesis template that is at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 31, at least 32, at least 33, at least 34, at least 35, at least 36, at least 37, at least 38, at least 39, at least 40, at least 41, at least 42, at least 43, at least 44, at least 45, at least 46, at least 47, at least 48, at least 49, or at least 50 nucleotides in length. In some embodiments, the PEgRNA comprises a DNA synthesis template that is 19, 22, 23, 24, 25, 26, 29, 30, 32, 34, 37, 38, 39, 40, 42, or 46 nucleotides in length. The DNA synthesis template comprises the desired edit, which can be a substitution of one or more nucleotides, a deletion of one or more nucleotides, or an addition of one or more nucleotides.

The extension arm of the PEgRNA can be formed from RNA or DNA. In the case of an RNA extension, the polymerase of the polymerase editor can be an RNA-dependent DNA polymerase (such as a reverse transcriptase). In the case of a DNA extension, the polymerase of the polymerase editor may be a DNA-dependent DNA polymerase.

The replacement strand containing the desired edit (e.g., substitution, deletion, or addition) shares the same sequence as the non-target strand of the target sequence to be edited (with the exception that it includes the desired edit). Through DNA repair and/or replication machinery, the non-target strand of the target sequence is replaced by the newly synthesized replacement strand containing the desired edit. In some cases, polymerase editing may be thought of as a "search-and-replace" genome editing technology since the polymerase editors not only search and locate the desired target sequence to be edited, but at the same time, encode a replacement strand containing a desired edit which is installed in place of the corresponding non-target strand of the target sequence. Thus, in some embodiments, a guide RNA of the disclosure comprises an extension comprising an edit template for polymerase editing.

An RTT of the present disclosure can comprise a nucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as any one of SEQ ID NOs: 226-233. In some embodiments, an RTT of the present disclosure comprises the nucleotide sequence set forth as any one of SEQ ID NOs: 226-233. A PBS of the present disclosure can comprise a nucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as any one of SEQ ID NOs: 221-223. In some embodiments, a PBS of the present disclosure comprises the nucleotide sequence set forth as any one of SEQ ID NOs: 221-223. A PEgRNA of the present disclosure can comprise a nucleotide sequence having at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the

nucleotide sequence set forth as any one of SEQ ID NOs: 129-132, and 197-220. In some embodiments, a PEgRNA of the present disclosure comprises the nucleotide sequence set forth as any one of SEQ ID NOs: 129-132, and 197-220.

3. Nicking guide RNA

In order to reduce the possibility that the edit introduced by a polymerase editor is removed due to mismatch repair of the edited strand, a nicking guide RNA can be used. A "nicking guide RNA" is a guide RNA that targets a sequence within the unedited strand at a site nearby and opposite to the original nick and guides the RGN nickase of the PE system to this unedited strand to introduce a single-stranded nick. The nicking guide RNA can be designed to match the edited sequence introduced by the PEgRNA, but not the original unedited sequence, to ensure that the nicking occurs after the editing event on the non-target strand takes place.

VII. Nucleotides Encoding RNA-guided nucleases, base editing polypeptides, polymerases, fusion proteins, base editors, polymerase editors, CRISPR RNA, tracrRNA, and/or guide RNA

The present disclosure provides polynucleotides comprising the presently disclosed CRISPR RNAs, tracrRNAs, and/or sgRNAs and polynucleotides comprising a nucleotide sequence encoding the presently disclosed variant LPG10145 RGNs, CRISPR RNAs, tracrRNAs, sgRNAs, and/or PEgRNAs. Systems of the disclosure can comprise polynucleotides comprising or encoding guide RNAs or PEgRNAs and polynucleotides comprising a nucleotide sequence encoding variant LPG10145 RGN polypeptides, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, and/or PEs. Presently disclosed polynucleotides include those comprising or encoding a CRISPR repeat comprising the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof that when comprised within a guide RNA is capable of directing the sequence-specific binding of an associated RNA-guided nuclease to a target sequence of interest. Also disclosed are polynucleotides comprising or encoding a tracrRNA comprising the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof that when comprised within a guide RNA is capable of directing the sequence-specific binding of an associated RNA-guided nuclease to a target sequence of interest.

Polynucleotides are also provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. Polynucleotides are provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871,

872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. Polynucleotides are provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

Polynucleotides are also provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. Polynucleotides are provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. Polynucleotides are provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

Polynucleotides are also provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, polynucleotides are provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

Polynucleotides are also provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, polynucleotides are also provided that encode an RNA-guided nuclease comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, a polynucleotide is provided that encodes an RNA-guided nuclease comprising the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid

position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

In some embodiments, a polynucleotide is provided that encodes an RNA-guided nuclease comprising the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set

forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, and active fragments or variants thereof that retain the ability to bind to a target nucleotide sequence in an RNA-guided sequence-specific manner.

The use of the term "polynucleotide" or "nucleic acid molecule" is not intended to limit the present disclosure to polynucleotides comprising DNA. Those of ordinary skill in the art will recognize that polynucleotides can comprise ribonucleotides (RNA) and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. These include peptide nucleic acids (PNAs), PNA-DNA chimers, locked nucleic acids (LNAs), and phosphothiorate linked sequences. The polynucleotides disclosed herein also encompass all forms of sequences including, but not limited to, single-stranded forms, double-stranded forms, DNA-RNA hybrids, triplex structures, stem-and-loop structures, and the like.

In some embodiments, the polynucleotide encoding a presently disclosed variant LPG10145 RGN polypeptide, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptides, base editors, and/or PE is an mRNA (messenger RNA) molecule. An mRNA refers to any polynucleotide which encodes a polypeptide of interest and which is capable of being translated to produce the encoded polypeptide of interest *in vitro*, *in vivo*, *in situ*, or *ex vivo*. In embodiments, the basic components of an mRNA molecule include at least a coding region, a 5′UTR, a 3′UTR, a 5′ cap and a poly-A tail. A 5′ UTR, situated 5′ of a coding sequence and transcribed as part of an mRNA, may comprise various regulatory elements, including, e.g., 5′ cap structure, G-quadruplex structure (G4), stem-loop structure, and internal ribosome entry sites (IRES), which can control translation initiation of the mRNA. A 3′ UTR, situated 3′ of a coding sequence and transcribed as part of an mRNA, can be involved in numerous regulatory processes including transcript cleavage, stability and polyadenylation, translation, and mRNA localization. The 3′ UTR can serve as a binding site for numerous regulatory proteins and small non-coding RNAs, e.g., microRNAs. A 5′ UTR and/or a 3′ UTR heterologous to an mRNA originates from an organism or species that is different from that of the mRNA, or if from the same organism or species as the mRNA, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention.

In some embodiments, inclusion of a 5′ UTR and/or a 3′ UTR heterologous to an mRNA encoding a variant LPG10145 RGN polypeptide, fusion protein comprising same, polymerase (e.g., RT), and/or PE of the disclosure improves polypeptide synthesis from the mRNA in a tissue (e.g., liver, or cells *in vitro*, such as stem cells, hepatocytes or lymphocytes). Heterologous 5′ UTRs and/or 3′ UTRs may, for example, increase protein synthesis by increasing the time that the mRNA remains in translating polysomes (message stability) and/or the rate at which ribosomes initiate translation on the mRNA (message translation efficiency). Thus, inclusion of a 5′ UTR and/or a 3′ UTR heterologous to an mRNA encoding variant LPG10145 RGN polypeptide, fusion protein comprising the same, polymerase (e.g., RT), base editing

polypeptides, base editors, and/or PE of the disclosure can lead to prolonged and/or increased polypeptide synthesis, enabling improved editing of a target polynucleotide by the PE or PE system. In some embodiments, the enhanced polypeptide synthesis from an mRNA occurs in a tissue-specific manner. Heterologous UTR sequences are described, for example, in US 2023/0050143 and US 2017/0252461.

In some embodiments, an mRNA encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptides, base editors, and/or PE useful in the presently disclosed methods and compositions can include one or more structural and/or chemical modifications or alterations which impart useful properties to the polynucleotide. For instance, a useful property of an mRNA includes the lack of a substantial induction of the innate immune response of a cell into which the mRNA is introduced. A “structural” feature or modification is one in which two or more linked nucleotides are inserted, deleted, duplicated, inverted or randomized in an mRNA without significant chemical modification to the nucleotides themselves. Because chemical bonds will necessarily be broken and reformed to effect a structural modification, structural modifications are of a chemical nature and hence are chemical modifications. However, structural modifications will result in a different sequence of nucleotides. Chemical modifications to mRNA can involve inclusion of 5-methylcytosine, N1-methyl-pseudouridine, pseudouridine, 2-thiouridine, 4-thiouridine, 5-methoxyuridine, 2′Fluoroguanosine, 2′Fluorouridine, 5-bromouridine, 5-(2-carbomethoxyvinyl) uridine, 5-[3(1-E-propenylamino)] uridine, α-thiocytidine, N6-methyladenosine, 5-methylcytidine, N4-acetylcytidine, 5-formylcytidine, or combinations thereof, in an mRNA.

The nucleic acid molecules encoding variant LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, and/or PEs can be codon optimized for expression in an organism of interest. A "codon-optimized” coding sequence is a polynucleotide coding sequence having its frequency of codon usage designed to mimic the frequency of preferred codon usage or transcription conditions of a particular host cell. Expression in the particular host cell or organism is enhanced as a result of the alteration of one or more codons at the nucleic acid level such that the translated amino acid sequence is not changed. Nucleic acid molecules can be codon optimized, either wholly or in part. Codon tables and other references providing preference information for a wide range of organisms are available in the art (*see*, *e.g.*, Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of plant-preferred codon usage). Methods are available in the art for synthesizing plant-preferred genes or mammalian (for example human) codon-optimized coding sequences. See, for example, U.S. Patent Nos. 5,380,831, and 5,436,391, and Murray *et al.* (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

Polynucleotides encoding the variant LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, crRNAs, tracrRNAs, sgRNAs, and/or PEgRNAs provided herein can be provided in expression cassettes for *in vitro* expression or expression in a cell, organelle, embryo, or organism of interest. The cassette will include 5' and 3' regulatory sequences

operably linked to a polynucleotide encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptides, base editors, PE, crRNA, tracrRNAs, sgRNAs, and/or PEgRNAs provided herein that allows for expression of the polynucleotide. The cassette may additionally contain at least one additional gene or genetic element to be cotransformed into the organism. Where additional genes or elements are included, the components are operably linked. The term "operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a promoter and a coding region of interest (*e.g.*, region coding for an RGN, crRNA, tracrRNAs, sgRNAs, and/or PEgRNAs) is a functional link that allows for expression of the coding region of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by "operably linked" or "operably fused" is intended that the coding regions are in the same reading frame, even if one is inserted into another. In some embodiments, polypeptides that are "operably fused" or "operably linked" means that the structure and/or biological activity of each individual peptide is also present in the fusion. Alternatively, the additional gene(s) or element(s) can be provided on multiple expression cassettes. For example, the nucleotide sequence encoding a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE can be present on one expression cassette, whereas the nucleotide sequence encoding a crRNA, tracrRNA, or complete guide RNA (or PEgRNA) can be on a separate expression cassette. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of the polynucleotides to be under the transcriptional regulation of the regulatory regions. The expression cassette may additionally contain a selectable marker gene.

The expression cassette will include in the 5'-3' direction of transcription, a transcriptional (and, in some embodiments, translational) initiation region (*i.e.*, a promoter), a variant LPG10145 RGN-, a fusion protein-, a polymerase- (e.g., RT-), base editing polypeptide-, base editor-, PE-, crRNA-, tracrRNA-, sgRNA-, and/or PEgRNA- encoding polynucleotide of the invention, and a transcriptional (and in some embodiments, translational) termination region (*i.e.*, termination region) functional in the organism of interest. The promoters of the invention are capable of directing or driving expression of a coding sequence in a host cell. The regulatory regions (*e.g.*, promoters, transcriptional regulatory regions, and translational termination regions) may be endogenous or heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a regulatory region that is heterologous to another regulatory region or to the host cell, is a regulatory region that is not found with another regulatory region or in the host cell in nature. The heterologous regulatory region can originate from a foreign species or from the same species. The heterologous regulatory region can be in its native form or is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, a chimeric gene comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence. Similarly, a nucleic acid molecule that is heterologous to another nucleic acid molecule is a nucleic acid molecule that is not found with another nucleic acid molecule in nature. The heterologous

nucleic acid molecule can originate from a foreign species or from the same species. The heterologous nucleic acid molecule can be in its native form or is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. For example, an mRNA encoding a variant LPG10145 RGN polypeptide can comprise a heterologous 5' UTR, wherein the 5' UTR is not present as part of the mRNA encoding the variant LPG10145 RGN polypeptide in nature.

Convenient termination regions include ones from simian virus (SV40), human growth hormone (hGH), bovine growth hormone (BGH), and rabbit beta-globin (rbGlob). See also Proudfoot (1991) *Cell* 64:671-674; Munroe *et al.* (1990) *Gene* 91:151-158; Schek *et al.* (1992) *Molecular and Cellular Biology* 12(12):5386-5393; Gil and Proudfoot (1987) *Cell* 49(3):399-406; Goodwin and Rottman (1992) *The Journal of Biological Chemistry* 267(23):16330-16334; and Lanoix and Acheson (1988) *EMBO J.* 7(8): 2515-2522. Additional termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau *et al.* (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon *et al.* (1991) *Genes Dev.* 5:141-149; Mogen *et al.* (1990) *Plant Cell* 2:1261-1272; Munroe *et al.* (1990) *Gene* 91:151-158; Ballas *et al.* (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi *et al.* (1987) *Nucleic Acids Res.* 15:9627-9639.

Additional regulatory signals include, but are not limited to, transcriptional initiation start sites, operators, activators, enhancers, other regulatory elements, ribosomal binding sites, an initiation codon, termination signals, and the like. See, for example, U.S. Pat. Nos. 5,039,523 and 4,853,331; EPO 0480762A2; Sambrook et al. (1992) Molecular Cloning: A Laboratory Manual, ed. Maniatis et al. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), hereinafter "Sambrook 11"; Davis et al., eds. (1980) Advanced Bacterial Genetics (Cold Spring Harbor Laboratory Press), Cold Spring Harbor, N.Y., and the references cited therein.

In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, *in vitro* mutagenesis, primer repair, restriction, annealing, resubstitutions, *e.g.*, transitions and transversions, may be involved.

A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, inducible, growth stage-specific, cell type-specific, tissue-preferred, tissue-specific, or other promoters for expression in the organism of interest. See, for example, promoters set forth in WO 99/43838 and in US Patent Nos: 8,575,425; 7,790,846; 8,147,856; 8,586832; 7,772,369; 7,534,939; 6,072,050; 5,659,026; 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611; herein incorporated by reference.

Exemplary constitutive promoters for expression in cells of the present disclosure include: an SV40 early promoter; a mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter; a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE); a rous sarcoma virus (RSV) promoter; a human ubiquitin C promoter (UBC); a human U6 small nuclear promoter (U6); an enhanced U6 promoter; a human H1 promoter from RNA polymerase III (H1); a human elongation factor 1α promoter (EF1A); a human beta-actin promoter (ACTB); a human or mouse phosphoglycerate kinase 1 promoter (PGK); a chicken β-Actin promoter coupled with CMV early enhancer (CAGG); a yeast transcription elongation factor promoter (TEF1); and the like. See, for example, Miyagishi *et al.* (2002) *Nature Biotechnology* 20:497-500; Xia *et al.* (2003) Nucleic Acids Res. 31(17):e100-e100; Pasleau *et al.* (1985) *Gene* 38:227–232; Martin-Gallardo *et al.* (1988) *Gene* 70: 51–56; Oellig and Seliger (1990) *J Neurosci Res* 26: 390–396; Manthorpe *et al.* (1993) *Hum Gene Ther* 4: 419–431; Yew *et al.* (1997) *Hum Gene Ther* 8: 575–584; Xu *et al.* (2001) *Gene* 272: 149–156; Nguyen *et al.* (2008) *J Surg Res* 148: 60–66; Costa *et al.* (2005) *Nat Meth.* 2:259–260; Lam and Truong (2020) *ACS Synth. Biol.* 9(10):2625–2631.

For expression in plants, constitutive promoters also include CaMV 35S promoter (Odell *et al.* (1985) *Nature* 313:810-812); rice actin (McElroy *et al.* (1990) *Plant Cell* 2:163-171); ubiquitin (Christensen *et al.* (1989) *Plant Mol. Biol.* 12:619-632 and Christensen *et al.* (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.* (1991) *Theor. Appl. Genet.* 81:581-588); and MAS (Velten *et al.* (1984) *EMBO J.* 3:2723-2730).

Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, the PPDK promoter and the pepcarboxylase promoter which are both inducible by light. Also useful are promoters which are chemically inducible, such as the In2-2 promoter which is safener induced (U.S. Pat. No. 5,364,780), the Axig1 promoter which is auxin induced and tapetum specific but also active in callus (PCT US01/22169), the steroid-responsive promoters (see, for example, the ERE promoter which is estrogen induced, and the glucocorticoid-inducible promoter in Schena *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis *et al.* (1998) *Plant J.* 14(2):247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz *et al.* (1991) *Mol. Gen. Genet.* 227:229-237, and U.S. Pat. Nos. 5,814,618 and 5,789,156), herein incorporated by reference.

Tissue-specific or tissue-preferred promoters can be utilized to target expression of an expression construct within a particular tissue. In some embodiments, the tissue-specific or tissue-preferred promoters are active in mammalian tissue. Examples of tissue-specific or tissue-preferred promoters include promoters that initiate transcription preferentially in certain tissues, such as white blood cells (e.g., CD4 T cell), heart, kidney, liver, CNS, eye, pancreas, skeletal muscle, and testis. In certain embodiments, the tissue-specific or tissue-preferred promoters are active in plant tissue. Examples of promoters under developmental control in plants include promoters that initiate transcription preferentially in certain tissues, such as leaves, roots, fruit, seeds, or flowers. A "tissue specific" promoter is a promoter that initiates transcription only in certain

tissues. Unlike constitutive expression of genes, tissue-specific expression is the result of several interacting levels of gene regulation. As such, promoters from homologous or closely related plant species can be preferable to use to achieve efficient and reliable expression of transgenes in particular tissues. In some embodiments, the expression comprises a tissue-preferred promoter. A "tissue preferred" promoter is a promoter that initiates transcription preferentially, but not necessarily entirely or solely in certain tissues.

In some embodiments, the nucleic acid molecules encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, sgRNA, and/or PEgRNA comprise a cell type-specific promoter. A "cell type specific" promoter is a promoter that primarily drives expression in certain cell types in one or more organs. Some examples of cells in which cell type specific promoters may be primarily active include, for example, a primary cell, a neuronal cell, a glial cell, an adipocyte, a cardiomyocyte, a smooth muscle cell, a photoreceptor cell, and a retinal ganglia cell. Some examples of plant cells in which cell type specific promoters functional in plants may be primarily active include, for example, BETL cells, vascular cells in roots, leaves, stalk cells, and stem cells. The nucleic acid molecules can also include cell type preferred promoters. A "cell type preferred" promoter is a promoter that primarily drives expression mostly, but not necessarily entirely or solely in certain cell types in one or more organs. Some examples of cells in which cell type preferred promoters may be preferentially active include, for example, a primary cell, a neuron, an adipocyte, a cardiomyocyte, a smooth muscle cell, and a photoreceptor cell. Some examples of plant cells in which cell type preferred promoters functional in plants may be preferentially active include, for example, BETL cells, vascular cells in roots, leaves, stalk cells, and stem cells.

The nucleic acid sequences encoding the variant LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, crRNAs, tracrRNAs, sgRNAs, and/or PEgRNAs can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for example, for *in vitro* mRNA synthesis. In such embodiments, the *in vitro*-transcribed RNA can be purified for use in the methods described herein. For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. In such embodiments, the expressed protein and/or RNAs can be purified for use in the methods of genome modification described herein.

In certain embodiments, the polynucleotide encoding the variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, sgRNA, and/or PEgRNA also can be linked to a polyadenylation signal (*e.g.*, SV40 polyA signal and other signals functional in plants) and/or at least one transcriptional termination sequence. Additionally, the sequence encoding the RGN also can be linked to sequence(s) encoding at least one nuclear localization signal, at least one cell-penetrating domain, and/or at least one signal peptide capable of trafficking proteins to particular subcellular locations, as described elsewhere herein.

The polynucleotide encoding the variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, sgRNA, and/or PEgRNA can be present in a vector or multiple vectors. A "vector" refers to a polynucleotide composition for transferring, delivering, or introducing a nucleic acid into a host cell. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors (*e.g.*, lentiviral vectors, adeno-associated viral vectors, baculoviral vector). The vector can comprise additional expression control sequences (*e.g.*, enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences), selectable marker sequences (*e.g.*, antibiotic resistance genes), origins of replication, and the like. Additional information can be found in "Current Protocols in Molecular Biology" Ausubel *et al.*, John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001.

The vector can also comprise a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D). Marker genes can include genes that allow selection for growth on a particular nutrient or substance, such as dihydrofolate reductase (DHFR; Simonsen and Levinson (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:2495-2499), histidinol dehydrogenase (hisD; Hartman and Mulligan (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85:8047-8051), puromycin-N-acetyl transferase (PAC or puro; de la Luna *et al.* (1988) *Gene* 62:121- 126), thymidine kinase (TK; Littlefield (1964) *Science* 145:709-710), and xanthine-guanine phosphoribosyltransferase (XGPRT or gpt; Mulligan and Berg (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78:2072- 2076).

In some embodiments, the expression cassette or vector comprising the polynucleotide encoding the variant LPG10145 RGN polypeptide, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, and/or PE, can further comprise a polynucleotide encoding a crRNA and/or a tracrRNA, or the crRNA and tracrRNA combined to create a sgRNA or combined to be a part of a PEgRNA. The polynucleotide sequence(s) encoding the crRNA, tracrRNA, gRNA, and/or PEgRNA can be operably linked to at least one transcriptional control sequence for expression of the crRNA, tracrRNA, gRNA, and/or PEgRNA in the organism or host cell of interest. For example, the polynucleotide encoding the crRNA, tracrRNA, gRNA, and/or PEgRNA can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters and rice U6 and U3 promoters, such as the human U6 promoter set forth as SEQ ID NO: 41, as well as the promoters disclosed in International Publication No. WO 2022/261394, which is herein incorporated by reference in its entirety, including those set forth herein as SEQ ID NOs: 114-123.

As indicated, expression constructs comprising nucleotide sequences encoding the variant

LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, and/or PEs, crRNAs, tracrRNAs, sgRNAs, and/or PEgRNAs can be used to transform organisms of interest. Methods for transformation involve introducing a nucleotide construct into an organism of interest. By "introducing" is intended to introduce the nucleotide construct to the host cell in such a manner that the construct gains access to the interior of the host cell. The methods of the invention do not require a particular method for introducing a nucleotide construct to a host organism, only that the nucleotide construct gains access to the interior of at least one cell of the host organism. The host cell can be a eukaryotic or prokaryotic cell. In particular embodiments, the eukaryotic host cell is a plant cell, a mammalian cell, an avian cell, or an insect cell. In some embodiments, the eukaryotic cell that comprises or expresses a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, or that has been modified by a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, is a human cell. In some embodiments, the eukaryotic cell that comprises or expresses a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, or that has been modified by a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, is a primary cell. The term "primary cell" refers to a cell isolated directly from a multicellular organism. Primary cells typically have undergone very few population doublings and are therefore more representative of the main functional component of the tissue from which they are derived in comparison to continuous (tumor or artificially immortalized) cell lines. In some cases, primary cells are cells that have been isolated and then used immediately. In other cases, primary cells cannot divide indefinitely and thus cannot be cultured for long periods of time *in vitro*. In some embodiments, a primary cell is a primary T cell. In some embodiments, the eukaryotic cell that comprises or expresses a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, or that has been modified by a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, is a cell of hematopoietic origin, such as an immune cell (i.e., a cell of the innate or adaptive immune system) including but not limited to a B cell, a T cell, a natural killer (NK) cell, a pluripotent stem cell, an induced pluripotent stem cell, a chimeric antigen receptor T (CAR-T) cell, a monocyte, a macrophage, and a dendritic cell. In some embodiments, the eukaryotic cell that comprises or expresses a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, or that has been modified by a presently disclosed variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE, is an ocular cell, muscle cell (e.g., skeletal muscle cell), epithelial cell (e.g., lung epithelial cell), diseased cell (e.g., tumor cell).

Methods for introducing nucleotide constructs into plants and other host cells are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

The methods result in a transformed organism, such as a plant, including whole plants, as well as plant organs (*e.g.*, leaves, stems, roots, etc.), seeds, plant cells, propagules, embryos and progeny of the same. Plant cells can be differentiated or undifferentiated (*e.g.* callus, suspension culture cells, protoplasts, leaf cells, root cells, phloem cells, pollen).

"Transgenic organisms" or "transformed organisms" or "stably transformed" organisms or cells or tissues refers to organisms that have incorporated or integrated a polynucleotide encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, gRNA, and/or PEgRNA of the invention. It is recognized that other exogenous or endogenous nucleic acid sequences or DNA fragments may also be incorporated into the host cell. *Agrobacterium*-and biolistic-mediated transformation remain the two predominantly employed approaches for transformation of plant cells. However, transformation of a host cell may be performed by infection, transfection, microinjection, electroporation, microprojection, biolistics or particle bombardment, electroporation, silica/carbon fibers, ultrasound mediated, PEG mediated, calcium phosphate co-precipitation, polycation DMSO technique, DEAE dextran procedure, and viral mediated, liposome mediated and the like. Viral-mediated introduction of a polynucleotide encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, gRNA, and/or PEgRNA includes retroviral, lentiviral, adenoviral, and adeno-associated viral mediated introduction and expression, as well as the use of Caulimoviruses, Geminiviruses, and RNA plant viruses.

Transformation protocols as well as protocols for introducing polypeptides or polynucleotide sequences into plants may vary depending on the type of host cell (*e.g.*, monocot or dicot plant cell) targeted for transformation. Methods for transformation are known in the art and include those set forth in US Patent Nos: 8,575,425; 7,692,068; 8,802,934; 7,541,517; each of which is herein incorporated by reference. See, also, Rakoczy-Trojanowska, M. (2002) *Cell Mol Biol Lett.* 7:849-858; Jones *et al.* (2005) *Plant Methods* 1:5; Rivera *et al.* (2012) *Physics of Life Reviews* 9:308-345; Bartlett *et al.* (2008) *Plant Methods* 4:1-12; Bates, G.W. (1999) *Methods in Molecular Biology* 111:359-366; Binns and Thomashow (1988) *Annual Reviews in Microbiology* 42:575-606; Christou, P. (1992) *The Plant Journal* 2:275-281; Christou, P. (1995) *Euphytica* 85:13-27; Tzfira *et al.* (2004) *TRENDS in Genetics* 20:375-383; Yao *et al.* (2006) *Journal of Experimental Botany* 57:3737-3746; Zupan and Zambryski (1995) *Plant Physiology* 107:1041-1047; Jones *et al.* (2005) *Plant Methods* 1:5.

Transformation may result in stable or transient incorporation of the nucleic acid into the cell. "Stable transformation" is intended to mean that the nucleotide construct introduced into a host cell integrates into the genome of the host cell and is capable of being inherited by the progeny thereof.

"Transient transformation" is intended to mean that a polynucleotide is introduced into the host cell and does not integrate into the genome of the host cell.

Methods for transformation of chloroplasts are known in the art. See, for example, Svab et al. (1990) *Proc. Nail. Acad. Sci. USA* 87:8526-8530; Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917; Svab and Maliga (1993) *EMBO J.* 12:601-606. The method relies on particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination. Additionally, plastid transformation can be accomplished by transactivation of a silent plastid-borne transgene by tissue-preferred expression of a nuclear-encoded and plastid-directed RNA polymerase. Such a system has been reported in McBride et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:7301-7305.

The cells that have been transformed may be grown into a transgenic organism, such as a plant, in accordance with conventional ways. See, for example, McCormick et al. (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having constitutive expression of the desired phenotypic characteristic identified. Two or more generations may be grown to ensure that the polynucleotide encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, gRNA, and/or PEgRNA is stably maintained and inherited and then seeds harvested to ensure the presence of the polynucleotide encoding a variant LPG10145 RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, PE, crRNA, tracrRNA, gRNA, and/or PEgRNAs. In this manner, the present invention provides a transformed plant or plant part having a nucleotide construct of the invention, for example, an expression cassette of the invention, stably incorporated into their genome. Seed having an expression cassette of the disclosure stably incorporated into their genome can be referred to as "transgenic seed".

Alternatively, cells that have been transformed may be introduced into an organism. These cells could have originated from the organism, wherein the cells are transformed in an *ex vivo* approach. These cells can be autologous (originated and returned to the same subject), allogeneic (the donor and recipient subjects are of the same species).

The sequences provided herein may be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plants of interest include, but are not limited to, corn (maize), sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, and oilseed rape, Brassica sp., alfalfa, rye, millet, safflower, peanuts, sweet potato, cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, papaya, cashew, macadamia, almond, oats, vegetables, ornamentals, and conifers.

Vegetables include, but are not limited to, tomatoes, lettuce, green beans, lima beans, peas, and members of the genus Curcumis such as cucumber, cantaloupe, and musk melon. Ornamentals include, but are not limited to, azalea, hydrangea, hibiscus, roses, tulips, daffodils, petunias, carnation, poinsettia, and

chrysanthemum. Preferably, plants of the present invention are crop plants (for example, maize, sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, oilseed rape, etc.).

As used herein, the term plant includes plant cells, plant protoplasts, plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced polynucleotides. Further provided is a processed plant product or byproduct that retains the sequences disclosed herein, including for example, soymeal.

The polynucleotides encoding the variant LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, crRNAs, tracrRNAs, gRNAs, and/or PEgRNAs or comprising the crRNAs, tracrRNAs, gRNAs, and/or PEgRNAs can also be used to transform any prokaryotic species, including but not limited to, archaea and bacteria (*e.g.*, *Bacillus* sp., *Klebsiella* sp. *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., Agrobacterium, *Lactobacillus* sp.).

The polynucleotides encoding the variant LPG10145 RGNs, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, crRNAs, tracrRNAs, gRNAs, and/or PEgRNAs or comprising the crRNAs, tracrRNAs, gRNAs, and/or PEgRNAs can be used to transform any eukaryotic species, including but not limited to animals (*e.g.*, mammals, insects, fish, birds, and reptiles), fungi, amoeba, algae, and yeast.

Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian, insect, or avian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of an RGN, base editing, or PE system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256: 808- 813 (1992); Nabel & Feigner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10): 1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and Bohm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid: nucleic acid conjugates, naked

DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam ™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration). The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291- 297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Viral. 66:2731-2739 (1992); Johann et al., J. Viral. 66:1635-1640 (1992); Sommnerfelt et al., Viral. 176:58-59 (1990); Wilson et al., J. Viral. 63:2374-2378 (1989); Miller et al., *1.* Viral. 65:2220-2224 (1991); WO 1994/026877).

In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors

may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Katin, Human Gene Therapy 5:793-801 (1994); Muzyczka, *1.* Clin. Invest. 94:1351 (1994).

The term "adeno-associated virus" or "AAV" as used herein refers to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae. Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. At least 11, sequentially numbered, have been described. Non-limiting exemplary serotypes useful in the compositions and methods disclosed herein include any of the 11 serotypes (e.g., AAV2, AAV5, AAV6, AAV8, AAV9), or variant serotypes, e.g., AAV-DJ. AAV is advantageous over other viral vectors for *in vivo* delivery of genes (e.g., encoding gene editing components) due to their low toxicity and low probability of causing insertional mutagenesis because it typically does not integrate into the host genome. AAV has a packaging limit of about 4.5 to 4.75 Kb.

Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., *1.* Viral. 63:03822-3828 (1989). Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψJ2 cells or PA317 cells, which package retrovirus.

Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences.

The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject.

In some embodiments, a cell that is transfected is a eukaryotic cell. In some embodiments, the eukaryotic cell is an animal cell (*e.g.*, mammals, humans, insects, fish, birds, and reptiles). In some embodiments, a cell that is transfected is a human cell. In some embodiments, a cell that is transfected is a cell of hematopoietic origin, such as an immune cell (i.e., a cell of the innate or adaptive immune system) including but not limited to a B cell, a T cell, a natural killer (NK) cell, a pluripotent stem cell, an induced pluripotent stem cell, a chimeric antigen receptor T (CAR-T) cell, a monocyte, a macrophage, and a dendritic cell.

In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. In some embodiments, the cell or cell line is prokaryotic. In some embodiments, the cell or cell line is eukaryotic. In some embodiments, the cell or cell line may be mammalian, such as for example human, monkey, mouse, cow, swine, goat, hamster, rat, cat, or dog. In further embodiments, the cell or cell line is derived from insect, avian, plant, or fungal species. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3, Huhl, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TFl, CTLL-2, CIR, Rat6, CVI, RPTE, AlO, T24, 182, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388Dl, SEM-K2, WEHI- 231, HB56, TIB55, lurkat, *145.01,* LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4. COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-Ll, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-I cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10Tl/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-Kl, CHO-K2, CHO-T, CHO Dhfr-/-, COR-L23, COR-L23/CPR, COR-L235010, CORL23/ R23, COS-7, COV-434, CML Tl, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, lurkat, *lY* cells, K562 cells, Ku812, KCL22, KGl, KYOl, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-l0A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/ 0.2R, MONO-MAC 6, MTD-lA, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-lA/ PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THPl cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, 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, a cell transfected with one or more polynucleotides or vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of an RGN, base editing, or PE system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of an RGN, base editing, or PE system, is used to establish a new cell line

comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, hamster, rabbit, cow, or pig. In some embodiments, the transgenic animal is a bird, such as a chicken or a duck. In some embodiments, the transgenic animal is an insect, such as a mosquito or a tick.

VIII. Variants and Fragments of Polypeptides and Polynucleotides

The present disclosure provides active engineered variants of a naturally-occurring (*i.e.*, wild-type) LPG10145 RNA-guided nuclease. In some embodiments, the wild-type LPG10145 RGN comprises the amino acid sequence set forth as SEQ ID NO: 1. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1 at one or more corresponding amino acid residues.

When referring to a variant LPG10145 RGN that "comprises an amino acid sequence having at least x% (e.g., 85%) sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs", it is intended to mean that the variant LPG10145 RGN comprises an amino acid sequence that has at least x% (e.g., 85%) sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, and the amino acid residues at the recited positions differ from the corresponding amino acid residues in SEQ ID NO: 1. When referring to a variant LPG10145 RGN that "comprises an amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs", it is intended to mean that the variant LPG10145 RGN comprises an amino acid sequence that is identical to the amino acid sequence set forth in SEQ ID NO: 1, except that amino acid residues at the recited positions differ from the corresponding amino acid residues in SEQ ID NO: 1.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472,

533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid

position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some embodiments, a variant LPG10145 RGN of the disclosure comprises an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some embodiments, the present disclosure provides active variants and fragments of naturally-occurring CRISPR repeats, such as the sequence set forth as SEQ ID NO: 33, 244, or 245, and active variants and fragments of naturally-occurring tracrRNAs, such as the sequence set forth as SEQ ID NO: 34, 246, 247, or 248, and polynucleotides encoding the same.

While the activity of a variant or fragment may be altered compared to the polynucleotide or polypeptide of interest, the variant and fragment should retain the functionality of the polynucleotide or polypeptide of interest. For example, a variant or fragment may have increased activity, decreased activity, different spectrum of activity or any other alteration in activity when compared to the polynucleotide or polypeptide of interest.

Variant LPG10145 RGN polypeptides, such as those disclosed herein, will retain sequence-specific, RNA-guided DNA-binding activity. In particular embodiments, fragments and variants of variant LPG10145 RGN polypeptides, such as those disclosed herein, will retain nuclease activity (single-stranded or double-stranded).

The binding and/or cleaving activity of an active variant or fragment of a variant LPG10145 RGN disclosed herein can be dependent upon recognizing a protospacer adjacent motif (PAM) adjacent and 3' to the target sequence. In some embodiments, the PAM comprises a consensus sequence of NNGG.

In some embodiments, the present disclosure provides an active variant or fragment of a variant LPG10145 RGN that has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285 has increased nuclease activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein has nuclease activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the nuclease activity of a reference LPG10145 RGN. In some embodiments, a reference LPG10145 RGN has the amino acid sequence set forth as SEQ ID NO: 1. In

some embodiments, a reference LPG10145 RGN is a variant of SEQ ID NO: 1 that lacks the corresponding mutations of the active variant or fragment thereof that has nuclease activity, e.g., as described above. In some embodiments, a reference LPG10145 RGN is a non-identical variant LPG10145 RGN. In some embodiments, a reference LPG10145 RGN is a non-identical variant LPG10145 RGN. Nuclease activity can be measured by assays known to one of skill in the art, including but not limited to, Tracking of Indels by DEcomposition (TIDE) analysis, flow cytometry, *in vitro* or *in vivo* cleavage assays wherein cleavage is confirmed using PCR, sequencing, or gel electrophoresis, with or without the attachment of an appropriate label (*e.g.*, radioisotope, fluorescent substance) to the target sequence to facilitate detection of degradation products. In some embodiments, the nicking triggered exponential amplification reaction (NTEXPAR) assay can be used (see, *e.g.*, Zhang et al. (2016) *Chem. Sci.* 7:4951-4957). *In vivo* cleavage can be evaluated using the Surveyor assay (Guschin et al. (2010) *Methods Mol Biol* 649:247-256). In some embodiments, the efficiency of cleaving a target sequence is assessed by measuring the percentage of a target sequence or cells comprising the target sequence that comprise altered expression of the target sequence or of a polypeptide encoded by the target sequence. In some embodiments, the expression is measured by quantitative PCR, microarray, RNA-seq, flow cytometry, immunoblot, enzyme-linked immunosorbent assay (ELISA), protein immunoprecipitation, immunostaining, high performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC/MS), mass spectrometry, or a combination thereof. In some embodiments, the target sequence encodes a cell surface expressed protein, and the efficiency of cleaving the target sequence is assessed by measuring the percentage of cells comprising a reduction of the cell surface expressed protein as measured by flow cytometry.

In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein, when operably fused to a deaminase, has increased base editing activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285, when operably fused to a deaminase, has increased base editing activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285, when operably fused to a deaminase, has increased base editing activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein, when operably fused to a deaminase, has base editing activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least 103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least

111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the base editing activity of a reference LPG10145 RGN, when operably fused to the same deaminase. In some embodiments, a reference LPG10145 RGN has the amino acid sequence set forth as SEQ ID NO: 1. In some embodiments, a reference LPG10145 RGN is a variant of SEQ ID NO: 1 that lacks the corresponding mutations of the active variant or fragment thereof that has base editing activity, e.g., as described above. In some embodiments, a reference LPG10145 RGN is a non-identical variant LPG10145 RGN. Base editing activity can be measured by assays known to one of skill in the art, including but not limited to, transfection of mammalian cells with a base editor (e.g., comprising a variant LPG10145 RGN and a deaminase) and a guide RNA and detecting base editing by PCR amplification of the target sequence and next generation sequencing (NGS), as described in Example 4 of the present disclosure. Base editing assays are also described in International Publication No. WO 2022/056254, which is herein incorporated by reference in its entirety. In some embodiments, the efficiency of base editing a target sequence is assessed by measuring the percentage of a target sequence or cells comprising the target sequence that comprise altered expression of the target sequence or of a polypeptide encoded by the target sequence. In some embodiments, the expression is measured by quantitative PCR, microarray, RNA-seq, flow cytometry, immunoblot, enzyme-linked immunosorbent assay (ELISA), protein immunoprecipitation, immunostaining, high performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC/MS), mass spectrometry, or a combination thereof.

In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein, when operably fused to a polymerase (e.g., RT), has increased editing activity as compared to SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 2-16, 182-196, and 271-285, when operably fused to a polymerase (e.g., RT), has increased editing activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein comprising the amino acid sequence of any one of SEQ ID NOs: 2-16, 182-196, and 271-285, when operably fused to a polymerase (e.g., RT), has increased editing activity as compared to the RGN of SEQ ID NO: 1. In some embodiments, an active variant or fragment of a variant LPG10145 RGN disclosed herein, when operably fused to a polymerase (e.g., RT), has editing activity that is from about 80% to about 500%, from about 80% to about 200%, or from about 90% to about 150%, or from about 95% to about 120%, or is at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, at least 100%, at least 101%, at least 102%, at least

103%, at least 104%, at least 105%, at least 106%, at least 107%, at least 108%, at least 109%, at least 110%, at least 111%, at least 112%, at least 113%, at least 114%, at least 115%, at least 116%, at least 117%, at least 118%, at least 119%, at least 120%, at least 125%, at least 130%, at least 135%, at least 140%, at least 145%, at least 150%, at least 160%, at least 170%, at least 180%, at least 190%, at least 200%, at least 250%, at least 300%, at least 350%, at least 400%, at least 450%, at least 500%, or more, of the editing activity of a reference LPG10145 RGN, when operably fused to the same polymerase. In some embodiments, a reference LPG10145 RGN has the amino acid sequence set forth as SEQ ID NO: 1. In some embodiments, a reference LPG10145 RGN is a variant of SEQ ID NO: 1 that lacks the corresponding mutations of the active variant or fragment thereof that has editing activity, e.g., as described above. In some embodiments, a reference LPG10145 RGN is a non-identical variant LPG10145 RGN. Editing activity can be measured by assays known to one of skill in the art, including but not limited to, transfection of mammalian cells with a polymerase editor (e.g., comprising a variant LPG10145 RGN and a reverse transcriptase) and a PEgRNA and detecting editing by PCR amplification of the target sequence and next generation sequencing (NGS), as described in Examples 5 and 6 of the present disclosure. The editing rate (RT Edit %) can be determined by calculating the percent of read counts assigned to alleles containing the desired edit and, optionally, no additional edits. In some embodiments, the efficiency of editing a target sequence is assessed by measuring the percentage of a target sequence or cells comprising the target sequence that comprise altered expression of the target sequence or of a polypeptide encoded by the target sequence. In some embodiments, the expression is measured by quantitative PCR, microarray, RNA-seq, flow cytometry, immunoblot, enzyme-linked immunosorbent assay (ELISA), protein immunoprecipitation, immunostaining, high performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC/MS), mass spectrometry, or a combination thereof.

Active variants and fragments of CRISPR repeats, such as those disclosed herein, will retain the ability, when part of a guide RNA (comprising a tracrRNA), to bind to and guide an RNA-guided nuclease or base editor or PE comprising the same (complexed with the guide RNA) to a target nucleotide sequence in a sequence-specific manner.

Active variants and fragments of tracrRNAs, such as those disclosed herein, will retain the ability, when part of a guide RNA (comprising a CRISPR RNA), to guide an RNA-guided nuclease or base editor or PE comprising the same (complexed with the guide RNA) to a target nucleotide sequence in a sequence-specific manner.

Active variants and fragments of PEgRNAs disclosed herein, will retain the ability to bind and guide a PE to a target nucleotide sequence in a sequence-specific manner.

Active variants and fragments of base editors disclosed herein, will retain the ability to, when associated with a guide RNA, chemically modify (e.g., deaminate) a nucleobase, resulting in conversion from one nucleobase to another.

Active variants and fragments of PEs disclosed herein, will retain the ability to, when associated with a PEgRNA, edit a double-stranded polynucleotide through the replacement of a target sequence using the template sequence of the PEgRNA.

The term “fragment” refers to a portion of a polynucleotide or polypeptide sequence of the invention. "Fragments", “active fragments”, or "biologically active portions" include polynucleotides comprising a sufficient number of contiguous nucleotides to retain the biological activity (*i.e.*, binding to and directing an RGN in a sequence-specific manner to a target nucleotide sequence when comprised within a guide RNA). "Fragments" or "biologically active portions" include polypeptides comprising a sufficient number of contiguous amino acid residues to retain the biological activity (*i.e.*, binding to a target nucleotide sequence in a sequence-specific manner when complexed with a guide RNA). Fragments of the RGN proteins include those that are shorter than the full-length sequences due to the use of an alternate downstream start site. Such biologically active portions can be prepared by recombinant techniques and evaluated for activity.

A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. A biologically active portion of an

RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 are positively charged amino acid residues.

A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. A biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, a biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the

amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16.

In some embodiments, a biologically active portion of an RGN protein can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or more contiguous amino acid residues of an amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence

set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16. Such biologically active portions can be prepared by recombinant techniques and evaluated for sequence-specific, RNA-guided DNA-binding activity.

A biologically active fragment of a CRISPR repeat can comprise at least 8 contiguous amino acids of SEQ ID NO: 33, 244, or 245. A biologically active portion of a CRISPR repeat can be a polynucleotide that comprises, for example, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 contiguous nucleotides of SEQ ID NO: 33, 244, or 245. A biologically active portion of a tracrRNA can be a polynucleotide that comprises, for example, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more contiguous nucleotides of SEQ ID NO: 34, 246, 247, or 248. A biologically active fragment of a guide RNA backbone (comprising the CRISPR repeat and tracrRNA, and optionally a nucleotide linker) can be a polynucleotide that comprises, for example, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more contiguous nucleotides of any one of SEQ ID NOs: 249-252. A biologically active fragment of a PEgRNA can be a polynucleotide that comprises, for example, 100, 105, 110, 115, 120, 125, 130, or more contiguous nucleotides of any one of SEQ ID NOs: 129-132, and 197-220.

A biologically active portion of a polymerase (e.g., RT) can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, or more contiguous amino acid residues of SEQ ID NO: 254 or 255.

A biologically active portion of a base editing polypeptide (e.g., deaminase) can be a polypeptide that comprises, for example, 10, 25, 50, 75, 100, 125, or more contiguous amino acid residues of any one of SEQ ID NOs: 42-113, and 257.

A biologically active portion of a base editor can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, or more contiguous amino acid residues of any one of SEQ ID NOs: 258-270.

A biologically active portion of a PE can be a polypeptide that comprises, for example, 10, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, or more contiguous amino acid residues of any one of SEQ ID NOs: 135, and 162-181.

In general, "variants" or “active variant” is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a deletion and/or addition of one or more nucleotides at one or more internal sites within the native polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" or “wild type” polynucleotide or

polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. For polynucleotides, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the native amino acid sequence of the gene of interest. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis but which still encode the polypeptide or the polynucleotide of interest. Generally, variants of a particular polynucleotide disclosed herein will have at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to that particular polynucleotide as determined by sequence alignment programs and parameters described elsewhere herein.

Variants of a particular polynucleotide disclosed herein (*i.e.*, the reference polynucleotide) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant polynucleotide and the polypeptide encoded by the reference polynucleotide. Percent sequence identity between any two polypeptides can be calculated using sequence alignment programs and parameters described elsewhere herein. Where any given pair of polynucleotides disclosed herein is evaluated by comparison of the percent sequence identity shared by the two polypeptides they encode, the percent sequence identity between the two encoded polypeptides is at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the

amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differs from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differs from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

97%, 98%, 99%, or greater identity to the amino acid sequence set forth as SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position

647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R; (ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R; (hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

In some embodiments, the presently disclosed polynucleotides encode an RNA-guided nuclease polypeptide comprising an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater identity to the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position

778 is substituted by an R; (c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and (o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

A biologically active variant of a polymerase (e.g., RT) can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set forth as SEQ ID NO: 254 or 255.

A biologically active variant of a base editing polypeptide (e.g., deaminase) can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set forth as any one of SEQ ID NOs: 42-113, and 257.

A biologically active variant of a base editor can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set forth as any one of SEQ ID NOs: 258-270.

A biologically active variant of a PE can comprise an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the amino acid sequence set forth as any one of SEQ ID NOs: 135, and 162-181.

A biologically active variant of an RGN polypeptide, a fusion protein comprising an RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a base editor, or a PE of the invention may differ by as few as about 1-15 amino acid residues, as few as about 1-10, such as about 6-10, as few as 5, as few as 4, as few as 3, as few as 2, or as few as 1 amino acid residue. In specific embodiments, the polypeptides can comprise an N-terminal or a C-terminal truncation, which can comprise at least a deletion of 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100 amino acids or more from either the N or C terminus of the polypeptide.

In certain embodiments, the presently disclosed polynucleotides comprise or encode a CRISPR repeat comprising a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245. The presently disclosed polynucleotides can comprise or encode a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides. In some embodiments, the CRISPR repeat comprises a nucleotide sequence that differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide.

The presently disclosed polynucleotides can comprise or encode a tracrRNA comprising a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248.

The presently disclosed polynucleotides can comprise or encode a guide RNA backbone comprising a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as any one of SEQ ID NOs: 249-252.

The presently disclosed polynucleotides can comprise or encode a PEgRNA comprising a nucleotide sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%,

86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or greater sequence identity to the nucleotide sequence set forth as any one of SEQ ID NOs: 129-132, and 197-220.

Biologically active variants of a CRISPR repeat, tracrRNA, guide RNA backbone, or PEgRNA of the invention may differ by as few as about 1-15 nucleotides, as few as about 1-10, such as about 6-10, as few as 5, as few as 4, as few as 3, as few as 2, or as few as 1 nucleotide. In specific embodiments, the polynucleotides can comprise a 5' or 3' truncation, which can comprise at least a deletion of 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 95, 100, 105, 110 nucleotides or more from either the 5' or 3' end of the polynucleotide.

Variants of guide RNAs disclosed herein include guide RNAs that have modified nucleotides, sugars, phosphate backbone, and/or nucleobases. Variant guide RNAs can include modifications including: 2'-O-methyl (2'-O-Me) modification; 2'-fluoro (2'-F) modification; 2'F-4'Cα-OMe modification; 2',4'-di-Cα-OMe modification; 2'-O-methyl 3'-phosphorothioate (MS) modification; 2'-O-methyl 3'thiophosphonoacetate (MSP; 2'-O-methyl 3'thioPACE) modification; 2'-O-methyl 3'phosphonoacetate (MP) modification; phosphorothioate (PS) modification; bridged nucleic acid (BNA) modification (e.g., 2',4' BNA, locked nucleic acid (LNA), N-methyl substituted bridged nucleic acid BNA^{NC}[N-Me], 2'-O,4'-C-ethylene bridged nucleic acid (2',4'-ENA), and S-constrained ethyl (cEt)); or a combination thereof. Chemical modifications of spacers, crRNA repeats, crRNAs, tracrRNAs, and guide RNAs are described in International Application Publication No. WO 2024/042489, which is hereby incorporated by reference in its entirety herein.

It is recognized that modifications may be made to the RGN polypeptides, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, CRISPR repeats, tracrRNAs, guide RNAs, and PEgRNAs provided herein creating variant proteins and polynucleotides. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques. Alternatively, native, as yet-unknown or as yet unidentified polynucleotides and/or polypeptides structurally and/or functionally-related to the sequences disclosed herein may also be identified that fall within the scope of the present invention. Conservative amino acid substitutions may be made in nonconserved regions that do not alter the function of the RGN proteins, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, or PEs. Alternatively, modifications may be made that improve the activity of the RGN proteins, fusion proteins comprising the same, polymerases (e.g., RTs), base editing polypeptides, base editors, or PEs.

Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more variant LPG10145 RGN protein, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE disclosed herein is manipulated to create a new RGN protein, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population

of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined *in vitro* or *in vivo*. For example, using this approach, sequence motifs encoding a domain of interest may be shuffled between the RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE sequences provided herein and other known RGN, fusion protein comprising the same, polymerase (e.g., RT), base editing polypeptide, base editor, or PE genes to obtain a new gene coding for a protein with an improved property of interest, such as an increased K_m in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri *et al.* (1997) *Nature Biotech.* 15:436-438; Moore *et al.* (1997) *J. Mol. Biol.* 272:336-347; Zhang *et al. (1997) Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri *et al.* (1998) *Nature* 391:288-291; and U.S. Patent Nos. 5,605,793 and 5,837,458. A "shuffled" nucleic acid is a nucleic acid produced by a shuffling procedure such as any shuffling procedure set forth herein. Shuffled nucleic acids are produced by recombining (physically or virtually) two or more nucleic acids (or character strings), for example in an artificial, and optionally recursive, fashion. Generally, one or more screening steps are used in shuffling processes to identify nucleic acids of interest; this screening step can be performed before or after any recombination step. In some (but not all) shuffling embodiments, it is desirable to perform multiple rounds of recombination prior to selection to increase the diversity of the pool to be screened. The overall process of recombination and selection are optionally repeated recursively. Depending on context, shuffling can refer to an overall process of recombination and selection, or, alternately, can simply refer to the recombinational portions of the overall process.

As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. It is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (*e.g.*, charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically, this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference

sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix; or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (*e.g.*, BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, *e.g.*, in Dayhoff et al. (1978) "A model of evolutionary change in proteins." In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff et al. (1992) Proc. Natl. Acad. Sci. USA 89:10915-10919. The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acid positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences, so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, *e.g.*, gapped BLAST 2.0, described in Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402, and made available to the public at the National Center for Biotechnology Information Website (www.ncbi.nlm.nih.gov). Optimal alignments, including multiple alignments, can be prepared using, *e.g.*, PSI-BLAST, available through www.ncbi.nlm.nih.gov and described by Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402.

With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue "corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general

the amino acid residue number in a test sequence as determined by simply counting from the N-terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

IX. Antibodies

Also encompassed in the disclosure are antibodies recognizing: engineered variant LPG10145 RGN polypeptides, fusion proteins comprising the RGN polypeptides, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, or ribonucleoproteins comprising the RGN polypeptides or fusion proteins of the present disclosure. Engineered variant LPG10145 RGN polypeptides include those having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

Antibodies of the disclosure bind to the RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides of the present disclosure, including those having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence selected from: the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ

ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or active variants or fragments thereof.

In some embodiments, antibodies of the present disclosure bind to RGN polypeptides or ribonucleoproteins comprising the RGN polypeptides having the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or active variants or fragments thereof.

Methods for producing antibodies are well known in the art (see, for example, Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; and U.S. Pat. No. 4,196,265). These antibodies can be used in kits for the detection and isolation of RGN polypeptides, fusion proteins comprising the RGN polypeptides, polymerases (e.g., RTs), base editing polypeptides, base editors, PEs, or ribonucleoproteins comprising the RGN polypeptides or fusion proteins of the disclosure. Thus, this disclosure provides kits comprising antibodies that specifically bind to the polypeptides or ribonucleoproteins described herein, including, for example, polypeptides comprising the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the polypeptides comprise the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the polypeptides comprise the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

The polypeptides can comprise the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, the polypeptides comprise the

amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, the polypeptides comprise the amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

The kits can comprise antibodies that specifically bind to the polypeptides comprising the amino acid sequence selected from: the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid

sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16.

X. Systems and Ribonucleoprotein Complexes for Binding, Cleaving, or Modifying a Target Sequence of Interest and Methods of Making the Same

The present disclosure provides a system for binding a target sequence of interest, wherein the system comprises at least one guide RNA or a nucleotide sequence encoding the same, and at least one variant LPG10145 RNA-guided nuclease or a nucleotide sequence encoding the same. The guide RNA hybridizes to the target sequence of interest and also forms a complex with the variant LPG10145 RGN polypeptide, thereby directing the variant LPG10145 RGN polypeptide to bind to the target sequence.

The present disclosure also provides a base editing system for binding and editing a target sequence of interest, wherein the base editing system comprises an engineered variant LPG10145 RGN polypeptide (or a polynucleotide encoding the same), a base editing polypeptide (e.g., deaminase) (or a polynucleotide encoding the same), along with one or more guide RNA (or one or more polynucleotide sequences encoding the same). The base editing system may comprise a base editor wherein the base editor comprises a single fusion protein (or a polynucleotide encoding the same) comprising the variant LPG10145 RGN polypeptide and the base editing polypeptide (e.g., deaminase), or a base editor wherein the variant LPG10145 RGN polypeptide and the base editing polypeptide (e.g., deaminase) are two separate polypeptides. In some embodiments, a base editing system can comprise a plurality of engineered variant LPG10145 RGN polypeptides (or polynucleotides encoding the same), a plurality of base editing polypeptide (e.g., deaminases) (or polynucleotides encoding the same), and/or a plurality of guide RNAs (or polynucleotides encoding the same). A base editor of the disclosure can comprise an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to any one of SEQ ID NOs: 258-270. In some embodiments, a base editor disclosed herein comprises the amino acid sequence set forth as any one of SEQ ID NOs: 258-270.

The present disclosure also provides a PE system for binding and editing a target sequence of interest, wherein the PE system comprises an engineered variant LPG10145 RGN polypeptide (or a polynucleotide encoding the same), a polymerase (e.g., RT) (or a polynucleotide encoding the same), along with one or more PEgRNAs (or one or more polynucleotide sequences encoding the same). The PE system may comprise a PE wherein the PE comprises a single fusion protein (or a polynucleotide encoding the

same) comprising the variant LPG10145 RGN polypeptide and the polymerase (e.g., RT), or a PE wherein the variant LPG10145 RGN polypeptide and the polymerase (e.g., RT) are two separate polypeptides. A PE of the disclosure can comprise an amino acid sequence having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to any one of SEQ ID NOs: 135, and 162-181. In some embodiments, a PE disclosed herein comprises the amino acid sequence set forth as any one of SEQ ID NOs: 135, and 162-181.

In some embodiments, a dual PE (DPE) system for dual polymerase editing is provided, wherein the system comprises: a) a first PE system comprising a first PE (or one or more polynucleotides encoding the same; or a combination of protein(s) and encoding polynucleotide(s)) and a first PEgRNA (or a polynucleotide encoding the same); and b) a second PE system comprising a second PE (or one or more polynucleotides encoding the same; or a combination of protein(s) and encoding polynucleotide(s)) and a second PEgRNA (or a polynucleotide encoding the same). The first and second PE can be the same or different and can be any one of the PEs described herein. The DPE system can also comprise a PE (or one or more polynucleotides encoding the same; or a combination of protein(s) and encoding polynucleotide(s)) and a first PEgRNA (or a polynucleotide encoding the same) and a second PEgRNA (or a polynucleotide encoding the same). The first and second PEgRNAs can bind to opposite strands of a target polynucleotide such that the intended repair event is installed on both strands of the target DNA. The DNA synthesis template of the first PEgRNA and the DNA synthesis template of the second PEgRNA each encode a single-stranded DNA sequence that is complementary (full or partial) to each other so that the region of the DNA between the two nicked sites is replaced.

For large insertions, the DPE system comprises PEgRNAs wherein the DNA synthesis templates are designed such that the replacement sequence (comprising the complementary single-stranded DNA sequences encoded by the first and second DNA synthesis templates) comprises a first recombinase site, as well as a donor DNA comprising a second recombinase site and the corresponding site-specific recombinase that recognizes the first and second recombinase site. The recombination of the replacement sequence and donor DNA results in an insertion of exogenous DNA.

As used herein, the term "recombinase" refers to a site-specific enzyme that catalyzes the recombination of DNA between recombinase sites that results in the excision, integration, inversion, or exchange of DNA fragments between the recombinase sites. Non-limiting examples of recombinases are serine recombinases (e.g., Hin, Gin, Tn3, β-six, CinH, ParA, γδ, Bxb1, ϕC31, TP901, TG1, φBT1, R4, φRV1, φFC1, MR11, A118, U153, and gp29) and tyrosine recombinases (e.g., Cre, FLP, R, Lambda, HK101, HK022, and pSAM2). See, e.g., Brown et al., "Serine recombinases as tools for genome engineering." Methods.2011;53(4):372-9; Hirano et al., "Site-specific recombinases as tools for heterologous gene integration." Appl. Microbiol. Biotechnol.2011; 92(2):227-39; Chavez and Calos, "Therapeutic applications of the ΦC31 integrase system." Curr. Gene Ther.2011;11(5):375-81; Turan and

Bode, "Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications." FASEB J.2011; 25(12):4088-107; Venken and Bellen, "Genome-wide manipulations of Drosophila melanogaster with transposons, Flp recombinase, and ΦC31 integrase." Methods Mol. Biol.2012; 859:203-28; Murphy, "Phage recombinases and their applications." Adv. Virus Res.2012; 83:367-414; Zhang et al., "Conditional gene manipulation: Cre-ating a new biological era." J. Zhejiang Univ. Sci. B.2012; 13(7):511-24; Karpenshif and Bernstein, "From yeast to mammals: recent advances in genetic control of homologous recombination." DNA Repair (Amst).2012; 1;11(10):781-8; each of which are hereby incorporated by reference in its entirety. Serine and tyrosine recombinases derive their name from the conserved nucleophilic amino acid residue that it uses to attack the DNA and is covalently linked to the DNA during strand exchange.

As used herein, the term "recombinase site" refers to a target nucleotide sequence recognized by a recombinase that undergoes strand exchange with another nucleotide sequence having a similar recombinase site. Non-limiting examples of recombinase sites are the attB/attP sites recognized by the HK022 and ϕC31 recombinases.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some embodiments, the variant LPG10145 RGNs comprise an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some of these embodiments, the variant LPG10145 RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid

sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some of these embodiments, the variant LPG10145 RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In various embodiments, the guide RNA comprises a CRISPR repeat sequence comprising the nucleotide sequence set forth as SEQ ID NO: 33, 245, or 245, or an active variant or fragment thereof. In particular embodiments, the guide RNA comprises a tracrRNA comprising the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof. The guide RNA of the system can be a single guide RNA or a dual-guide RNA. In particular embodiments, the system comprises an RNA-guided nuclease that is heterologous to the guide RNA, wherein the RGN and guide RNA are not found complexed to one another (i.e., bound to one another) in nature.

The system for binding a target sequence of interest provided herein can be a ribonucleoprotein complex, which is at least one molecule of an RNA bound to at least one protein. The ribonucleoprotein (RNP) complexes provided herein can comprise at least one guide RNA as the RNA component and an RNA-guided nuclease as the protein component. An RNP complex provided herein can comprise at least one guide RNA as the RNA component and a base editor as the protein component. An RNP complex provided herein can comprise at least one PEgRNA as the RNA component and a PE as the protein component. Such

ribonucleoprotein complexes can be purified from a cell or organism that naturally expresses the protein component (e.g., an RGN polypeptide) and has been engineered to express a particular guide RNA that is specific for a target sequence of interest. Alternatively, the ribonucleoprotein complex can be purified from a cell or organism that has been transformed with polynucleotides that encode a protein component (e.g., an RGN polypeptide, a base editor, or a PE) and an RNA component (e.g., a guide RNA or a PEgRNA) (or a polynucleotide that comprises the RNA component) and cultured under conditions to allow for the expression of the protein and RNA components (e.g., RGN polypeptide and guide RNA, base editor and guide RNA, or PE and PEgRNA). Thus, methods are provided for making an RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, or a PE, or an RNP complex comprising these polypeptides. Such methods comprise culturing a cell comprising a nucleotide sequence encoding a protein component (e.g., an RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, or PE), and in some embodiments a nucleotide sequence encoding an RNA component (e.g., a guide RNA or PEgRNA), under conditions in which the protein component (e.g., RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, or PE) (and in some embodiments, the RNA component (e.g., guide RNA or PEgRNA) is expressed. The RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, a PE, or RNP complex can then be purified from a lysate of the cultured cells. In embodiments, the nucleotide sequence encoding an RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, or PE includes a mRNA (messenger RNA). In some embodiments, methods for assembling an RNP complex comprise combining one or more of the presently disclosed guide RNAs and one or more of the presently disclosed RGN polypeptides, base editors, or PEs under conditions suitable for formation of the RNP complex. In some embodiments, methods for assembling an RNP complex comprise combining one or more of the presently disclosed RNA components (e.g., guide RNAs or PEgRNAs) and one or more of the presently disclosed protein components (e.g., RGN polypeptide, base editor, PE) under conditions suitable for formation of the RNP complex.

Methods for purifying an RGN polypeptide, a polymerase (e.g., RT), a base editing polypeptide, a fusion protein comprising an RGN, a base editor, a PE, or RNP complex from a lysate of a biological sample are known in the art (*e.g.*, size exclusion and/or affinity chromatography, 2D-PAGE, HPLC, reversed-phase chromatography, immunoprecipitation). In particular methods, the RGN polypeptide, polymerase (e.g., RT), base editing polypeptide, fusion protein comprising an RGN, base editor, PE, or RNP complex is recombinantly produced and comprises a purification tag to aid in its purification, including but not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG (e.g., 3X FLAG tag), HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, 10xHis, biotin carboxyl carrier protein (BCCP), and calmodulin. Generally, the tagged RGN

polypeptide, polymerase (e.g., RT), base editing polypeptide, fusion protein comprising an RGN, base editor, PE, or RGN ribonucleoprotein complex is purified using immobilized metal affinity chromatography. It will be appreciated that other similar methods known in the art may be used, including other forms of chromatography or for example immunoprecipitation, either alone or in combination.

An "isolated" or "purified" polypeptide, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the polypeptide as found in its naturally occurring environment. Thus, an isolated or purified polypeptide is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, optimally culture medium represents less than about 30%, 20%, 10%, 5%, or 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals. Similarly, an “isolated” polynucleotide or nucleic acid molecule is removed from its naturally occurring environment. An isolated polynucleotide is substantially free of chemical precursors or other chemicals when chemically synthesized or has been removed from a genomic locus via the breaking of phosphodiester bonds. An isolated polynucleotide can be part of a vector, a composition of matter or can be contained within a cell so long as the cell is not the original environment of the polynucleotide.

Particular methods provided herein for binding, cleaving, and/or editing a target sequence of interest involve the use of an *in vitro* assembled RNP complex. *In vitro* assembly of an RNPcomplex can be performed using any method known in the art in which a protein component (e.g., an RGN polypeptide, base editor, or PE) is contacted with an RNA component (e.g., guide RNA or PEgRNA) under conditions to allow for binding of the protein component (e.g., RGN polypeptide, base editor, or PE) to the RNA component (e.g., guide RNA or PEgRNA). As used herein, "contact", contacting", or "contacted" refer to making one entity in touch with one or more other entities with or without any intermediate means. In some embodiments, “contact”, “contacting”, or “contacted” include but is not limited to placing the components of a desired reaction together under conditions suitable for carrying out the desired reaction, or placing the entities in the same container or the same solution, or transfection, transformation, viral delivery, or LNP delivery of the one entity and the one or more other entities. The RGN polypeptide, base editor, or PE can be purified from a biological sample, cell lysate, or culture medium, produced via *in vitro* translation, or chemically synthesized. The guide RNA or PEgRNA can be purified from a biological sample, cell lysate, or culture medium, transcribed *in vitro*, or chemically synthesized. The protein component (e.g., RGN polypeptide, base editor, or PE) and RNA component (e.g., guide RNA or PEgRNA) can be brought into contact in solution (*e.g.*, buffered saline solution) to allow for *in vitro* assembly of the RNP complex.

The disclosure provides kits containing any one or more of the elements of disclosed compositions described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit.

In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a DNA sequence encoding the crRNA sequence and one or more insertion sites for inserting a guide sequence upstream of the encoded crRNA sequence, wherein when expressed, the guide sequence directs sequence-specific binding of an RGN complex to a target sequence in a eukaryotic cell, wherein the RGN complex comprises an RGN enzyme complexed with the guide RNA polynucleotide; and/or (b) a second regulatory element operably linked to an enzyme coding sequence encoding said RGN enzyme comprising a nuclear localization sequence. The kits can include engineered variant LPG10145 RGN polypeptides, polymerases (e.g., RTs), base editing polypeptides, fusion proteins comprising variant LPG10145 RGN polypeptides, base editors, PEs, guide RNAs, PEgRNAs, or polynucleotides encoding the same; cells; and complete RGN, base editing, or PE systems. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube.

In some embodiments, the kit includes instructions in one or more languages. In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g. in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10.

In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template polynucleotide. In some embodiments, the disclosure provides methods for using one or more elements of an RGN, base editing, or PE system. The RGN, base editing, or PE system of the invention provides an effective means for modifying a target polynucleotide. The RGN, base editing, or PE system of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating, base editing) a target polynucleotide in a multiplicity of cell types. As such the RGN, base editing, or PE system of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary RGN system, or RGN complex, comprises an RGN enzyme complexed with a guide RNA hybridized to a target sequence within the target polynucleotide. An exemplary base editing system, or RNP complex comprising a base editor, comprises a base editor complexed with a guide RNA hybridized to a target sequence within the target polynucleotide. An exemplary PE system, or RNP complex comprising a PE, comprises a PE complexed with a PEgRNA hybridized to a target sequence within the target polynucleotide.

XI. Methods of Binding, Cleaving, or Modifying a Target Sequence

The present disclosure provides methods for binding, cleaving, and/or modifying a target nucleotide sequence of interest. The methods include delivering a system comprising at least one guide RNA or a polynucleotide encoding the same, and at least one variant LPG10145 RGN polypeptide or a polynucleotide encoding the same to the target sequence or a cell, organelle, or embryo comprising the target sequence. In some embodiments, the methods include delivering a system comprising at least one guide RNA or a polynucleotide encoding the same, and at least one base editor comprising a variant LPG10145 RGN polypeptide (or one or more polynucleotides encoding the same, or a combination of protein(s) and encoding polynucleotide(s)) to the target sequence or a cell, organelle, or embryo comprising the target sequence. In some embodiments, the methods include delivering a system comprising at least one PEgRNA or a polynucleotide encoding the same, and at least one PE comprising a variant LPG10145 RGN polypeptide (or one or more polynucleotides encoding the same, or a combination of protein(s) and encoding polynucleotide(s)) to the target sequence or a cell, organelle, or embryo comprising the target sequence. The methods for modifying a target polynucleotide (e.g., target DNA) of interest comprising a target sequence can be performed *ex vivo* or *in vitro*. In some embodiments, the methods for modifying a target polynucleotide (e.g., target DNA) of interest are not methods for treatment of the human or animal body by therapy or are not processes for editing the germ line genetic identity of a human being.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some of these embodiments, the RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid

sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj) the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some of these embodiments, the RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some embodiments, the guide RNA comprises a CRISPR repeat comprising the nucleotide sequences set forth as SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof. In some embodiments, the guide RNA comprises a tracrRNA comprising the nucleotide sequence set forth as SEQ ID NO: 34, 246, 247, or 248, or an active variant or fragment thereof. The guide RNA of the system can be a single guide RNA or a dual-guide RNA.

The RGN of the system may be a nuclease dead RGN, have nickase activity, or may be a fusion polypeptide. In some embodiments, the fusion polypeptide comprises a base-editing polypeptide, for example a cytosine deaminase or an adenine deaminase. In other embodiments, the RGN fusion protein is a PE. The PE can comprise, for example, a polymerase (e.g., reverse transcriptase). In other embodiments, the RGN fusion protein comprises a polypeptide that recruits members of a functional nucleic acid repair complex, such as a member of the nucleotide excision repair (NER) or transcription coupled-nucleotide excision repair (TC-NER) pathway (Wei *et al.*, 2015, *PNAS USA* 112(27):E3495-504 ; Troelstra *et al.*, 1992, *Cell* 71:939-953; Marnef *et al.*, 2017, *J Mol Biol* 429(9):1277-1288), as described in U.S. Provisional

Application No. 62/966,203, which was filed on January 27, 2020, and is incorporated by reference in its entirety. In some embodiments, the RGN fusion protein comprises CSB (van den Boom *et al*., 2004, *J Cell Biol* 166(1):27-36; van Gool *et al*., 1997, *EMBO J* 16(19):5955-65; an example of which is set forth as SEQ ID NO: 39), which is a member of the TC-NER (nucleotide excision repair) pathway and functions in the recruitment of other members. In further embodiments, the RGN fusion protein comprises an active domain of CSB, such as the acidic domain of CSB which comprises amino acid residues 356-394 of SEQ ID NO: 39 (Teng *et al*., 2018, *Nat Commun* 9(1):4115).

In particular embodiments, the RGN, base editor, PE, and/or guide RNA is heterologous to the cell, organelle, or embryo to which the RGN, base editor, PE, and/or guide RNA (or polynucleotide(s) encoding at least one of the RGN and guide RNA) are introduced.

Delivery of a polynucleotide encoding a guide RNA, a PEgRNA, an RGN polypeptide, a base editor, and/or a PE, to a cell or embryo can be delivered to a cell or embryo *ex vivo*, *in vitro*, or *in vivo*. The cell or embryo can then be cultured under conditions in which the guide RNA, PEgRNA, RGN polypeptide, base editor, and/or PEare expressed. In various embodiments, the method comprises contacting a target polynucleotide with an RNP complex. The contacting can be *ex vivo*, *in vitro*, or *in vivo*. The RNP complex may comprise an RGN that is nuclease dead or has nickase activity. In some embodiments, the RGN of the RNP complex is a fusion polypeptide comprising a base-editing polypeptide. In some embodiments, the RGN of the RNP complex is a fusion polypeptide comprising a polymerase (e.g., RT). In certain embodiments, the method comprises introducing into a cell, organelle, or embryo comprising a target polynucleotide an RNP complex. The RNP complex can be one that has been purified from a biological sample, recombinantly produced and subsequently purified, or *in vitro*-assembled as described herein. In those embodiments wherein the RNP complex that is contacted with the target sequence or a cell organelle, or embryo has been assembled *in vitro*, the method can further comprise the *in vitro* assembly of the complex prior to contact with the target polynucleotide, cell, organelle, or embryo.

A purified or *in vitro* assembled RNP complex can be introduced into a cell, organelle, or embryo using any method known in the art, including, but not limited to electroporation. Alternatively, an RGN polypeptide, base editor, PE, and/or polynucleotide encoding or comprising the guide RNA can be introduced into a cell, organelle, or embryo using any method known in the art (*e.g.*, electroporation). In some embodiments, delivery of a polynucleotide encoding a guide RNA, PEgRNA, RGN polypeptide, base editor, and/or PE to a cell or embryo is not a method for treatment of the human or animal body by therapy or is not a process for modifying the germ line genetic identity of a human being. In some embodiments, a method comprising contacting a target polynucleotide with an RNP complex is not a method for treatment of the human or animal body by therapy or is not a process for modifying the germ line genetic identity of a human being. In some embodiments, the embryo is a non-human embryo.

Upon delivery to or contact with the target polynucleotide or cell, organelle, or embryo comprising the target polynucleotide, the guide RNA directs the RGN, base editor, or PE to bind to the target sequence

in a sequence-specific manner. In those embodiments wherein the RGN has nuclease activity, the RGN polypeptide cleaves the target sequence of interest upon binding. The target sequence can subsequently be modified via endogenous repair mechanisms, such as non-homologous end joining, or homology-directed repair with a provided donor polynucleotide. In those embodiments wherein the base editor comprises an RGN nickase, a single strand of a double-stranded DNA target sequence is nicked and the base editing polypeptide (e.g., deaminase) of the base editor edits a nucleobase in the target sequence. In those embodiments wherein the PE comprises an RGN nickase, a single strand of a double-stranded DNA target sequence is nicked and the DNA synthesis template within the extension arm of the associated PEgRNA is used as a template to introduce a desired edit into the target sequence.

Methods to measure binding of an RGN polypeptide to a target sequence are known in the art and include chromatin immunoprecipitation assays, gel mobility shift assays, DNA pull-down assays, reporter assays, microplate capture and detection assays. Likewise, methods to measure cleavage or modification of a target sequence are known in the art and include *in vitro* or *in vivo* cleavage assays wherein cleavage is confirmed using PCR, sequencing, or gel electrophoresis, with or without the attachment of an appropriate label (*e.g.*, radioisotope, fluorescent substance) to the target sequence to facilitate detection of degradation products. Alternatively, the nicking triggered exponential amplification reaction (NTEXPAR) assay can be used (see, *e.g.*, Zhang et al. (2016) *Chem. Sci.* 7:4951-4957). *In vivo* cleavage can be evaluated using the Surveyor assay (Guschin et al. (2010) *Methods Mol Biol* 649:247-256).

In some embodiments, the methods involve the use of a single type of RGN, base editor, or PE complexed with more than one guide RNA. The more than one guide RNA can target different regions of a single gene or can target multiple genes.

In some embodiments, the methods comprise dual polymerase editing wherein a single PE (or a polynucleotide encoding the same) and two PEgRNAs (or polynucleotides encoding the same) that associate with the PE are contacted with a target DNA molecule for the programmable replacement or excision of DNA sequences. Alternatively, a first PE (or a polynucleotide encoding the same) and its associated first PEgRNA (or a polynucleotide encoding the same) and a second PE (or a polynucleotide encoding the same) and its associated second PEgRNA (or a polynucleotide encoding the same) are contacted with a target DNA molecule, wherein the first and second PEs are different from each other. In both scenarios, the two PEgRNAs bind to opposite strands of a target DNA so that the intended repair event is installed on both strands of the target DNA. The DNA synthesis template of the first PEgRNA and the DNA synthesis template of the second PEgRNA each encode a single-stranded DNA sequence that is complementary (full or partial) to each other so that the region of the DNA between the two nicked sites is replaced. For large insertions, the DPE DNA synthesis templates are designed such that the replacement sequence comprises a first recombinase site, and a donor DNA comprising a second recombinase site is introduced, as well as the corresponding site-specific recombinase that recognizes the first and second recombinase site, which results in site-specific integration of exogenous DNA into the target DNA molecule.

In those embodiments wherein a donor polynucleotide is not provided, a double-stranded break introduced by an RGN polypeptide can be repaired by a non-homologous end-joining (NHEJ) repair process. Due to the error-prone nature of NHEJ, repair of the double-stranded break can result in a modification to the target sequence. As used herein, a "modification" in reference to a nucleic acid molecule refers to a change in the nucleotide sequence of the nucleic acid molecule, which can be a deletion, insertion, or substitution of one or more nucleotides, or a combination thereof. Modification of the target sequence can result in the expression of an altered protein product or inactivation of a coding sequence.

In those embodiments wherein a donor polynucleotide is present, the donor sequence in the donor polynucleotide can be integrated into or exchanged with the target nucleotide sequence during the course of repair of the introduced double-stranded break, resulting in the introduction of the exogenous donor sequence. A donor polynucleotide thus comprises a donor sequence that is desired to be introduced into a target sequence of interest. In some embodiments, the donor sequence alters the original target nucleotide sequence such that the newly integrated donor sequence will not be recognized and cleaved by the RGN. Integration of the donor sequence can be enhanced by the inclusion within the donor polynucleotide of flanking sequences, referred to herein as "homology arms" that have substantial sequence identity with the sequences flanking the target nucleotide sequence, allowing for a homology-directed repair process. In some embodiments, homology arms have a length of at least 30 base pairs, at least 35 base pairs, at least 40 base pairs, at least 45 base pairs, at least 50 base pairs, at least 55 base pairs, at least 60 base pairs, at least 65 base pairs, at least 70 base pairs, at least 75 base pairs, at least 80 base pairs, at least 85 base pairs, at least 90 base pairs, at least 95 base pairs, at least 100 base pairs, and up to 2000 base pairs or more, and have at least 90%, at least 95%, or more, sequence homology to their corresponding sequence within the target nucleotide sequence.

In those embodiments wherein the RGN polypeptide introduces double-stranded staggered breaks, the donor polynucleotide can comprise a donor sequence flanked by compatible overhangs, allowing for direct ligation of the donor sequence to the cleaved target nucleotide sequence comprising overhangs by a non-homologous repair process during repair of the double-stranded break.

In those embodiments wherein the method involves the use of an RGN that is a nickase (*i.e.*, is only able to cleave a single strand of a double-stranded polynucleotide), the method can comprise introducing two RGN nickases that target identical or overlapping target sequences and cleave different strands of the polynucleotide. For example, an RGN nickase that only cleaves the positive (+) strand of a double-stranded polynucleotide can be introduced along with a second RGN nickase that only cleaves the negative (-) strand of a double-stranded polynucleotide.

In various embodiments, a method is provided for binding a target nucleotide sequence and detecting the target sequence, wherein the method comprises introducing into a cell, organelle, or embryo at least one guide RNA or a polynucleotide encoding the same, and at least one RGN polypeptide or a polynucleotide encoding the same, expressing the guide RNA and/or RGN polypeptide (if coding sequences

are introduced), wherein the RGN polypeptide is a nuclease-dead RGN and further comprises a detectable label, and the method further comprises detecting the detectable label. The detectable label may be fused to the RGN as a fusion protein (*e.g.*, fluorescent protein) or may be a small molecule conjugated to or incorporated within the RGN polypeptide that can be detected visually or by other means.

Also provided herein are methods for modulating the expression of a target sequence or a gene of interest under the regulation of a target sequence. The methods comprise introducing into a cell, organelle, or embryo at least one guide RNA or a polynucleotide encoding the same, and at least one RGN polypeptide or a polynucleotide encoding the same, expressing the guide RNA and/or RGN polypeptide (if coding sequences are introduced), wherein the RGN polypeptide is a nuclease-dead RGN. In some of these embodiments, the nuclease-dead RGN is a fusion protein comprising an expression modulator domain (*i.e.*, epigenetic modification domain, transcriptional activation domain or a transcriptional repressor domain) as described herein.

The present disclosure also provides methods for binding and/or modifying a target nucleotide sequence of interest. The methods include delivering a system comprising at least one guide RNA or a polynucleotide encoding the same, and at least one fusion polypeptide comprising an RGN of the invention and a base-editing polypeptide, for example a cytosine deaminase or an adenine deaminase, or a polynucleotide encoding the fusion polypeptide, to the target sequence or a cell, organelle, or embryo comprising the target sequence. The methods include delivering a system comprising at least one guide RNA or a polynucleotide encoding the same, and at least one fusion polypeptide comprising an RGN of the invention and a polymerase editing polypeptide, for example a DNA polymerase or a reverse transcriptase, or a polynucleotide encoding the fusion polypeptide, to the target sequence or a cell, organelle, or embryo comprising the target sequence.

In some embodiments wherein a fusion polypeptide comprising an RGN and a base-editing polypeptide is utilized, the binding of the fusion polypeptide to a target sequence results in the modification of nucleotide(s) adjacent to the target sequence. The nucleobase adjacent to the target sequence that is modified by the deaminase may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 base pairs from the 5' or 3' end of the target sequence.

In some embodiments wherein a PE or a PE system disclosed herein is utilized, the binding of the PE or PE system results in the modification of nucleotide(s) within or adjacent to the target sequence through the use of a DNA synthesis template of varying lengths such that the editing window using a polymerase editor can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 base pairs from the 5' or 3' end of the target sequence.

In order to minimize removal of the edit inserted by the polymerase editor by the DNA mismatch repair machinery, one or more components of the mismatch repair system can be inactivated using, for example, a dominant negative version thereof. A non-limiting example of a component of the mismatch

repair system that can be inactivated or have its activity reduced is MLH1. A dominant negative MLH1 or a polynucleotide encoding the same can be introduced into the cell, along with, preceding or subsequent to the introduction of the PE system. A non-limiting example of a dominant-negative MLH1 is the sequence set forth as SEQ ID NO: 256. In some embodiments, the DNA construct encoding the PE comprises the dominant-negative MLH1-encoding sequence on its N-terminus or its C-terminus and the dominant-negative MLH1-encoding sequence can connect to the PE-encoding sequence through a peptide linker and/or a NLS.

One of ordinary skill in the art will appreciate that any of the presently disclosed methods can be used to target a single target sequence or multiple target sequences. Thus, methods comprise the use of a single RGN polypeptide in combination with multiple, distinct guide RNAs, which can target multiple, distinct sequences within a single gene and/or multiple genes. Also encompassed herein are methods wherein multiple, distinct guide RNAs are introduced in combination with multiple, distinct RGN polypeptides. These guide RNAs and guide RNA/RGN polypeptide systems can target multiple, distinct sequences within a single gene and/or multiple genes.

Also provided herein is a method of increasing efficiency of cleaving and/or modifying a nucleic acid molecule comprising a target sequence. In some embodiments, the method comprises delivering an RGN system, a base editing system, a PE system, or an RNP complex described herein to a target sequence or to a cell comprising the target sequence. The RGN system, base editing system, PE system, or the RNP complex comprises a variant LPG10145 RGN described herein such that cleavage or modification of the target nucleic acid molecule occurs at greater efficiency as compared to cleavage or modification of the target nucleic acid molecule by a method comprising delivering to the same target nucleic acid molecule a reference RGN system, base editing system, PE system, or RNP complex. In some embodiments, the reference RGN system, base editing system, PE system, or RNP complex does not comprise a variant LPG10145 RGN described herein. In some embodiments, the reference RGN system, base editing system, PE system, or RNP complex comprises an LPG10145 RGN having the amino acid sequence set forth as SEQ ID NO: 1. In some embodiments, the reference RGN system, base editing system, PE system, or RNP complex comprises a variant LPG10145 RGN that is not identical to the variant LPG10145 RGN of the RGN system, base editing system, PE system, or RNP complex being tested for efficiency of cleaing and/or modifying a nucleic acid molecule.

In some embodiments, the efficiency of cleaving and/or modifying a target sequence is increased by 2-fold to 100-fold, or 2-fold to 80-fold, or 2-fold to 5-fold. In some embodiments, the efficiency of cleaving and/or modifying a target sequence is increased by 2-fold, 2.5-fold, 3-fold, 3.5-fold, 4-fold, 4.5-fold, 5-fold, or more. In some embodiments, the efficiency of cleaving and/or modifying a target sequence is increased by at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, or more. In some embodiments, gene editing efficiency is measured by next generation sequencing, Tracking of Indels by DEcomposition (TIDE) analysis, flow cytometry, or a combination thereof. In some embodiments, the efficiency of cleaving and/or modifying the target sequence

is assessed by measuring the percentage of a target sequence or cells comprising the target sequence that comprise altered expression of the target sequence or of a polypeptide encoded by the target sequence. In some embodiments, the expression is measured by quantitative PCR, microarray, RNA-seq, flow cytometry, immunoblot, enzyme-linked immunosorbent assay (ELISA), protein immunoprecipitation, immunostaining, high performance liquid chromatography (HPLC), liquid chromatography-mass spectrometry (LC/MS), mass spectrometry, or a combination thereof. In some embodiments, the target sequence encodes a cell surface expressed protein, and the efficiency of cleaving and/or modifying the target sequence is assessed by measuring the percentage of cells comprising a reduction of the cell surface expressed protein as measured by flow cytometry.

XII. Target Polynucleotides

The disclosure provides for methods of modifying a target polynucleotide comprising a target sequence or modifying the expression of a target polynucleotide in a eukaryotic cell, which may be *in vivo*, *ex vivo*, or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal or plant (including microalgae) and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant (including micro-algae).

Using natural variability, plant breeders combine most useful genes for desirable qualities, such as yield, quality, uniformity, hardiness, and resistance against pests. These desirable qualities also include growth, day length preferences, temperature requirements, initiation date of floral or reproductive development, fatty acid content, insect resistance, disease resistance, nematode resistance, fungal resistance, herbicide resistance, tolerance to various environmental factors including drought, heat, wet, cold, wind, and adverse soil conditions including high salinity The sources of these useful genes include native or foreign varieties, heirloom varieties, wild plant relatives, and induced mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome for sources of useful genes, and in varieties having desired characteristics or traits employ the present invention to induce the rise of useful genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

The target polynucleotide of an RGN, base editing, or PE system can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the RGN, base editing, or PE system.

The precise sequence and length requirements for the PAM differ depending on the RGN used, but PAMs are typically 2-7 base pair sequences adjacent to the protospacer (that is, the target sequence).

The target polynucleotide of an RGN, base editing, or PE system may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides. Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease (e.g., a causal mutation). The causal mutation may be an expanded repeat (present at higher numbers than those in normal individuals) that leads to instability of transcribed mRNA, altered splicing, or instability of a translated product or a translated product with reduced activity. The transcribed or translated products may be known or unknown, and further may be at a normal or abnormal level. In some embodiments, the disease may be an animal disease. In some embodiments, the disease may be an avian disease. In other embodiments, the disease may be a mammalian disease. In further embodiments, the disease may be a human disease. Examples of disease-associated genes and polynucleotides in humans are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

Although RGN systems are particularly useful for their relative ease in targeting to genomic sequences of interest, there still remains an issue of what the RGN can do to address a causal mutation. One approach is to produce a fusion protein between an RGN (preferably an inactive or nickase variant of the RGN) and a base-editing enzyme or the active domain of a base editing enzyme, such as a cytosine deaminase or an adenine deaminase base editor (U.S. Patent No. 9,840,699, herein incorporated by reference). In some embodiments, the methods comprise contacting a DNA molecule with (a) a fusion protein comprising a variant LPG10145 RGN of the invention or a nickase variant thereof and a base-editing polypeptide such as a deaminase; and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA molecule; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleobase. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleobase results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence resides in an allele of a crop plant, wherein the

particular allele of the trait of interest results in a plant of lesser agronomic value. The deamination of the nucleobase results in an allele that improves the trait and increases the agronomic value of the plant.

A fusion protein can be generated comprising an RGN (preferably an inactive or nickase variant of the RGN) and a polymerase editing polypeptide, such as a DNA polymerase or reverse transcriptase as described herein. In some embodiments, the methods comprise contacting a DNA molecule with (a) a fusion protein comprising a variant LPG10145 RGN of the disclosure or a nickase variant thereof and a polymerase editing polypeptide such as a reverse transcriptase; and (b) a PEgRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA molecule; wherein the DNA molecule is contacted with the fusion protein and the PEgRNA in an amount effective and under conditions suitable for the editing of the target nucleotide sequence. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the editing of the target DNA sequence results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence resides in an allele of a crop plant, wherein the particular allele of the trait of interest results in a plant of lesser agronomic value. The editing of the target DNA sequence results in an allele that improves the trait and increases the agronomic value of the plant.

In some embodiments, the target DNA sequence comprises a T→C or A→G point mutation associated with a disease or disorder, and wherein the deamination of the mutant C or G base or editing of the target DNA sequence comprising the point mutation results in a sequence that is not associated with a disease or disorder. In some embodiments, the deamination or editing corrects a point mutation in the sequence associated with the disease or disorder.

In some embodiments, the sequence associated with the disease or disorder encodes a protein, and the deamination or editing introduces a stop codon into the sequence associated with the disease or disorder, resulting in a truncation of the encoded protein. In some embodiments, the contacting is performed *in vivo* in a subject susceptible to having, having, or diagnosed with the disease or disorder. In some embodiments, the disease or disorder is a disease associated with a point mutation, or a single-base mutation, in the genome. In some embodiments, the disease is a genetic disease, a cancer, a metabolic disease, or a lysosomal storage disease.

XIII. Pharmaceutical Compositions and Methods of Treatment

Pharmaceutical compositions of the disclosure can comprise: the presently disclosed variant LPG10145 RGN polypeptides and active variants and fragments thereof (or base editors or PEs comprising the same), as well as polynucleotides encoding the same; the presently disclosed gRNAs, or active variants and fragments thereof, or polynucleotides encoding the same; the presently disclosed RGN systems comprising variant LPG10145 RGN polypeptides and/or gRNAs; base editing systems comprising variant LPG10145 RGN polypeptides and/or gRNAs described herein; PE systems comprising variant LPG10145 RGN polypeptides and/or PEgRNAs described herein (including DPE systems); or cells comprising any of

the variant LPG10145 RGN polypeptides (or base editors or PEs comprising the same) or variant LPG10145 RGN-encoding, polymerase-encoding, base editing polypeptide-encoding polynucleotides, gRNA or gRNA-encoding polynucleotides; or the RGN, base editing, or PE systems; and a pharmaceutically acceptable carrier are provided.

A pharmaceutical composition is a composition that is employed to prevent, reduce in intensity, cure or otherwise treat a target condition or disease that comprises an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) and a pharmaceutically acceptable carrier.

As used herein, a "pharmaceutically acceptable carrier" refers to a material that does not cause significant irritation to an organism and does not abrogate the activity and properties of the active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these). Carriers must be of sufficiently high purity and of sufficiently low toxicity to render them suitable for administration to a subject being treated. The carrier can be inert, or it can possess pharmaceutical benefits. In some embodiments, a pharmaceutically acceptable carrier comprises one or more compatible solid or liquid filler, diluents or encapsulating substances which are suitable for administration to a human or other vertebrate animal. In some embodiments, the pharmaceutically acceptable carrier is not naturally-occurring. In some embodiments, the pharmaceutically acceptable carrier and the active ingredient are not found together in nature.

Pharmaceutical compositions used in the presently disclosed methods can be formulated with suitable carriers, excipients, and other agents that provide suitable transfer, delivery, tolerance, and the like. A multitude of appropriate formulations are known to those skilled in the art. See, *e.g.*, Remington, The Science and Practice of Pharmacy (21st ed. 2005). Suitable formulations include, for example, powders, pastes, ointments, jellies, waxes, oils, lipids, lipid (cationic or anionic) containing vesicles (such as LIPOFECTIN vesicles), lipid nanoparticles, DNA conjugates, anhydrous absorption pastes, oil-in-water and water-in-oil emulsions, emulsions carbowax (polyethylene glycols of various molecular weights), semi-solid gels, and semi-solid mixtures containing carbowax. Pharmaceutical compositions for oral or parenteral use may be prepared into dosage forms in a unit dose suited to fit a dose of the active ingredients. Such dosage forms in a unit dose include, for example, tablets, pills, capsules, injections (ampoules), suppositories, etc.

The disclosure provides for pharmaceutical compositions comprising lipid-based formulations including an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these). The lipid-based formulations can include liposomes. The lipid-based formulations can include

lipid nanoparticles (LNPs). In some embodiments, an active ingredient is encapsulated in the lipid particle and/or disposed on the surface of the lipid particle. In some embodiments, an active ingredient is covalently attached to the lipid particle. In some embodiments, an active ingredient is non-covalently associated with the lipid particle. A covalent attachment includes the sharing of electrons in a chemical bond. Non-covalent interactions include dispersed electromagnetic interactions such as hydrogen bonds, ionic bonds, van der Waals interactions, and hydrophobic bonds.

In some embodiments, an active ingredient is encapsulated in the lipid particle. The term "encapsulate" means to enclose, surround or encase. As it relates to the formulation of the compounds of the disclosure, encapsulation may be substantial, complete or partial. The term "substantially encapsulated" or "substantial encapsulation" means that greater than 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or greater of the pharmaceutical composition or active ingredient of the disclosure may be enclosed, surrounded, or encased within a delivery agent (e.g., liposome or LNP). The term "partially encapsulated" or "partial encapsulation" means that less than 50%, 40%, 30%, 20%, 10%, or less of the pharmaceutical composition or active ingredient of the disclosure may be enclosed, surrounded, or encased within the delivery agent. Encapsulation may be determined by measuring the escape or the activity of the pharmaceutical composition or active ingredient of the disclosure using fluorescence and/or electron microscopy. For example, at least 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or greater of the pharmaceutical composition or active ingredient of the disclosure is encapsulated in a delivery agent (e.g., liposome or LNP).

Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro (2011) Journal of drug delivery 2011).

Liposomes can be made from several different types of lipids (e.g. ionizable lipids, structural lipids, helper lipids, and pegylated lipids); however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro (2011) Journal of drug delivery 2011).

A conventional liposome formulation is mainly comprised of natural phospholipids and phospholipids such as 1,2-distearoryl-*sn*-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, and monosialoganglioside. In some embodiments, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE) increases stability of a liposome.

Additives may be added to liposomes in order to modify their structure and properties. In some embodiments, cholesterol and/or sphingomyelin may be added to a liposomal mixture to help stabilize the

liposomal structure and to prevent leakage of the liposomal inner cargo. In some embodiments, addition of cholesterol to a conventional liposome formulation reduces rapid release of the encapsulated active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) into the plasma. In some embodiments, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate. In some embodiments, mean liposome vesicle size is adjusted to about 50 or 100 nm. In some embodiments, Trojan Horse liposomes (also known as Molecular Trojan Horses or PEGylated immunoliposomes) may be used in pharmaceutical compositions for delivery of an active ingredient across the blood brain barrier (BBB) (described on World Wide Web at cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long). Without being bound by any theory, it is believed that neutral lipid particles with specific antibodies conjugated to the surface allows crossing of the BBB via endocytosis. In some embodiments, pharmaceutical compositions comprising Trojan Horse liposomes may be used to deliver an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) to the brain via an intravascular injection.

In some embodiments, liposomes include stable nucleic-acid-lipid particles (SNALP) (see, e.g., Morrissey et al. (2005) Nature Biotechnology 23(8):1002-1007; Zimmerman et al. (2006) Nature 441: 111-114). SNALPs include a mixture of cationic and fusogenic lipids and coated with polyethylene glycol (PEG) that allow cellular uptake and endosomal release of an active ingredient cargo. In some embodiments, a SNALP is a class of LNP and includes an ionizable lipid that is cationic at low pH (e.g., DLinDMA, COATSOME® SS-OC), a neutral helper lipid (e.g. DSPC), cholesterol, and a diffusible polyethylene glycol (PEG)-lipid (e.g. Brij S100). In some embodiments, a SNALP formulation includes the following lipids: 3-*N*-(-methoxy poly(ethylene glycol)2000) carbamoyl-1,2-dimyrestyloxy-propylamine (PEG-cDMA); 1,2-dilinoleyloxy-*N*,*N*-dimethyl-3-aminopropane (DLinDMA); 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC); and cholesterol. In some embodiments, a SNALP includes synthetic cholesterol, dipalmitoylphosphatidylcholine (DOPC), PEG-cDMA, and DLinDMA (see, e.g., Geisbert et al. (2010) Lancet 375:1896-1905). In some embodiments, a SNALP includes synthetic cholesterol, DSPC, PEG-cDMA, and DLinDMA (see, e.g., Judge et al. (2009) J. Clin. Invest. 119:661-673). In some embodiments, a SNALP formulation includes COATSOME® SS-OC, DSPC, Brij S100, and cholesterol. In some embodiments, a SNALP formulation includes an ionizable lipid, DSPC, cholesterol, and a PEG lipid. In some embodiments, a SNALP formulation includes the ionizable lipids and/or PEG lipids disclosed in WO2022173531 or WO2022173531 each of which is herein incorporated by reference in its entirety.

In some embodiments, SNALP liposomes are about 80-100 nm in size. SNALPs have been used as effective delivery molecules to highly vascularized HepG2-derived liver tumors (see, e.g., Li et al. (2012) Gene Therapy 19:775- 780).

Without being bound by any one theory, during formulation of SNALPs, the ionizable lipid serves to condense lipid with an active ingredient (e.g., a polynucleotide) during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid may mediate the fusion of a SNALP with the endosomal membrane, enabling release of the active ingredient into the cytoplasm. The PEG-lipid may stabilize the particle and reduce aggregation during formulation, and subsequently may provide a neutral hydrophilic exterior that improves pharmacokinetic properties. In some embodiments, SNALP liposomes are prepared by formulating DLinDMA and PEG-cDMA with DSPC, cholesterol and an active ingredient using a 25:1 lipid: active ingredient ratio and a 48:40:10:2 molar ratio of cholesterol: DLinDMA: DSPC: PEG-cDMA.

In some embodiments, a pharmaceutical composition of the disclosure includes LNPs. In some embodiments, lipids may be formulated with an active ingredient of the present disclosure to form LNPs. An LNP comprises a plurality of lipid molecules physically associated with each other by intermolecular forces. In some embodiments, LNPs include liposomes. In some embodiments, LNPs differ from liposomes in not having a continuous lipid bilayer. In some embodiments, LNPs comprise solid particles having a mixture of solid and liquid lipids. In some embodiments, LNPs include dendrimer lipid nanoparticles (DLNPs), SNALPs, and lipid-like nanoparticles (LLNPs). In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nanometers (nm). In some embodiments, nanoparticles have a diameter of 500 nm or less. In some embodiments, nanoparticles have a diameter ranging between 25 nm and 200 nm, or 100 nm or less. In some embodiments, nanoparticles have a diameter ranging between 35 nm and 60 nm. In some embodiments, an LNP includes a lipid particle between about 1 and about 100 nm in size.

LNPs include four components: ionizable cationic lipids, fusogenic zwitterionic phospholipids, cholesterol, and PEGylated (PEG) lipids. In some embodiments, the ionizable cationic lipid component complexes a negatively charged polynucleotide and enhances endosomal escape). In some embodiments, the phospholipid component functions in modifying lipid bilayer structure. In some embodiments, the cholesterol component helps to stabilize an LNP. In some embodiments, the PEG lipid component decreases LNP aggregation and non-specific uptake.

In some embodiments, the LNP includes an ionizable lipid that is cationic at low pH (e.g., DLinDMA, COATSOME® SS-OC), a neutral helper lipid (e.g. DSPC), cholesterol, and a diffusible polyethylene glycol (PEG)-lipid (e.g. Brij S100). In some embodiments, the LNP includes the following lipids: 3-*N*-(-methoxy poly(ethylene glycol)2000) carbamoyl-1,2-dimyrestyloxy-propylamine (PEG-cDMA); 1,2-dilinoleyloxy-*N*,*N*-dimethyl-3-aminopropane (DLinDMA); 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC); and cholesterol. In some embodiments, the LNP includes synthetic cholesterol, dipalmitoylphosphatidylcholine (DOPC), PEG-cDMA, and DLinDMA (see, e.g., Geisbert et al. (2010)

Lancet 375:1896-1905). In some embodiments, the LNP includes synthetic cholesterol, DSPC, PEG-cDMA, and DLinDMA (see, e.g., Judge et al. (2009) J. Clin. Invest. 119:661-673).

Ionizable cationic lipids useful in LNPs include: COATSOME® SS-OC, 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP); DLinDMA; l,2-dilinoleyloxy-keto-*N*,*N*-dimethyl-3-aminopropane (DlinK-DMA); 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DlinKC2-DMA); 5A2-SC8 (Zhou et al. (2016) Proc. Natl Acad. Sci. USA 113:520–525); C12-200 (Love et al. (2010) Proc. Natl Acad. Sci. USA 107:1864–1869); 246C10 (Kim et al. (2021) Sci Adv 7(9): eabf4398); cKK-E12 (Fenton et al. (2016) Advanced Materials 28(15):2939-2943); 1,2-distearyloxy-*N*,*N*-dimethyl-3-aminopropane (DSDMA); 1,2-dioleyloxy-*N*,*N* -dimethyl-3-aminopropane (DODMA); 1,2-dilinolenyloxy-*N*,*N* -dimethyl-3-aminopropane (DLenDMA); and dilinoleylmethyl-4-dimethylaminobutyrate (Dlin-MC3-DMA; Jayaraman et al. (2012) Angew Chem Int Ed Engl. 51(34): 8529–8533). Cationic lipids are further described in International Publication Nos. WO2012040184, WO2011153120, WO2011149733, WO2011090965, WO2011043913, WO2011022460, WO2012061259, WO2012054365, WO2012044638, WO2010080724, WO201021865, WO2022173531, WO2022150485 and WO2008103276, US Patent Nos. 7,893,302 and 7,404,969 and US Patent Publication No. US20100036115, each of which is herein incorporated by reference in its entirety.

Zwitterionic phospholipids useful for LNPs include DSPC, DOPE, and DOPC. PEG lipids useful for LNPs include: 1,2-dimyristoyl-rac-glycero-3-methoxypolyethylene glycol (PEG-DMG); (3-o-[2''-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG); R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG); and C16 PEG-ceramide. In some embodiments, an LNP includes 50: 10: 38.5: 1.5 molar ratio of DlinKC2-DMA or C12-200: DSPC: cholesterol: PEG-DMG (see, e.g., Basha et al. (2011) Molecular Therapy 19(12):2186-2200). In some embodiments, an LNP includes 26.5: 20: 52: 1.5 ionizable lipid: DOPE: cholesterol: PEG lipid (see, e.g., Han et al. (2022) Sci Adv 8(3): eabj6901; Kim et al. (2021) Sci Adv 7(9): eabf4398). PEG lipids are further described in WO2012099755, WO2022173531 and WO2022150485 each of which is herein incorporated by reference in its entirety. In some embodiments, the ratio of PEG in the LNP formulations may be increased or decreased and/or the carbon chain length of the PEG lipid may be modified from C14 to C18 to alter the pharmacokinetics and/or biodistribution of the LNP formulations.

In some embodiments, a LNP formulation includes COATSOME® SS-OC, DSPC, Brij S100, and cholesterol. In some embodiments, a LNP formulation includes an ionizable lipid, DSPC, cholesterol, and a PEG lipid. In some embodiments, a LNP formulation includes the ionizable lipids and/or PEG lipids disclosed in WO2022173531 or WO2022150485 each of which is herein incorporated by reference in its entirety. In some embodiments, a LNP formulation includes the ionizable lipids and/or PEG lipids disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety), DSPC, and cholesterol. In some embodiments, a LNP formulation includes any one of CAT1-CAT35 and any one of CHM-001 to CHM-016 disclosed in WO2022173531 or WO2022150485 each of which is herein incorporated by reference in its entirety. In some embodiments, a LNP formulation includes any one of

CAT1-CAT35 and any one of CHM-001 to CHM-016 disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety), DSPC, and cholesterol.

In some embodiments, the LNP of the disclosure comprises 44-60 mol % of the cationic lipid, 19-25 mol % of the helper lipid, 25-33 mol % of the structural lipid, and 0.2-0.8 mol % of the PEG-lipid, inclusive of the endpoints. In some embodiments, the LNP of the disclosure comprises 44-54 mol % of the cationic lipid, 19-25 mol % of the helper lipid, 24-32 mol % of the structural lipid, and 1.2-1.8 mol % of the PEG-lipid, inclusive of the endpoints. In some embodiments, the LNP of the disclosure comprises 44-54 mol % of the cationic lipid, 8-14 mol % of the helper lipid, 35-43 mol % of the structural lipid, and 1.2-1.8 mol % of the PEG-lipid, inclusive of the endpoints. In some embodiments, the LNP of the disclosure comprises 45-55 mol % of the cationic lipid, 5-9 mol % of the helper lipid, 36-44 mol % of the structural lipid, and 2.5-3.5 mol % of the PEG-lipid, inclusive of the endpoints.

In some embodiments, the LNP of the disclosure comprises 49 mol % of the cationic lipid, 22 mol % of the helper lipid, 28.5 % of the structural lipid, and .05 mol % of the PEG-lipid. In some embodiments, the LNP of the disclosure comprises 49 mol % of SS-OC, 22 mol % DSPC, 28.5 mol % cholesterol, and .05 mol % Brij S100.

In some embodiments, the LNP of the disclosure comprises 50 mol % of the cationic lipid, 7 mol % of the helper lipid, 40 % of the structural lipid, and 3 mol % of the PEG-lipid. In some embodiments, the LNP of the disclosure comprises 50 mol % of the ionizable lipid, 7 mol % DSPC, 40 mol % cholesterol, and .05 mol % the PEG-lipid. In some embodiments, the LNP of the disclosure comprises 50 mol % any one of CAT1-CAT35 disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety), 7 mol % DSPC, 40 mol % cholesterol, and .05 mol % any one of CHM-001 – CHM-016 disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety). In some embodiments, the LNP of the disclosure comprises 50 mol % CAT7 disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety), 7 mol % DSPC, 40 mol % cholesterol, and .05 mol % CHM-006 disclosed in WO2022173531 or WO2022150485 (each of which is herein incorporated by reference in its entirety).

In some embodiments, LNPs are about 80-100 nm in size. In some embodiments, the LNPs are about 80 nm, about 81 nm, about 82 nm, about 83 nm, about 84 nm, about 85 nm, about 86 nm, about 87 nm, about 88 nm, about 89 nm, about 90 nm, about 91 nm, about 92 nm, about 93 nm, about 94 nm, about 95 nm, about 96 nm, about 97 nm, about 98 nm, about 99 nm, or about 100 nm in size. The individual LNPs in a population of LNPs may vary in size by about +5 nm.

In some embodiments, the charge of an LNP is taken into consideration. Cationic lipids may combine with negatively charged lipids to induce non-bilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Basha et al. (2011) Molecular Therapy 19(12):2186-2200). Negatively charged polymers such as polynucleotides may be loaded into LNPs at low

pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times.

Preparation of LNPs and encapsulation of an active ingredient are described in e.g., Basha et al. (2011) Molecular Therapy 19(12):1286-2200; Han et al. (2022) Sci Adv 8(3): eabj6901; Kim et al. (2021) Sci Adv 7(9): eabf4398; Finn et al. (2018) Cell Reports 22:2227–2235; Wei et al. (2020) Nature Communications 11:3232; WO2011127255; and WO2008103276. Lipids are commercially available (e.g., from Tekmira Pharmaceuticals, Vancouver, Canada; Avanti Polar Lipids, Inc., Alabaster, AL) or may be synthesized (e.g., Kim et al. (2021) Sci Adv 7(9): eabf4398). Synthesis of cationic lipids are also described in International Publication Nos. WO2012040184, WO2011153120, WO2011149733, WO2011090965, WO2011043913, WO2011022460, WO2012061259, WO2012054365, WO2012044638, WO2010080724 and WO201021865. Cholesterol is commercially available (e.g., from Sigma-Aldrich, St Louis, MO).

In some embodiments, encapsulation may be performed by dissolving lipid mixtures comprising cationic lipid (e.g., Dlin-DMA): phospholipid (e.g., DSPC, DOPE): cholesterol: PEG-lipid (e.g., at 40: 10: 40: 10 molar ratio) in ethanol. An active ingredient (e.g., a polynucleotide comprising or encoding a guide RNA or RGN of the disclosure) may be dissolved in an acidic buffer (e.g., citrate, acetate), pH 3 or 4. In some embodiments, the lipid solution and active ingredient solution may be mixed using a microfluidics system (Chen et al. (2012) J. Amer. Chem. Soc. 134:6948-6951; e.g., NanoAssemblr from Precision Nanosystems) or by dropwise addition of the lipid solution to the active ingredient solution. Removal of ethanol and neutralization of formulation buffer may be performed by dialysis for, e.g., 16 hours or overnight, against phosphate-buffered saline (PBS) using dialysis cassettes (e.g., 3500 molecular weight cut-off cassettes from Life Technologies). Dynamic light scattering may be used to assess LNP size, polydispersity index (PDI), and zeta potential. Encapsulation efficiency of an active ingredient such as RNA may be determined by assays such as Quant-it™ Ribogreen Assay (Thermo Fisher). In some embodiments wherein the encapsulated active ingredient is a polynucleotide, the polynucleotide may be extracted from the eluted nanoparticles and quantified at 260 nm. LNP pKa may be assessed using a 2-(*p*-toluidino)-6-napthalene sulfonic acid (TNS) assay (Zhang et al. (2011) Langmuir 27(5):1907–1914). In some embodiments, a final lipid: active ingredient weight ratio includes 12: 1, 11: 1, 10: 1, 9: 1, 8: 1, 7: 1, 6: 1, and 5: 1.

In some embodiments where a pharmaceutical composition comprises a ribonucleoprotein (RNP) complex encapsulated in an LNP, inclusion of an additional permanent cationic lipid (e.g., 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP)) allows formation of LNPs comprising RNP by mixing an ethanol solution of lipids with a solution of RNP at physiological pH (e.g., PBS buffer; Wei et al. (2020) Nature Communications 11:3232). In some embodiments, the permanent cationic lipid is included at 10 to 20 mole % of total lipids in LNPs.

In some embodiments, the LNP formulations described herein may additionally comprise a permeability enhancer molecule. Non-limiting permeability enhancer molecules are described in US2005/0222064.

In some embodiments, the LNP compositions are biodegradable, in that they do not accumulate to cytotoxic levels *in vivo* at a therapeutically effective dose. LNP formulations may be improved by replacing the cationic lipid with a biodegradable cationic lipid which is known as a rapidly eliminated lipid nanoparticle (reLNP). In some embodiments, the rapid metabolism of the rapidly eliminated lipids can improve the tolerability and therapeutic index of LNPs by an order of magnitude from a 1 mg/kg dose to a 10 mg/kg dose in rat. Inclusion of an enzymatically degraded ester linkage can improve the degradation and metabolism profile of the cationic component, while still maintaining the activity of the reLNP formulation. The ester linkage can be internally located within the lipid chain or it may be terminally located at the terminal end of the lipid chain. The internal ester linkage may replace any carbon in the lipid chain.

In some embodiments, the LNP compositions do not cause an innate immune response that leads to substantial adverse effects at a therapeutic dose level. In some embodiments, the LNP compositions provided herein do not cause toxicity at a therapeutic dose level.

In some embodiments, the active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) is formulated as a solid lipid nanoparticle. A solid lipid nanoparticle (SLN) may be spherical with an average diameter between 10 to 1000 nm. SLN possess a solid lipid core matrix that can solubilize lipophilic molecules and may be stabilized with surfactants and/or emulsifiers. In some embodiments, the lipid nanoparticle may be a self-assembly lipid-polymer nanoparticle (see, e.g., Zhang et al. (2008) ACS Nano 2(8):1696-1702).

In some embodiments, a lipid-based formulation including an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) can be formulated for controlled release and/or targeted delivery. As used herein, "controlled release" refers to a pharmaceutical composition or compound release profile that conforms to a particular pattern of release to effect a therapeutic outcome.

In some embodiments, a lipid-based formulation including an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) includes at least one controlled release coating. Controlled release coatings include: OPADRY® (Colorcon Inc., Harleysville, PA); polyvinylpyrrolidone/vinyl acetate copolymer; polyvinylpyrrolidone; hydroxypropyl methylcellulose;

hydroxypropyl cellulose; hydroxyethyl cellulose; EUDRAGIT RL® (Evonik, Essen, Germany); EUDRAGIT RS® (Evonik, Essen, Germany); and cellulose derivatives such as ethylcellulose aqueous dispersions (AQUACOAT® and SURELEASE®, Colorcon Inc., Harleysville, PA). In some embodiments, the controlled release and/or targeted delivery formulation may comprise at least one degradable polyester which may contain polycationic side chains. Degradable polyesters include poly(serine ester), poly(L-lactide-*co*-L-lysine), poly(4-hydroxy-L-proline ester), and combinations thereof. In some embodiments, the degradable polyesters may include a PEG conjugation to form a PEGylated polymer.

In some embodiments, LNP formulations may be prepared such that they passively or actively are directed to different cell types *in vivo*, including hepatocytes, immune cells, tumor cells, endothelial cells, antigen presenting cells, and leukocytes (Akinc et al. (2010) Mol Ther. 18: 1357-1364; Song et al. (2005) Nat Biotechnol. 23:709-717; Judge et al. (2009) J Clin Invest. 119:661-673; Kaufmann et al. (2010) Microvasc Res 80:286-293; Santel et al. (2006) Gene Ther 13:1222-1234; Santel et al. (2006) Gene Ther 13:1360-1370; Gutbier et al. (2010) Pulm Pharmacol. Ther. 23:334-344; Basha et al. (2011) Mol. Ther. 19:2186-2200; Fenske and Cullis (2008) Expert Opin Drug Deliv. 5:25-44; Peer et al. (2008) Science 319:627-630; Peer and Lieberman (2011) Gene Ther. 18:1127-1133; all of which are incorporated herein by reference in their entirety). One example of passive targeting of formulations to liver cells includes the Dlin-DMA, Dlin-KC2-DMA and MC3 -based lipid nanoparticle formulations which have been shown to bind to apolipoprotein E and promote binding and uptake of these formulations into hepatocytes *in vivo* (Akinc et al. (2010) Mol Ther. 18: 1357-1364).

LNP formulations can also be selectively targeted through expression of different ligands on their surface such as folate, transferrin, N-acetylgalactosamine (GalNAc), and antibody targeted approaches (Kolhatkar et al. (2011) Curr Drug Discov Technol. 8:197-206; Musacchio and Torchilin (2011) Front Biosci. 16:1388-1412; Yu et al. (2010) Mol Membr Biol. 27:286-298; Patil et al. (2008) Crit Rev Ther Drug Carrier Syst. 25:1- 61; Benoit et al. (2011) Biomacromolecules. 12:2708-2714; Zhao et al. (2008) Expert Opin Drug Deliv. 5:309-319; Akinc et al. (2010) Mol Ther. 18: 1357-1364; Srinivasan et al. (2012) Methods Mol Biol. 820:105-116; Ben-Arie et al. (2012) Methods Mol Biol. 757:497- 507; Peer, D (2010) J of controlled release 148(1):63-68; Peer et al. (2007) Proc Natl Acad Sci USA. 104:4095-4100; Kim et al. (2011) Methods Mol Biol. 721:339-353; Subramanya et al. (2010) Mol Ther. 18:2028-2037; Song et al. (2005) Nat Biotechnol. 23:709-717; Peer et al. (2008) Science 319:627-630; Peer and Lieberman (2011) Gene Ther. 18:1127-1133; all of which are incorporated herein by reference in their entirety).

In some embodiments, an active ingredient (i.e., RGN polypeptides, RGN-encoding polynucleotides, base editors, base editor-encoding polynucleotides, PEs, PE-encoding polynucleotides, gRNA, gRNA-encoding polynucleotides, RNP complexes, RGN systems, base editing systems, PE systems, or cells comprising any one of these) may be encapsulated into an LNP and the LNP may then be encapsulated into a polymer, polymer matrix, hydrogel and/or surgical sealant described herein and/or known in the art. In some embodiments, the polymer, hydrogel or surgical sealant includes: poly(lactic-*co*-

glycolic acid (PLGA); ethylene vinyl acetate (EVAc); poloxamer; GELSITE® (Nanotherapeutics, Inc. Alachua, FL); HYLENEX® (Halozyme Therapeutics, San Diego CA); surgical sealants such as fibrinogen polymers (Ethicon Inc., Cornelia, GA) and TISSELL® (Baxter International, Inc Deerfield, IL); PEG-based sealants; and COSEAL® (Baxter International, Inc Deerfield, IL).

LNPs and LNP formulations are further described in, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658; European Pat. Nos. 1766035; 1519714; 1781593; and 1664316.

In embodiments wherein cells comprising or modified with the presently disclosed RGN polypeptides, base editors, PEs, gRNAs, RNP complexes, RGN systems, base editing systems, PE systems, or polynucleotides encoding the same are administered to a subject, the cells are administered as a suspension with a pharmaceutically acceptable carrier. One of skill in the art will recognize that a pharmaceutically acceptable carrier to be used in a cell composition will not include buffers, compounds, cryopreservation agents, preservatives, or other agents in amounts that substantially interfere with the viability of the cells to be delivered to the subject. A formulation comprising cells can include e.g., osmotic buffers that permit cell membrane integrity to be maintained, and optionally, nutrients to maintain cell viability or enhance engraftment upon administration. Such formulations and suspensions are known to those of skill in the art and/or can be adapted for use with the cells described herein using routine experimentation.

A cell composition can also be emulsified or presented as a liposome composition, provided that the emulsification procedure does not adversely affect cell viability. The cells and any other active ingredient can be mixed with excipients that are pharmaceutically acceptable and compatible with the active ingredient, and in amounts suitable for use in the therapeutic methods described herein.

Additional agents included in a cell composition can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids, such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases, such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

Physiologically tolerable and pharmaceutically acceptable carriers are well known in the art. Exemplary liquid carriers are sterile aqueous solutions that contain no materials in addition to the active ingredients and water, or contain a buffer such as sodium phosphate at physiological pH value, physiological saline or both, such as phosphate-buffered saline. Still further, aqueous carriers can contain more than one buffer salt, as well as salts such as sodium and potassium chlorides, dextrose, polyethylene glycol and other solutes. Liquid compositions can also contain liquid phases in addition to and to the exclusion of water. Exemplary of such additional liquid phases are glycerin, vegetable oils such as cottonseed oil, and water-oil

emulsions. The amount of an active compound used in the cell compositions that is effective in the treatment of a particular disorder or condition can depend on the nature of the disorder or condition, and can be determined by standard clinical techniques.

The presently disclosed RGN polypeptides, base editors, PEs, guide RNAs, RNP complexes, RGN systems, base editing systems, PE systems, or polynucleotides encoding the same can be formulated with pharmaceutically acceptable excipients such as carriers, solvents, stabilizers, adjuvants, diluents, etc., depending upon the particular mode of administration and dosage form. In some embodiments, these pharmaceutical compositions are formulated to achieve a physiologically compatible pH, and range from a pH of about 3 to a pH of about 11, about pH 3 to about pH 7, depending on the formulation and route of administration. In some embodiments, the pH can be adjusted to a range from about pH 5.0 to about pH 8. In some embodiments, the compositions can comprise a therapeutically effective amount of at least one compound as described herein, together with one or more pharmaceutically acceptable excipients. In some embodiments, the compositions comprise a combination of the compounds described herein, or include a second active ingredient useful in the treatment or prevention of bacterial growth (for example and without limitation, anti-bacterial or anti-microbial agents), or include a combination of reagents of the present disclosure.

Suitable excipients include, for example, carrier molecules that include large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Other exemplary excipients can include antioxidants (for example and without limitation, ascorbic acid), chelating agents (for example and without limitation, EDTA), carbohydrates (for example and without limitation, dextrin, hydroxyalkylcellulose, and hydroxyalkylmethylcellulose), stearic acid, liquids (for example and without limitation, oils, water, saline, glycerol and ethanol), wetting or emulsifying agents, pH buffering substances, and the like.

In some embodiments, the formulations are provided in unit-dose or multi-dose containers, for example sealed ampules and vials, and may be stored in a freeze-dried (lyophilized) condition requiring the addition of the sterile liquid carrier, for example, saline, water-for-injection, a semi-liquid foam, or gel, immediately prior to use. Extemporaneous injection solutions and suspensions may be prepared from sterile powders, granules and tablets of the kind previously described. In some embodiments, the active ingredient is dissolved in a buffered liquid solution that is frozen in a unit-dose or multi-dose container and later thawed for injection or kept/stabilized under refrigeration until use.

The therapeutic agent(s) may be contained in controlled release systems. In order to prolong the effect of a drug, it often is desirable to slow the absorption of the drug from subcutaneous, intrathecal, or intramuscular injection. This may be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug then depends upon its rate of dissolution which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form is accomplished by dissolving or suspending the drug in

an oil vehicle. In some embodiments, the use of a long-term sustained release implant may be particularly suitable for treatment of chronic conditions. Long-term sustained release implants are well-known to those of ordinary skill in the art.

Methods of treating a disease in a subject in need thereof are provided herein. The methods comprise administering to a subject in need thereof an effective amount of: a presently disclosed RGN polypeptide or active variant or fragment thereof or a polynucleotide encoding the same; a base editor, or active variant or fragment thereof, described herein, or a polynucleotide encoding the same; a PE, or active variant or fragment thereof, described herein, or a polynucleotide encoding the same; a presently disclosed gRNA or a polynucleotide encoding the same; a presently disclosed RGN, base editing, or PE system; or a cell modified by or comprising any one of these compositions.

In some embodiments, the treatment comprises *in vivo* gene editing by administering a presently disclosed RGN polypeptide, base editor, PE, gRNA, or system (e.g., RGN, base editing, or PE), or polynucleotide(s) encoding the same. In some embodiments, the treatment comprises *ex vivo* gene editing wherein cells are genetically modified *ex vivo* with a presently disclosed RGN polypeptide, base editor, PE, gRNA, or system(e.g., RGN, base editing, or PE), or polynucleotide(s) encoding the same and then the modified cells are administered to a subject. In some embodiments, the genetically modified cells originate from the subject that is then administered the modified cells, and the transplanted cells are referred to herein as autologous. In some embodiments, the genetically modified cells originate from a different subject (i.e., donor) within the same species as the subject that is administered the modified cells (i.e., recipient), and the transplanted cells are referred to herein as allogeneic. In some examples described herein, the cells can be expanded in culture prior to administration to a subject in need thereof.

In some embodiments, the disease to be treated with the presently disclosed compositions is one that can be treated with immunotherapy, such as with a chimeric antigen receptor (CAR) T cell. Such diseases include but are not limited to cancer.

In some embodiments, the disease to be treated with the presently disclosed compositions is associated with a genetic defect, *e.g.*, a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. Thus, in some embodiments, compositions of the disclosure are used to edit a genomic sequence causal for the disease or disorder or symptoms of such disease or disorder. For example, the deamination of a target nucleobase or editing of a target polynucleotide results in the correction of a genetic defect such that loss of function of a gene product is restored.

In some embodiments, the disease to be treated with the presently disclosed compositions is associated with a mutation or causal mutation. As used herein, a "causal mutation" or "mutation associated with a disease" refers to a particular nucleotide, nucleotides, or nucleotide sequence in the genome that contributes to the severity or presence of a disease or disorder in a subject. The correction of the causal mutation leads to the improvement of at least one symptom resulting from a disease or disorder. In some embodiments, the causal mutation is adjacent to a PAM site recognized by a variant LPG10145 RGN

disclosed herein. The causal mutation can be corrected with a presently disclosed variant LPG10145 RGN or a fusion polypeptide comprising a presently disclosed variant LPG10145 RGN and a base-editing polypeptide (i.e., a base editor) or a fusion polypeptide comprising a presently disclosed variant LPG10145 RGN and a polymerase (i.e., a PE). Non-limiting examples of diseases associated with a mutation or causal mutation include cystic fibrosis, Hurler syndrome, Friedreich's Ataxia, Huntington's Disease, and sickle cell disease. Non-limiting examples of disease-associated genes and mutations are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a RGN, base editing, or PE system to introduce a deactivating point mutation into an oncogene (*e.g.*, in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, *e.g.*, a truncated protein lacking the function of the full-length protein. In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. The RGN, base editing, or PE system can be validated for gene editing-based human therapeutics *in vitro*, *e.g.*, by correcting a disease associated mutation in human cell culture. It will be understood by the skilled artisan that a base editor comprising an RGN and a base editing polypeptide, or a PE comprising an RGN and a polymerase, can be used to correct a single point mutation.

In some embodiments, a method of treating a disease in a subject in need thereof comprises creating an induced pluripotent stem cell (iPSC) or isolating a mesenchymal stem cell from the subject, contacting the iPSC or mesenchymal stem cell with any one of the RGN polypeptides, base editors, PE, RGN systems, base editing systems, PE systems, compositions comprising the same, or pharmaceutical compositions disclosed herein in order to genetically modify a target polynucleotide within the cell, differentiating the modified iPSC or the modified mesenchymal stem cell into a genetically modified mature cell or precursor thereof, and administering the genetically modified mature cell or precursor thereof into the subject. In some embodiments, the iPSC or the mesenchymal stem cell is an autologous or an allogeneic cell. In some embodiments, the iPSC or the mesenchymal stem cell is derived from a donor that is a perfect human leukocyte antigen (HLA) match for the subject. In some embodiments, the subject is administered a myeloablative therapy prior to administration of the modified cells.

Any method known in the art for creating patient specific iPS cells can be used, including but not limited to that described in Takahashi and Yamanaka, Cell 126(4):663-76, 2006. For example, the creating step can comprise: a) isolating a somatic cell, such as a skin cell or fibroblast, from the subject; and b) introducing a set of pluripotency-associated genes into the somatic cell in order to induce the cell to become

a pluripotent stem cell. The set of pluripotency-associated genes can be one or more of the genes selected from the group consisting of OCT4, SOX1, SOX2, SOX3, SOX15, SOX18, NANOG, KLF1, KLF2, KLF4, KLF5, c-MYC, n-MYC, REM2, TERT and LIN28. Mesenchymal stem cells can be isolated according to any method known in the art, such as from a patient's bone marrow or peripheral blood. For example, marrow aspirate can be collected into a syringe with heparin. Cells can be washed and centrifuged on a Percoll. The cells can be cultured in Dulbecco's modified Eagle's medium (DMEM) (low glucose) containing 10% fetal bovine serum (FBS) (Pittinger M F, Mackay A M, Beck S C et al., Science 1999; 284:143-147).

Genetically modified cells of the disclosure administered to a subject include autologous and allogeneic cells. Allogeneic cells refer to cells that are from a donor or donors (i.e., an individual or individuals from which the genetically modified cells are derived). Autologous cells refer to cells that are from the subject undergoing treatment (i.e., the recipient of the genetically modified cells). Due to the risk of transplant rejection, an effort is made to optimize the degree of major histocompatibility complex (MHC)/human leukocyte antigen (HLA) matching between donor tissue and recipient. HLA are found on the surface of cells and help the body in identifying self versus non-self, so that the body can attack foreign entities such as bacteria and viruses. HLA typing of donor tissue and the recipient concerns determining the genotype of six HLA antigens or alleles between a donor(s) and recipient to assess the degree to which the six HLA match. HLA alleles usually refer to two each at the loci HLA-A, HLA-B and HLA-DR, or one each at the loci HLA-A, HLA-B and HLA-C and one each at the loci HAL-DRB1, HLA-DQB1 and HLA-DPB1 (see e.g., Kawase et al., 2007, Blood 110:2235-2241). In some embodiments, 4 of 6 HLA matching between donor(s) and recipient are sufficient for administration to the recipient of cells derived from a donor. In some embodiments, 5 of 6 HLA matching between donor(s) and recipient are sufficient for administration to the recipient of cells derived from a donor(s). In some embodiments, 6 of 6 HLA are matched between donor(s) and recipient for administration to the recipient of cells derived from the donor(s). In general, a 4/6, 5/6, or a 6/6 HLA match is the standard of clinical care. When all 6 HLA match between donor(s) and recipient, the match is referred to as being a perfect match.

As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In particular embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For

example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their prevention or recurrence.

The term "effective amount" or "therapeutically effective amount" refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, and the delivery system in which it is carried.

The term "administering" refers to the placement of an active ingredient into a subject, by a method or route that results in at least partial localization of the introduced active ingredient at a desired site, such as a site of injury or repair, such that a desired effect(s) is produced. In some embodiments, the disclosure provides methods comprising delivering any of the RGN polypeptides, base editors, PEs, nucleic acid molecules, ribonucleoprotein complexes, vectors, pharmaceutical compositions and/or gRNAs described herein. In some embodiments, the disclosure further provides cells produced by such methods, and organisms (such as animals or plants) comprising or produced from such cells. In some embodiments, a RGN polypeptide (or base editor or PE comprising the same) and/or nucleic acid molecules as described herein in combination with (and optionally complexed with) a guide sequence is delivered to a cell.

In those embodiments wherein cells are administered, the cells can be administered by any appropriate route that results in delivery to a desired location in the subject where at least a portion of the implanted cells or components of the cells remain viable. The period of viability of the cells after administration to a subject can be as short as a few hours, e.g., twenty-four hours, to a few days, to as long as several years, or even the life time of the patient, i.e., long-term engraftment. For example, in some aspects described herein, an effective amount of photoreceptor cells or retinal progenitor cells is administered via a systemic route of administration, such as an intraperitoneal or intravenous route.

In some embodiments, the administering comprises administering by viral delivery. Viral vectors comprising a nucleic acid encoding the RGN polypeptides, base editors, PEs, nucleic acid molecules, ribonucleoprotein complexes, or vectors disclosed herein may be administered directly to patients (i.e., *in vivo*) or they may be used to treat cells *in vitro*, and the modified cells may optionally be administered to patients (i.e., *ex vivo*). Conventional viral based systems may include, without limitation, retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. In applications

where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division.

In embodiments where the administering comprises administering by AAV vector delivery, a therapeutically effective dosage to a human can be in the range of from about 20 to about 50 ml of saline solution containing from about 1 x 10^{10} to about 1 x 10^{50} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In some embodiments, the AAV dose can be from about 1 x 10^{5} to 1 x 10^{50} genomes AAV, from about 1 x 10^{8} to 1 x 10^{20} genomes AAV, from about 1 x 10^{10} to about 1 x 10^{16} genomes, or about 1 x 10^{11} to about 1 x 10^{16} genomes AAV. A human dosage may be about 1 x 10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2.

In some embodiments, the administering comprises administering by other non-viral delivery of nucleic acids. Exemplary non-viral delivery methods, without limitation, include RNP complexes, lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipidmucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos.5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO1991/17424; WO 1991/16024. Delivery can be to cells (e.g., in vitro or ex vivo administration) or target tissues (e.g., *in vivo* administration). In some embodiments, administration of a pharmaceutical composition of the disclosure includes daily intravenous injections of about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 mg/kg/day, or more of an active ingredient in a pharmaceutical composition comprising a liposome or LNP. In some embodiments, administration of a pharmaceutical composition comprising a liposome or LNP includes doses of about 0.01 to 1 mg per kg of body weight. In some embodiments, administration of a pharmaceutical composition comprising a liposome or LNP includes doses of about 1 to 10 mg per kg of body weight.

In some embodiments, the administering comprises administering by a method selected from the group consisting of: intravenously, subcutaneously, intramuscularly, orally, rectally, by aerosol, parenterally, ophthalmicly, pulmonarily, transdermally, vaginally, otically, nasally, and by topical administration, or any combination thereof. In some embodiments, for the delivery of cells, administration by injection or infusion is used.

As used herein, the term "subject" refers to any individual for whom diagnosis, treatment or therapy is desired. In some embodiments, the subject is an animal. In some embodiments, the subject is a mammal. In some embodiments, the subject is a human being.

The efficacy of a treatment can be determined by the skilled clinician. However, a treatment is considered an "effective treatment," if any one or all of the signs or symptoms of a disease or disorder are

altered in a beneficial manner (e.g., decreased by at least 10%), or other clinically accepted symptoms or markers of disease are improved or ameliorated. Efficacy can also be measured by failure of an individual to worsen as assessed by hospitalization or need for medical interventions (e.g., progression of the disease is halted or at least slowed). Methods of measuring these indicators are known to those of skill in the art. Treatment includes: (1) inhibiting the disease, e.g., arresting, or slowing the progression of symptoms; or (2) relieving the disease, e.g., causing regression of symptoms; and (3) preventing or reducing the likelihood of the development of symptoms.

Pharmaceutical compositions comprising the presently disclosed variant LPG10145 RGN polypeptides or polynucleotides encoding the same, the presently disclosed base editors or polynucleotides encoding the same, the presently disclosed PEs or polynucleotides encoding the same, the presently disclosed gRNAs or polynucleotides encoding the same, the presently disclosed systems (e.g., RGN, base editing, or PE), the presently disclosed ribonucleoprotein complexes or cells comprising any of the RGN polypeptides or RGN-encoding polynucleotides, base editors, PEs, gRNA or gRNA-encoding polynucleotides, or the systems (e.g., RGN, base editing, or PE), and a pharmaceutically acceptable carrier are provided.

As used herein, a "pharmaceutically acceptable carrier" refers to a material that does not cause significant irritation to an organism and does not abrogate the activity and properties of the active ingredient (e.g., an RGN polypeptide or nucleic acid molecule encoding the same). Carriers must be of sufficiently high purity and of sufficiently low toxicity to render them suitable for administration to a subject being treated. The carrier can be inert, or it can possess pharmaceutical benefits. In some embodiments, a pharmaceutically acceptable carrier comprises one or more compatible solid or liquid filler, diluents or encapsulating substances which are suitable for administration to a human or other vertebrate animal. In some embodiments, the pharmaceutical composition comprises a pharmaceutically acceptable carrier that is non-naturally occurring. In some embodiments, the pharmaceutically acceptable carrier and the active ingredient are not found together in nature and are thus, heterologous.

Pharmaceutical compositions used in the presently disclosed methods can be formulated with suitable carriers, excipients, and other agents that provide suitable transfer, delivery, tolerance, and the like. A multitude of appropriate formulations are known to those skilled in the art. See, *e.g.*, Remington, The Science and Practice of Pharmacy (21st ed. 2005). Non-limiting examples include a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. Administered intravenously, particular carriers are physiological saline or phosphate buffered saline (PBS). Pharmaceutical compositions for oral or parenteral use may be prepared into dosage forms in a unit dose suited to fit a dose of the active ingredients. Such dosage forms in a unit dose include, for example, tablets,

pills, capsules, injections (ampoules), suppositories, etc. These compositions also may contain adjuvants including preservative agents, wetting agents, emulsifying agents, and dispersing agents. Prevention of the action of microorganisms may be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, and the like. It also may be desirable to include isotonic agents, for example, sugars, sodium chloride and the like. Prolonged absorption of the injectable pharmaceutical form may be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin.

In some embodiments wherein cells comprising or modified with the presently disclosed variant LPG10145 RGNs, base editors, PEs, gRNAs, systems (e.g., RGN, base editing, or PE), or polynucleotides encoding the same are administered to a subject, the cells are administered as a suspension with a pharmaceutically acceptable carrier. One of skill in the art will recognize that a pharmaceutically acceptable carrier to be used in a cell composition will not include buffers, compounds, cryopreservation agents, preservatives, or other agents in amounts that substantially interfere with the viability of the cells to be delivered to the subject. A formulation comprising cells can include e.g., osmotic buffers that permit cell membrane integrity to be maintained, and optionally, nutrients to maintain cell viability or enhance engraftment upon administration. Such formulations and suspensions are known to those of skill in the art and/or can be adapted for use with the cells described herein using routine experimentation.

A cell composition can also be emulsified or presented as a liposome composition, provided that the emulsification procedure does not adversely affect cell viability. The cells and any other active ingredient can be mixed with excipients that are pharmaceutically acceptable and compatible with the active ingredient, and in amounts suitable for use in the therapeutic methods described herein.

Additional agents included in a cell composition can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids, such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases, such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

Suitable routes of administering the pharmaceutical compositions described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

In some embodiments, a pharmaceutical composition described herein is administered locally to a diseased site (e.g., the lung). In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, inhalation (e.g., of an aerosol), by means of a catheter, by means of a

suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber. In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, e.g., for gene editing.

In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, e.g., a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic aqueous buffer. Where necessary, the pharmaceutical can also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

In some embodiments, the pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration.

Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals or organisms of all sorts.

A. Modifying causal mutations using base-editing

An example of a genetically inherited disease which could be corrected using an approach that relies on an RGN-base editor fusion protein or RGN-polymerase fusion protein (PE) of the invention is Hurler Syndrome. Hurler Syndrome, also known as MPS-1, is the result of a deficiency of α-L-iduronidase (IDUA) resulting in a lysosomal storage disease characterized at the molecular level by the accumulation of dermatan sulfate and heparan sulfate in lysosomes. This disease is generally an inherited genetic disorder caused by mutations in the IDUA gene encoding α-L-iduronidase. Common IDUA mutations are W402X and Q70X, both nonsense mutations resulting in premature termination of translation. Such mutations are well addressed by precise genome editing (PGE) approaches, since reversion of a single nucleotide, for example by a base-editing approach, would restore the wild-type coding sequence and result in protein expression controlled by the endogenous regulatory mechanisms of the genetic locus. Additionally, since heterozygotes are known to be asymptomatic, a PGE therapy that targets one of these mutations would be useful to a large proportion of patients with this disease, as only one of the mutated alleles needs to be corrected (Bunge et al. (1994) Hum. Mol. Genet. 3(6): 861-866, herein incorporated by reference).

Current treatments for Hurler Syndrome include enzyme replacement therapy and bone marrow transplants (Vellodi et al. (1997) Arch. Dis. Child. 76(2): 92-99; Peters et al. (1998) Blood 91(7): 2601-2608, herein incorporated by reference). While enzyme replacement therapy has had a dramatic effect on the survival and quality of life of Hurler Syndrome patients, this approach requires costly and time-consuming weekly infusions. Additional approaches include the delivery of the IDUA gene on an expression vector or the insertion of the gene into a highly expressed locus such as that of serum albumin (U.S. Patent No. 9,956,247, herein incorporated by reference). However, these approaches do not restore the original IDUA locus to the correct coding sequence. A genome-editing strategy would have a number of advantages, most notably that regulation of gene expression would be controlled by the natural mechanisms present in healthy individuals. Additionally, using base editing does not necessitate causing a double stranded DNA breaks, which could lead to large chromosomal rearrangements, cell death, or oncogenecity by the disruption of tumor suppression mechanisms. A general strategy may be directed toward using RGN-base editor fusion proteins or PEs of the invention, for example those comprising variant LPG10145 RGNs, to target and correct certain disease-causing mutations in the human genome. It will be appreciated that similar approaches to target diseases that can be corrected by base-editing or polymerase editing may also be pursued. It will be further appreciated that similar approaches to target disease-causing mutations in other species, particularly common household pets or livestock, can also be deployed using the RGNs of the invention. Common household pets and livestock include dogs, cats, horses, pigs, cows, sheep, chickens, donkeys, snakes, ferrets, and fish including salmon and shrimp.

B. Modifying causal mutations by targeted deletion

RGNs of the invention could also be useful in human therapeutic approaches where the causal mutation is more complicated. For example, some diseases such as Friedreich's Ataxia and Huntington's Disease are the result of a significant increase in repeats of a three nucleotide motif at a particular region of a gene, which affects the ability of the expressed protein to function or to be expressed. Friedreich's Ataxia (FRDA) is an autosomal recessive disease resulting in progressive degeneration of nervous tissue in the spinal cord. Reduced levels of the frataxin (FXN) protein in the mitochondria cause oxidative damages and iron deficiencies at the cellular level. The reduced FXN expression has been linked to a GAA triplet expansion within the intron 1 of the somatic and germline FXN gene. In FRDA patients, the GAA repeat frequently consists of more than 70, sometimes even more than 1000 (most commonly 600-900) triplets, whereas unaffected individuals have about 40 repeats or less (Pandolfo et al. (2012) Handbook of Clinical Neurology 103: 275-294; Campuzano et al. (1996) Science 271: 1423-1427; Pandolfo (2002) Adv. Exp. Med. Biol. 516: 99-118; all herein incorporated by reference).

The expansion of the trinucleotide repeat sequence causing Friedreich's Ataxia (FRDA) occurs in a defined genetic locus within the FXN gene, referred to as the FRDA instability region. RNA guided nucleases (RGNs) may be used for excising the instability region in FRDA patient cells. This approach

requires 1) an RGN and guide RNA sequence that can be programmed to target the allele in the human genome; and 2) a delivery approach for the RGN and guide sequence. Many nucleases used for genome editing, such as the commonly used Cas9 nuclease from *S. pyogenes* (SpCas9), are too large to be packaged into adeno-associated viral (AAV) vectors, especially when considering the length of the SpCas9 gene and the guide RNA in addition to other genetic elements required for functional expression cassettes. This makes an approach using SpCas9 more difficult.

Variant LPG10145 RNA guided nucleases of the invention are well suited for packaging into an AAV vector along with a guide RNA. The present invention encompasses a strategy using RGNs of the invention in which a region of genomic instability is removed. Such a strategy is applicable to other diseases and disorders which have a similar genetic basis, such as Huntington's Disease. Similar strategies using RGNs of the invention may also be applicable to similar diseases and disorders in non-human animals of agronomic or economic importance, including dogs, cats, horses, pigs, cows, sheep, chickens, donkeys, snakes, ferrets, and fish including salmon and shrimp.

C. Modifying causal mutations by targeted mutagenesis

Variant LPG10145 RGNs of the invention could also be to introduce disruptive mutations that may result in a beneficial effect. Genetic defects in the genes encoding hemoglobin, particularly the beta globin chain (the HBB gene), can be responsible for a number of diseases known as hemoglobinopathies, including sickle cell anemia and thalassemias.

In adult humans, hemoglobin is a heterotetramer comprising two alpha (α)-like globin chains and two beta (β)-like globin chains and 4 heme groups. In adults the α2β2 tetramer is referred to as Hemoglobin A (HbA) or adult hemoglobin. Typically, the alpha and beta globin chains are synthesized in an approximate 1:1 ratio and this ratio seems to be critical in terms of hemoglobin and red blood cell (RBC) stabilization. In a developing fetus, a different form of hemoglobin, fetal hemoglobin (HbF), is produced which has a higher binding affinity for oxygen than Hemoglobin A such that oxygen can be delivered to the baby's system via the mother's blood stream. Fetal hemoglobin also contains two α globin chains, but in place of the adult β-globin chains, it has two fetal gamma (γ)-globin chains (i.e., fetal hemoglobin is α2γ2). The regulation of the switch from production of gamma- to beta-globin is quite complex, and primarily involves a down-regulation of gamma globin transcription with a simultaneous up-regulation of beta globin transcription. At approximately 30 weeks of gestation, the synthesis of gamma globin in the fetus starts to drop while the production of beta globin increases. By approximately 10 months of age, the newborn's hemoglobin is nearly all α2β2 although some HbF persists into adulthood (approximately 1-3% of total hemoglobin). In the majority of patients with hemoglobinopathies, the genes encoding gamma globin remain present, but expression is relatively low due to normal gene repression occurring around parturition as described above.

Sickle cell disease is caused by a V6E mutation in the β globin gene (HBB) (a GAG to GTG at the DNA level), where the resultant hemoglobin is referred to as "hemoglobinS" or "HbS." Under lower

oxygen conditions, HbS molecules aggregate and form fibrous precipitates. These aggregates cause the abnormality or 'sickling' of the RBCs, resulting in a loss of flexibility of the cells. The sickling RBCs are no longer able to squeeze into the capillary beds and can result in vaso-occlusive crisis in sickle cell patients. In addition, sickled RBCs are more fragile than normal RBCs, and tend towards hemolysis, eventually leading to anemia in the patient.

Treatment and management of sickle cell patients is a life-long proposition involving antibiotic treatment, pain management and transfusions during acute episodes. One approach is the use of hydroxyurea, which exerts its effects in part by increasing the production of gamma globin. Long term side effects of chronic hydroxyurea therapy are still unknown, however, and treatment gives unwanted side effects and can have variable efficacy from patient to patient. Despite an increase in the efficacy of sickle cell treatments, the life expectancy of patients is still only in the mid to late 50's and the associated morbidities of the disease have a profound impact on a patient's quality of life.

Thalassemias (alpha thalassemias and beta thalassemia) are also diseases relating to hemoglobin and typically involve a reduced expression of globin chains. This can occur through mutations in the regulatory regions of the genes or from a mutation in a globin coding sequence that results in reduced expression or reduced levels or functional globin protein. Treatment of thalassemias usually involves blood transfusions and iron chelation therapy. Bone marrow transplants are also being used for treatment of people with severe thalassemias if an appropriate donor can be identified, but this procedure can have significant risks.

One approach that has been proposed for the treatment of both sickle cell disease (SCD) and beta thalassemias is to increase the expression of gamma globin so that HbF functionally replaces the aberrant adult hemoglobin. As mentioned above, treatment of SCD patients with hydroxyurea is thought to be successful in part due to its effect on increasing gamma globin expression (DeSimone (1982) Proc Nat'l Acad Sci USA 79(14):4428-31; Ley, et al., (1982) N. Engl. J. Medicine, 307: 1469-1475; Ley, et al., (1983) Blood 62: 370-380; Constantoulakis et al., (1988) Blood 72(6):1961-1967, all herein incorporated by reference). Increasing the expression of HbF involves identification of genes whose products play a role in the regulation of gamma globin expression. One such gene is BCL11A. BCL11A encodes a zinc finger protein that expressed in adult erythroid precursor cells, and down-regulation of its expression leads to an increase in gamma globin expression (Sankaran et at (2008) Science 322: 1839, herein incorporated by reference). Use of an inhibitory RNA targeted to the BCL11A gene has been proposed (e.g., U.S. Patent Publication 2011/0182867, herein incorporated by reference) but this technology has several potential drawbacks, including that complete knock down may not be achieved, delivery of such RNAs may be problematic, and the RNAs must be present continuously, requiring multiple treatments for life.

Variant LPG10145 RGNs of the invention may be used to target the BCL11A enhancer region to disrupt expression of BCL11A, thereby increasing gamma globin expression. This targeted disruption can be achieved by non-homologous end joining (NHEJ), whereby an RGN of the invention targets to a particular sequence within the BCL11A enhancer region, makes a double-stranded break, and the cell's

machinery repairs the break, typically simultaneously introducing deleterious mutations. Similar to what is described for other disease targets, RGNs of the invention may have advantages over other known RGNs due to their relatively small size, which enables packaging expression cassettes for the RGN and its guide RNA into a single AAV vector for *in vivo* delivery. Similar strategies using RGNs of the invention may also be applicable to similar diseases and disorders in both humans and in non-human animals of agronomic or economic importance.

XI. Cells Comprising a Genetic Modification

Provided herein are cells and organisms comprising a target sequence of interest that has been modified using a process mediated by an RGN, base editor, PE, crRNA, tracrRNA, and/or gRNA, or systems comprising such, as described herein. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions differs from the corresponding amino acid residue in SEQ ID NO: 1: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is a positively charged amino acid residue: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at one or more of the following amino acid positions is an R: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 differs from the corresponding amino acid residue in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is a positively charged amino acid residue. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein the amino acid residue at amino acid position 778 and/or 969 is an R.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 differ from the corresponding amino acid residues in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 778 and 856 are positively charged amino acid residues.

In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 differ

from the corresponding amino acid residues in SEQ ID NO: 1. In some of these embodiments, the RGN comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, wherein amino acid residues at positions 55, 647, 778, and 969 are positively charged amino acid residues.

In some of these embodiments, the RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R; (b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R; (c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R; (d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R; (e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R; (f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R; (g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R; (h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R; (i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R; (j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R; (k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R; (l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R; (m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R; (n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R; (o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R; (p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R; (q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R; (r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R; (s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R; (t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R; (u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R; (v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R; (w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R; (x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R; (y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R; (z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R; (aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R; (bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R; (cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R; (dd) the amino acid sequence set forth as SEQ ID NO: 2; (ee) the amino acid sequence set forth as SEQ ID NO: 3; (ff) the amino acid sequence set forth as SEQ ID NO: 4; (gg) the amino acid sequence set forth as SEQ ID NO: 5; (hh) the amino acid sequence set forth as SEQ ID NO: 6; (ii) the amino acid sequence set forth as SEQ ID NO: 7; (jj)

the amino acid sequence set forth as SEQ ID NO: 8; (kk) the amino acid sequence set forth as SEQ ID NO: 9; (ll) the amino acid sequence set forth as SEQ ID NO: 10; (mm) the amino acid sequence set forth as SEQ ID NO: 11; (nn) the amino acid sequence set forth as SEQ ID NO: 12; (oo) the amino acid sequence set forth as SEQ ID NO: 13; (pp) the amino acid sequence set forth as SEQ ID NO: 14; (qq) the amino acid sequence set forth as SEQ ID NO: 15; and (rr) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In some of these embodiments, the RGN comprises the amino acid sequence selected from: (a) the amino acid sequence set forth as SEQ ID NO: 2; (b) the amino acid sequence set forth as SEQ ID NO: 3; (c) the amino acid sequence set forth as SEQ ID NO: 4; (d) the amino acid sequence set forth as SEQ ID NO: 5; (e) the amino acid sequence set forth as SEQ ID NO: 6; (f) the amino acid sequence set forth as SEQ ID NO: 7; (g) the amino acid sequence set forth as SEQ ID NO: 8; (h) the amino acid sequence set forth as SEQ ID NO: 9; (i) the amino acid sequence set forth as SEQ ID NO: 10; (j) the amino acid sequence set forth as SEQ ID NO: 11; (k) the amino acid sequence set forth as SEQ ID NO: 12; (l) the amino acid sequence set forth as SEQ ID NO: 13; (m) the amino acid sequence set forth as SEQ ID NO: 14; (n) the amino acid sequence set forth as SEQ ID NO: 15; and (o) the amino acid sequence set forth as SEQ ID NO: 16, or an active variant or fragment thereof.

In various embodiments, the guide RNA comprises a CRISPR repeat sequence comprising the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or an active variant or fragment thereof. In particular embodiments, the guide RNA comprises a tracrRNA comprising the nucleotide sequence set forth as SEQ ID NO:34, 246, 247, or 248, or an active variant or fragment thereof. The guide RNA of the system can be a single guide RNA or a dual-guide RNA.

The modified cells can be eukaryotic (*e.g.*, mammalian, plant, insect, avian cell) or prokaryotic. Also provided are organelles and embryos comprising at least one target nucleotide sequence that has been modified by a process utilizing an RGN, base editor, PE, crRNA, tracrRNA and/or gRNA, or systems comprising such, as described herein. The genetically modified cells, organisms, organelles, and embryos can be heterozygous or homozygous for the modified nucleotide sequence.

The chromosomal modification of the cell, organism, organelle, or embryo can result in altered expression (up-regulation or down-regulation), inactivation, or the expression of an altered protein product or an integrated sequence. In those embodiments wherein the chromosomal modification results in either the inactivation of a gene or the expression of a non-functional protein product, the genetically modified cell, organism, organelle, or embryo is referred to as a "knock out". The knock out phenotype can be the result of a deletion mutation (*i.e.*, deletion of at least one nucleotide), an insertion mutation (*i.e.*, insertion of at least one nucleotide), or a nonsense mutation (*i.e.*, substitution of at least one nucleotide such that a stop codon is introduced).

Alternatively, the chromosomal modification of a cell, organism, organelle, or embryo can produce a "knock in", which results from the chromosomal integration of a nucleotide sequence that encodes a

protein. In some of these embodiments, the coding sequence is integrated into the chromosome such that the chromosomal sequence encoding the wild-type protein is inactivated, but the exogenously introduced protein is expressed.

In other embodiments, the chromosomal modification using the presently disclosed RGNs, base editors, PEs, gRNAs, or systems comprising such, results in the production of a variant protein product. The expressed variant protein product can have at least one amino acid substitution and/or the addition or deletion of at least one amino acid. The variant protein product encoded by the altered chromosomal sequence can exhibit modified characteristics or activities when compared to the wild-type protein, including but not limited to altered enzymatic activity or substrate specificity.

In yet other embodiments, the chromosomal modification can result in an altered expression pattern of a protein. As a non-limiting example, chromosomal alterations in the regulatory regions controlling the expression of a protein product can result in the overexpression or downregulation of the protein product or an altered tissue or temporal expression pattern. In some embodiments, the mutation(s) introduced as a result of RGNs, base editors, PEs, gRNAs, or systems comprising such described herein yields a reduction or elimination in expression of a gene.

Cells that have been modified may be introduced into an organism. These cells could have originated from the same organism (e.g., person) in the case of autologous cellular transplants, wherein the cells are modified in an *ex vivo* approach. Alternatively, the cells originated from another organism within the same species (e.g., another person) in the case of allogeneic cellular transplants. The cells that have been modified can be grown into an organism, such as a plant, in accordance with conventional ways. See, for example, McCormick et al. (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same modified strain or different strains, and the resulting hybrid having the genetic modification. The present invention provides genetically modified seed. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the genetic modification. Further provided is a processed plant product or byproduct that retains the genetic modification, including for example, soymeal.

The methods provided herein may be used for modification of any plant species, including, but not limited to, monocots and dicots. Examples of plants of interest include, but are not limited to, corn (maize), sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, and oilseed rape, Brassica sp., alfalfa, rye, millet, safflower, peanuts, sweet potato, cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, papaya, cashew, macadamia, almond, oats, vegetables, ornamentals, and conifers.

Vegetables include, but are not limited to, tomatoes, lettuce, green beans, lima beans, peas, and members of the genus Curcumis such as cucumber, cantaloupe, and musk melon. Ornamentals include, but are not limited to, azalea, hydrangea, hibiscus, roses, tulips, daffodils, petunias, carnation, poinsettia, and chrysanthemum. Preferably, plants of the present invention are crop plants (for example, maize, sorghum,

wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, oilseed rape, etc.).

The methods provided herein can also be used to genetically modify any prokaryotic species, including but not limited to, archaea and bacteria (*e.g.*, *Bacillus* sp., *Klebsiella* sp. *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., Agrobacterium, *Lactobacillus* sp.).

The methods provided herein can be used to genetically modify any eukaryotic species or cells therefrom, including but not limited to animals (*e.g.*, mammals, insects, fish, birds, and reptiles), fungi, amoeba, algae, and yeast. In some embodiments, the cell that is modified by the presently disclosed methods include cells of hematopoietic origin, such as cells of the immune system including but not limited to B cells, T cells, natural killer (NK) cells, pluripotent stem cells, induced pluripotent stem cells, chimeric antigen receptor T (CAR-T) cells, monocytes, macrophages, and dendritic cells.

Cells that have been modified may be introduced into an organism. These cells could have originated from the same organism (e.g., person) in the case of autologous cellular transplants, wherein the cells are modified in an *ex vivo* approach. Alternatively, the cells originated from another organism within the same species (e.g., another person) in the case of allogeneic cellular transplants.

The article "a" and "an" are used herein to refer to one or more than one (*i.e.*, to at least one) of the grammatical object of the article. By way of example, "a polypeptide" means one or more polypeptides.

All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this disclosure pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended embodiments.

Non-limiting embodiments include:

1. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

2. The nucleic acid molecule of embodiment 1, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid

residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

3. The nucleic acid molecule of embodiment 1 or 2, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

4. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

5. The nucleic acid molecule of embodiment 4, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

6. The nucleic acid molecule of any one of embodiments 1-5, wherein said polynucleotide encoding an RGN polypeptide comprises an mRNA comprising a 5' untranslated region (UTR) and/or a 3' UTR, wherein the 5' UTR, the 3' UTR, or both are heterologous to the mRNA.

7. The nucleic acid molecule of any one of embodiments 1-5, wherein said polynucleotide encoding an RGN polypeptide is operably linked to a promoter heterologous to said polynucleotide.

8. The nucleic acid molecule of any one of embodiments 1-7, wherein said RGN polypeptide is capable of binding a target polynucleotide sequence in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target polynucleotide sequence.

9. The nucleic acid molecule of embodiment 8, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target polynucleotide sequence.

10. The nucleic acid molecule of embodiment 9, wherein said RGN polypeptide recognizes a PAM having a consensus nucleotide sequence set forth as NNGG.

11. The nucleic acid molecule of any one of embodiments 8-10, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence upon binding.

12. The nucleic acid molecule of embodiment 11, wherein said RGN polypeptide is capable of generating a double-stranded break.

13. The nucleic acid molecule of embodiment 11, wherein said RGN polypeptide is capable of generating a single-stranded break.

14. The nucleic acid molecule of any one of embodiments 1-10, wherein said RGN polypeptide is nuclease inactive or is a nickase.

15. The nucleic acid molecule of embodiment 14, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

16. The nucleic acid molecule of embodiment 14 or 15, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

17. The nucleic acid molecule of any one of embodiments 14-16, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

18. The nucleic acid molecule of any one of embodiments 1-17, wherein said RGN polypeptide is operably linked to a heterologous polypeptide.

19. The nucleic acid molecule of embodiment 18, wherein said heterologous polypeptide is operably linked to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

20. The nucleic acid molecule of embodiment 18 or 19, wherein said heterologous polypeptide is a polymerase editing polypeptide.

21. The nucleic acid molecule of embodiment 20, wherein said polymerase editing polypeptide comprises a DNA polymerase.

22. The nucleic acid molecule of embodiment 20, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

23. The nucleic acid molecule of embodiment 22, wherein said reverse transcriptase lacks an RNAse H domain.

24. The nucleic acid molecule of embodiment 22 or 23, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

25. The nucleic acid molecule of any one of embodiments 22-24, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

26. The nucleic acid molecule of embodiment 18 or 19, wherein the heterologous polypeptide is a base-editing polypeptide.

27. The nucleic acid molecule of embodiment 26, wherein the base-editing polypeptide is a deaminase.

28. The nucleic acid molecule of embodiment 27, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

29. The nucleic acid molecule of embodiment 27 or 28, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

30. The nucleic acid molecule of any one of embodiments 27-29, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

31. The nucleic acid molecule of embodiment 18 or 19, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

32. The nucleic acid molecule of embodiment 31, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

33. The nucleic acid molecule of any one of embodiments 1-32, wherein the RGN polypeptide comprises one or more nuclear localization signals.

34. The nucleic acid molecule of any one of embodiments 1-33, wherein the polynucleotide encoding the RGN polypeptide is codon optimized for expression in a eukaryotic cell.

35. A vector comprising the nucleic acid molecule of any one of embodiments 1-34.

36. The vector of embodiment 35, further comprising at least one nucleotide sequence encoding a guide RNA.

37. The vector of embodiment 36, wherein the guide RNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

38. The vector of embodiment 37, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides.

39. The vector of embodiment 37, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides.

40. The vector of embodiment 37, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides.

41. The vector of embodiment 37, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides.

42. The vector of embodiment 37, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide.

43. The vector of embodiment 37, wherein the CRISPR repeat comprises the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245.

44. The vector of any one of embodiments 36-43, wherein the guide RNA comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

45. The vector of embodiment 44, wherein the tracrRNA comprises a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, or 100% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

46. The vector of any one of embodiments 36-45, where said guide RNA is a single guide RNA.

47. The vector of embodiment 46, wherein said single guide RNA further comprises an extension comprising an edit template for polymerase editing.

48. The vector of any one of embodiments 36-45, wherein said guide RNA is a dual-guide RNA.

49. A cell comprising the nucleic acid molecule of any one of embodiments 1-34 or the vector of any one of embodiments 35-48.

50. The cell of embodiment 49, wherein the cell is a prokaryotic cell.

51. The cell of embodiment 49, wherein the cell is a eukaryotic cell.

52. The cell of embodiment 51, wherein the eukaryotic cell is a mammalian cell.

53. The cell of embodiment 52, wherein the mammalian cell is a human cell.

54. The cell of embodiment 53, wherein the human cell is an immune cell.

55. The cell of embodiment 53, wherein the human cell is a stem cell.

56. The cell of embodiment 55, wherein the stem cell is an induced pluripotent stem cell.

57. The cell of embodiment 51, wherein the eukaryotic cell is an insect or avian cell.

58. The cell of embodiment 51, wherein the eukaryotic cell is a fungal cell.

59. The cell of embodiment 51, wherein the eukaryotic cell is a plant cell.

60. A plant or plant part comprising the plant cell of embodiment 59.

61. A method for making an RGN polypeptide comprising culturing the cell of embodiment 49 under conditions in which the RGN polypeptide is expressed.

62. A method for making an RGN polypeptide comprising:

introducing into a cell a heterologous nucleic acid molecule comprising a nucleotide sequence encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

63. The method of embodiment 626, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

64. The method of embodiment 62 or 63, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

65. The method of embodiment 64, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R;

and culturing said cell under conditions in which the RGN polypeptide is expressed.

66. The method of any one of embodiments 62-65, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

67. The method of any one of embodiments 62-66, further comprising purifying said RGN polypeptide.

68. The method of any one of embodiments 62-67, wherein said cell further expresses one or more guide RNAs capable of binding to said RGN polypeptide to form an RGN ribonucleoprotein complex.

69. The method of embodiment 68, further comprising purifying said RGN ribonucleoprotein complex.

70. The method of any one of embodiments 62-69, wherein said heterologous nucleic acid molecule comprising a nucleotide sequence encoding an RNA-guided nuclease (RGN) polypeptide

comprises an mRNA comprising a 5' untranslated region (UTR) and/or a 3' UTR, wherein the 5' UTR, the 3' UTR, or both are heterologous to the mRNA.

71. The method of any one of embodiments 62-70, wherein said RGN polypeptide is capable of binding a target polynucleotide sequence in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target polynucleotide sequence.

72. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

73. The RGN polypeptide of embodiment 72, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

74. The RGN polypeptide of embodiment 72 or 73, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

75. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

76. The RGN polypeptide of embodiment 75, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

77. The RGN polypeptide of any one of embodiments 72-76, wherein said RGN polypeptide is an isolated RGN polypeptide.

78. The RGN polypeptide of any one of embodiments 72-77, wherein said RGN polypeptide is capable of binding a target polynucleotide sequence of a DNA molecule in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target polynucleotide sequence.

79. The RGN polypeptide of embodiment 78, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target polynucleotide sequence.

80. The RGN polypeptide of embodiment 79, wherein said RGN polypeptide recognizes a PAM having a consensus nucleotide sequence set forth as NNGG.

81. The RGN polypeptide of any one of embodiments 78-80, wherein the RGN polypeptide comprises a PAM-interacting domain comprising the amino acid sequence set forth as SEQ ID NO: 253.

82. The RGN polypeptide of any one of embodiments 78-81, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence upon binding.

83. The RGN polypeptide of embodiment 82, wherein cleavage by said RGN polypeptide generates a double-stranded break.

84. The RGN polypeptide of embodiment 82, wherein cleavage by said RGN polypeptide generates a single-stranded break.

85. The RGN polypeptide of any one of embodiments 72-81, wherein said RGN polypeptide is nuclease inactive or a nickase.

86. The RGN polypeptide of embodiment 85, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

87. The RGN polypeptide of embodiment 85 or 86, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

88. The RGN polypeptide of any one of embodiments 85-87, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

89. The RGN polypeptide of any one of embodiments 72-88, wherein said RGN polypeptide is operably fused to a heterologous polypeptide.

90. The RGN polypeptide of embodiment 89, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

91. The RGN polypeptide of embodiment 89 or 90, wherein said heterologous polypeptide is a polymerase editing polypeptide.

92. The RGN polypeptide of embodiment 91, wherein said polymerase editing polypeptide comprises a DNA polymerase.

93. The RGN polypeptide of embodiment 91, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

94. The RGN polypeptide of embodiment 93, wherein said reverse transcriptase lacks an RNAse H domain.

95. The RGN polypeptide of embodiment 93 or 94, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

96. The RGN polypeptide of any one of embodiments 93-95, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

97. The RGN polypeptide of embodiment 89 or 90, wherein the heterologous polypeptide is a base-editing polypeptide.

98. The RGN polypeptide of embodiment 97, wherein the base-editing polypeptide is a deaminase.

99. The RGN polypeptide of embodiment 98, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

100. The RGN polypeptide of embodiment 98 or 99, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

101. The RGN polypeptide of any one of embodiments 98-100, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

102. The RGN polypeptide of embodiment 89 or 90, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

103. The RGN polypeptide of embodiment 102, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

104. The RGN polypeptide of any one of embodiments 72-103, wherein the RGN polypeptide comprises one or more nuclear localization signals.

105. A ribonucleoprotein (RNP) complex comprising the RGN polypeptide of any one of embodiments 72-104 and the guide RNA bound to the RGN polypeptide.

106. A system, said system comprising:

A) one or more guide RNAs (gRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more gRNAs; and

B) an RNA-guided nuclease (RGN) polypeptide, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more

of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

107. The system of embodiment 106, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is a positively charged amino acid residue.

108. The system of embodiment 106 or 107, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

109. The system of embodiment 108, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

110. The system of any one of embodiments 106-109, wherein the RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

111. The system of any one of embodiments 106-110, wherein the polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide comprises an mRNA comprising a 5' untranslated region (UTR) and/or a 3' UTR, wherein the 5' UTR, the 3' UTR, or both are heterologous to the mRNA.

112. The system of any one of embodiments 106-110, wherein at least one of said nucleotide sequences encoding the one or more gRNAs and said nucleotide sequence encoding the RGN polypeptide is operably linked to a promoter heterologous to said nucleotide sequence.

113. The system of any one of embodiments 106-110, wherein the nucleotide sequences encoding the one or more gRNAs and the nucleotide sequence encoding the RGN polypeptide are located on one vector.

114. The system of any one of embodiments 106-111, wherein said RGN polypeptide and said one or more gRNAs are not found complexed to one another in nature.

115. The system of any one of embodiments 106-114, wherein said gRNA is a single guide RNA.

116. The system of embodiment 115, wherein said single guide RNA further comprises an extension comprising an edit template for polymerase editing.

117. The system of any one of embodiments 106-114, wherein said gRNA is a dual-guide RNA.

118. The system of any one of embodiments 106-117, wherein said gRNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

119. The system of embodiment 118, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides.

120. The system of embodiment 118, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides.

121. The system of embodiment 118, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides.

122. The system of embodiment 118, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides.

123. The system of embodiment 118, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide.

124. The system of embodiment 118, wherein the CRISPR repeat comprises the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245.

125. The system of any one of embodiments 118-124, wherein the guide RNA comprises a a tracrRNA having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

126. The system of embodiment 125, wherein said tracrRNA comprises a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

127. The system of any one of embodiments 106-126, wherein the one or more gRNAs are capable of hybridizing to a target polynucleotide sequence of a DNA molecule, and wherein the one or more guide RNAs are capable of forming a complex with the RGN polypeptide in order to direct said RGN polypeptide to bind to said target polynucleotide sequence of the DNA molecule.

128. The system of embodiment 127, wherein said target polynucleotide sequence is a eukaryotic target polynucleotide sequence.

129. The system of embodiment 127 or 128, wherein said target polynucleotide sequence is located adjacent and 5' to a protospacer adjacent motif (PAM).

130. The system of embodiment 129, wherein the PAM comprises a consensus nucleotide sequence of NNGG.

131. The system of any one of embodiments 127-130, wherein the target polynucleotide sequence is within a cell.

132. The system of any one of embodiments 127-131, wherein the complex comprising the one or more gRNAs and the RGN polypeptide directs cleavage of the target polynucleotide sequence.

133. The system of embodiment 132, wherein the cleavage generates a double-stranded break.

134. The system of embodiment 132, wherein the cleavage generates a single-stranded break.

135. The system of any one of embodiments 106-131, wherein said RGN polypeptide is nuclease inactive or is a nickase.

136. The system of embodiment 135, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

137. The system of embodiment 135 or 136, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

138. The system of any one of embodiments 135-137, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

139. The system of any one of embodiments 106-138, wherein said RGN polypeptide is operably linked to a heterologous polypeptide.

140. The system of embodiment 139, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

141. The system of embodiment 139 or 140, wherein the heterologous polypeptide is a polymerase editing polypeptide.

142. The system of embodiment 141, wherein the polymerase editing polypeptide comprises a DNA polymerase.

143. The system of embodiment 141, wherein the polymerase editing polypeptide comprises a reverse transcriptase.

144. The system of embodiment 143, wherein said reverse transcriptase lacks an RNAse H domain.

145. The system of embodiment 143 or 144, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

146. The nucleic acid molecule of any one of embodiments 143-145, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

147. The system of embodiment 139 or 140, wherein the heterologous polypeptide is a base-editing polypeptide.

148. The system of embodiment 147, wherein the base-editing polypeptide is a deaminase.

149. The system of embodiment 148, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

150. The system of embodiment 148 or 149, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

151. The system of any one of embodiments 148-150, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

152. The system of embodiment 139 or 140, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

153. The system of embodiment 152, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

154. The system of any one of embodiments 106-153, wherein the RGN polypeptide comprises one or more nuclear localization signals.

155. The system of any one of embodiments 106-154, wherein the polynucleotide comprising the nucleotide sequence encoding the RGN polypeptide is codon optimized for expression in a eukaryotic cell.

156. The system of any one of embodiments 106-155, wherein said system further comprises one or more donor polynucleotides.

157. A cell comprising the RGN polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, or the system of any one of embodiments 106-156.

158. The cell of embodiment 157, wherein the cell is a prokaryotic cell.

159. The cell of embodiment 157, wherein the cell is a eukaryotic cell.

160. The cell of embodiment 159, wherein the eukaryotic cell is a mammalian cell.

161. The cell of embodiment 160, wherein the mammalian cell is a human cell.

162. The cell of embodiment 161, wherein the human cell is an immune cell.

163. The cell of embodiment 161, wherein the human cell is a stem cell.

164. The cell of embodiment 163, wherein the stem cell is an induced pluripotent stem cell.

165. The cell of embodiment 159, wherein the eukaryotic cell is an insect or avian cell.

166. The cell of embodiment 159, wherein the eukaryotic cell is a fungal cell.

167. The cell of embodiment 159, wherein the eukaryotic cell is a plant cell.

168. A plant or plant part comprising the plant cell of embodiment 167.

169. A pharmaceutical composition comprising the nucleic acid molecule of any one of embodiments 1-34, the vector of any one of embodiments 35-48, the cell of embodiment 49 or 157, the RGN

polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, or the system of any one of embodiments 106-156, and a pharmaceutically acceptable carrier.

170. A method for binding a target polynucleotide sequence of a nucleic acid molecule comprising delivering a system according to any one of embodiments 106-156, to said target polynucleotide sequence or a cell comprising the target polynucleotide sequence.

171. The method of embodiment 170, wherein said RGN polypeptide or said one or more gRNAs further comprises a detectable label, thereby allowing for detection of said target polynucleotide sequence.

172. The method of embodiment 170 or 171, wherein said one or more gRNAs or said RGN polypeptide further comprises an expression modulator, thereby modulating expression of said target polynucleotide sequence or a gene under transcriptional control by said target polynucleotide sequence.

173. A method for cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule comprising delivering a system according to any one of embodiments 106-156 to said target polynucleotide sequence or a cell comprising the nucleic acid molecule, wherein cleavage or modification of said target polynucleotide sequence occurs.

174. The method of embodiment 173, wherein said modified target polynucleotide sequence comprises insertion of a heterologous polynucleotide sequence into the target polynucleotide.

175. The method of embodiment 173, wherein said modified target polynucleotide sequence comprises deletion of at least one nucleotide from the target polynucleotide sequence.

176. The method of embodiment 173, wherein said modified target polynucleotide sequence comprises mutation of at least one nucleotide in the target polynucleotide sequence.

177. A method for binding a target polynucleotide sequence of a nucleic acid molecule comprising:

a) assembling a ribonucleoprotein (RNP) complex under conditions suitable for formation of the RNP complex by combining:

i) one or more guide RNAs (gRNAs); and

ii) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1;

and

b) contacting said target polynucleotide sequence or a cell comprising said target polynucleotide sequence with the assembled RNP complex,

thereby binding said target polynucleotide sequence with said RNP complex.

178. The method of embodiment 177, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one

or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is a positively charged amino acid residue.

179. The method of embodiment 177 or 178, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

180. The method of embodiment 179, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

181. The method of embodiment 180, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

182. The method of any one of embodiments 177-181, wherein said RGN polypeptide or said one or more gRNAs further comprises a detectable label, thereby allowing for detection of said target polynucleotide sequence.

183. The method of any one of embodiments 177-182, wherein said one or more gRNAs or said RGN polypeptide further comprises an expression modulator, thereby allowing for the modulation of expression of said target polynucleotide sequence.

184. The method of any one of embodiments 177-183, wherein said RGN polypeptide further comprises a heterologous polypeptide.

185. The method of embodiment 184, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

186. The method of embodiment 184 or 185, wherein said heterologous polypeptide is a polymerase editing polypeptide, thereby allowing for the modification of said target polynucleotide sequence.

187. The method of embodiment 186, wherein said polymerase editing polypeptide comprises a DNA polymerase.

188. The method of embodiment 186, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

189. The method of embodiment 188, wherein said reverse transcriptase lacks an RNAse H domain.

190. The method of embodiment 188 or 189, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

191. The method of any one of embodiments 188-190, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

192. The method of embodiment 184 or 185, wherein said heterologous polypeptide is a base-editing polypeptide, thereby allowing for the modification of said target polynucleotide sequence.

193. The method of embodiment 192, wherein said base-editing polypeptide comprises a deaminase.

194. The method of embodiment 193, wherein said deaminase is a cytosine deaminase or an adenine deaminase.

195. The method of embodiment 193 or 194, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

196. The method of any one of embodiments 193-195, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

197. The method of embodiment 184 or 185, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

198. The method of embodiment 197, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

199. The method of any one of embodiments 177-182, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence, thereby allowing for the cleaving and/or modifying of said target polynucleotide sequence.

200. A method for cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule, comprising contacting the nucleic acid molecule with:

a) an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue

at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1; and

b) one or more guide RNAs (gRNAs) capable of targeting the RGN of (a) to the target polynucleotide sequence,

thereby cleaving and/or modifying said target polynucleotide sequence to generate a modified target polynucleotide sequence.

201. The method of embodiment 200, wherein the one or more gRNAs is capable of hybridizing to the target polynucleotide sequence and binding to said RGN polypeptide.

202. The method of embodiment 200 or 201, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

203. The method of any one of embodiments 200-202, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

204. The method of embodiment 203, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

205. The method of embodiment 204, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

206. The method of any one of embodiments 200-205, wherein cleavage by said RGN polypeptide generates a double-stranded break.

207. The method of any one of embodiments 200-205, wherein cleavage by said RGN polypeptide generates a single-stranded break.

208. The method of any one of embodiments 200-205, wherein said RGN polypeptide is nuclease inactive or is a nickase.

209. The method of embodiment 208, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

210. The method of embodiment 208 or 209, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

211. The method of any one of embodiments 208-210, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

212. The method of any one of embodiments 200-211, wherein said RGN polypeptide is operably fused to a heterologous polypeptide.

213. The method of embodiment 212, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

214. The method of embodiment 212 or 213, wherein said heterologous polypeptide is a polymerase editing polypeptide.

215. The method of embodiment 214, wherein said polymerase editing polypeptide comprises a DNA polymerase.

216. The method of embodiment 214, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

217. The method of embodiment 216, wherein said reverse transcriptase lacks an RNAse H domain.

218. The method of embodiment 216 or 217, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

219. The method of any one of embodiments 216-218, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

220. The method of embodiment 212 or 213, wherein said heterologous polypeptide is a base-editing polypeptide.

221. The method of embodiment 220, wherein the base-editing polypeptide is a deaminase.

222. The method of embodiment 221, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

223. The method of embodiment 221 or 222, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

224. The method of any one of embodiments 221-223, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

225. The method of any one of embodiments 200-219, wherein said modified target DNA sequence comprises insertion of heterologous DNA into the target DNA sequence.

226. The method of any one of embodiments 200-219, wherein said modified target DNA sequence comprises deletion of at least one nucleotide from the target DNA sequence.

227. The method of any one of embodiments 200-224, wherein said modified target DNA sequence comprises mutation of at least one nucleotide in the target DNA sequence.

228. The method of any one of embodiments 170-227, wherein said method is performed *in vitro*, *in vivo*, or *ex vivo*.

229. The method of any one of embodiments 170-228, wherein said target polynucleotide sequence is located adjacent and 5' to a protospacer adjacent motif (PAM).

230. The method of embodiment 229, wherein the PAM comprises a consensus sequence of NNGG.

231. The method of any one of embodiments 170-230, wherein said target polynucleotide sequence is a eukaryotic target DNA sequence.

232. The method of any one of embodiments 170-231, wherein said gRNA is a single guide RNA.

233. The method of embodiment 232, wherein said sgRNA further comprises an extension comprising an edit template for polymerase editing.

234. The method of any one of embodiments 170-231, wherein said gRNA is a dual-guide RNA.

235. The method of any one of embodiments 170-234, wherein said gRNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

236. The method of embodiment 235, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides.

237. The method of embodiment 235, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides.

238. The method of embodiment 235, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides.

239. The method of embodiment 235, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides.

240. The method of embodiment 235, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide.

241. The method of embodiment 235, wherein the CRISPR repeat comprises the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245.

242. The method of any one of embodiments 235-241, wherein the guide RNA comprises a tracrRNA having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

243. The method of any one of embodiments 235-242, wherein said tracrRNA comprises a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

244. The method of any one of embodiments 170-243, wherein the target polynucleotide sequence is within a cell.

245. The method of embodiment 244, wherein the cell is a eukaryotic cell.

246. The method of embodiment 245, wherein the eukaryotic cell is a mammalian cell.

247. The method of any one of embodiments 244-246, further comprising culturing the cell under conditions in which the RGN polypeptide is expressed and cleaves and modifies the target polynucleotide sequence to produce a nucleic acid molecule comprising a modified target polynucleotide sequence; and selecting a cell comprising said modified target polynucleotide sequence.

248. A cell comprising a modified target polynucleotide sequence produced according to the method of embodiment 247.

249. The cell of embodiment 248, wherein the cell is a prokaryotic cell.

250. The cell of embodiment 248, wherein the cell is a eukaryotic cell.

251. The cell of embodiment 250, wherein the eukaryotic cell is a mammalian cell.

252. The cell of embodiment 251, wherein the mammalian cell is a human cell.

253. The cell of embodiment 252, wherein the human cell is an immune cell.

254. The cell of embodiment 252, wherein the human cell is a stem cell.

255. The cell of embodiment 254, wherein the stem cell is an induced pluripotent stem cell.

256. The cell of embodiment 250, wherein said eukaryotic cell is an insect or avian cell.

257. The cell of embodiment 250, wherein the eukaryotic cell is a fungal cell.

258. The cell of embodiment 250, wherein the eukaryotic cell is a plant cell.

259. A plant or plant part comprising the plant cell of embodiment 258.

260. A pharmaceutical composition comprising the cell of any one of embodiments 49, 157, and 248, and a pharmaceutically acceptable carrier.

261. A method of treating a disease, disorder, or condition, said method comprising administering to a subject in need thereof an effective amount of a pharmaceutical composition of embodiment 169 or 260.

262. The method of embodiment 261, wherein said disease is associated with a causal mutation and said effective amount of said pharmaceutical composition corrects said causal mutation.

263. The method of embodiment 261 or 262, wherein said subject is at risk of developing said disease, disorder, or condition.

264. A method for treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

administering to the subject the nucleic acid molecule of any one of embodiments 1-34, the vector of any one of embodiments 35-48, the RGN polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, the system of any one of embodiments 106-156, the cell of any one of embodiments 49, 157 and 248, or the pharmaceutical composition of embodiment 169 or 260.

265. The method of embodiment 264, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

266. Use of the nucleic acid molecule of any one of embodiments 1-34, the vector of any one of embodiments 35-48, the RGN polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, the system of any one of embodiments 106-156, the cell of any one of embodiments 49, 157 and 248, or the pharmaceutical composition of embodiment 169 or 260 for the treatment of a disease, disorder, or condition in a subject having or at risk of developing said disease, disorder, or condition.

267. The use of embodiment 266, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

268. Use of the nucleic acid molecule of any one of embodiments 1-34, the vector of any one of embodiments 35-48, the RGN polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, the system of any one of embodiments 106-156, the cell of any one of embodiments 49, 157 and 248, or the pharmaceutical composition of embodiment 169 or 260 in the manufacture of a medicament useful for treating a disease, disorder, or condition.

269. The use of embodiment 268, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said medicament corrects said mutation.

270. The nucleic acid molecule of any one of embodiments 1-34, the vector of any one of embodiments 35-48, the RGN polypeptide of any one of embodiments 72-104, the RNP complex of embodiment 105, the system of any one of embodiments 106-156, the cell of any one of embodiments 49, 157 and 248, or the pharmaceutical composition of embodiment 169 or 260 for use in treating a disease, disorder, or condition.

271. The nucleic acid molecule, the vector, the RGN polypeptide, the RNP complex, the system, the cell, or the pharmaceutical composition of embodiment 270, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said nucleic acid molecule, vector, RGN polypeptide, RNP complex, system, cell, or pharmaceutical composition corrects said mutation.

272. A method of increasing efficiency of cleaving and/or modifying a nucleic acid molecule comprising a target sequence, the method comprising delivering the system of any one of embodiments 106-156 or the RNP complex of embodiment 105 to the target sequence or to a cell comprising the target

sequence, wherein cleavage or modification of the nucleic acid molecule occurs at greater efficiency as compared to cleavage or modification of the nucleic acid molecule by a method comprising delivering to the target sequence or to a cell comprising the target sequence a reference RGN system or RNP complex, wherein the reference RGN system or RNP complex does not comprise said RGN polypeptide.

273. The method of embodiment 272, wherein the efficiency of cleaving and/or modifying the target sequence is increased by at least 15%.

274. The method of embodiment 272 or 273, wherein the efficiency of cleaving and/or modifying the target sequence is measured by next generation sequencing, Tracking of Indels by DEcomposition (TIDE) analysis, flow cytometry, or a combination thereof.

275. A polypeptide having an amino acid sequence that differs from SEQ ID NO: 1 at one or more amino acid positions selected from: 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975.

276. The polypeptide of embodiment 275, wherein the polypeptide comprises an amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

277. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

278. The nucleic acid molecule of embodiment 277, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is a positively charged amino acid residue.

279. The nucleic acid molecule of embodiment 277 or 278, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is an R.

280. The nucleic acid molecule of embodiment 279, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

281. The nucleic acid molecule of embodiment 280, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2;

(b) the amino acid sequence set forth as SEQ ID NO: 3;

(c) the amino acid sequence set forth as SEQ ID NO: 4;

(d) the amino acid sequence set forth as SEQ ID NO: 5;

(e) the amino acid sequence set forth as SEQ ID NO: 6;

(f) the amino acid sequence set forth as SEQ ID NO: 7;

(g) the amino acid sequence set forth as SEQ ID NO: 8;

(h) the amino acid sequence set forth as SEQ ID NO: 9;

(i) the amino acid sequence set forth as SEQ ID NO: 10;

(j) the amino acid sequence set forth as SEQ ID NO: 11;

(k) the amino acid sequence set forth as SEQ ID NO: 12;

(l) the amino acid sequence set forth as SEQ ID NO: 13;

(m) the amino acid sequence set forth as SEQ ID NO: 14;

(n) the amino acid sequence set forth as SEQ ID NO: 15; and

(o) the amino acid sequence set forth as SEQ ID NO: 16.

282. A vector comprising the nucleic acid molecule of any one of embodiments 277-281.

283. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

284. The RGN polypeptide of embodiment 283, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is a positively charged amino acid residue.

285. The RGN polypeptide of embodiment 283 or 284, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is an R.

286. The RGN polypeptide of embodiment 285, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

287. The RGN polypeptide of embodiment 286, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2;

(b) the amino acid sequence set forth as SEQ ID NO: 3;

(c) the amino acid sequence set forth as SEQ ID NO: 4;

(d) the amino acid sequence set forth as SEQ ID NO: 5;

(e) the amino acid sequence set forth as SEQ ID NO: 6;

(f) the amino acid sequence set forth as SEQ ID NO: 7;

(g) the amino acid sequence set forth as SEQ ID NO: 8;

(h) the amino acid sequence set forth as SEQ ID NO: 9;

(i) the amino acid sequence set forth as SEQ ID NO: 10;

(j) the amino acid sequence set forth as SEQ ID NO: 11;

(k) the amino acid sequence set forth as SEQ ID NO: 12;

(l) the amino acid sequence set forth as SEQ ID NO: 13;

(m) the amino acid sequence set forth as SEQ ID NO: 14;

(n) the amino acid sequence set forth as SEQ ID NO: 15; and

(o) the amino acid sequence set forth as SEQ ID NO: 16.

288. The RGN polypeptide of any one of embodiments 283-287, wherein said RGN polypeptide is an isolated RGN polypeptide.

289. The RGN polypeptide of any one of embodiments 283-288, wherein said RGN polypeptide is capable of binding a target DNA sequence of a DNA molecule in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target DNA sequence.

290. The RGN polypeptide of embodiment 289, wherein said RGN polypeptide is capable of cleaving said target DNA sequence upon binding.

291. The RGN polypeptide of embodiment 290, wherein cleavage by said RGN polypeptide generates a double-stranded break.

292. The RGN polypeptide of embodiment 290, wherein cleavage by said RGN polypeptide generates a single-stranded break.

293. The RGN polypeptide of any one of embodiments 283-289, wherein said RGN polypeptide is nuclease inactive or a nickase.

294. The RGN polypeptide of embodiment 293, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

295. The RGN polypeptide of embodiment 293 or 294, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

296. The RGN polypeptide of any one of embodiments 293-295, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

297. The RGN polypeptide of any one of embodiments 283-296, wherein said RGN polypeptide is operably fused to a heterologous polypeptide.

298. The RGN polypeptide of embodiment 297, wherein said heterologous polypeptide is operably linked to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

299. The RGN polypeptide of embodiment 297 or 298, wherein said heterologous polypeptide is a polymerase editing polypeptide.

300. The RGN polypeptide of embodiment 299, wherein said polymerase editing polypeptide comprises a DNA polymerase.

301. The RGN polypeptide of embodiment 299, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

302. The RGN polypeptide of embodiment 301, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

303. The RGN polypeptide of embodiment 302, wherein said reverse transcriptase lacks an RNAse H domain.

304. The RGN polypeptide of embodiment 302 or 303, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

305. The RGN polypeptide of any one of embodiments 302-304, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

306. The RGN polypeptide of embodiment 297 or 298, wherein the heterologous polypeptide is a base-editing polypeptide.

307. The RGN polypeptide of embodiment 306, wherein the base-editing polypeptide is a deaminase.

308. The RGN polypeptide of embodiment 307, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

309. The RGN polypeptide of embodiment 307 or 308, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

310. The RGN polypeptide of any one of embodiments 307-309, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

311. The RGN polypeptide of embodiment 297 or 298, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

312. The RGN polypeptide of embodiment 311, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

313. The RGN polypeptide of any one of embodiments 283-312, wherein said target DNA sequence is located adjacent and 5' to a protospacer adjacent motif (PAM).

314. The RGN polypeptide of embodiment 313, wherein the PAM comprises NNGG.

315. The RGN polypeptide of any one of embodiments 283-314, wherein the RGN polypeptide comprises one or more nuclear localization signals.

316. A ribonucleoprotein (RNP) complex comprising the RGN polypeptide of any one of embodiments 283-315 and a guide RNA bound to the RGN polypeptide.

317. A system, said system comprising:

A) one or more guide RNAs (gRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more gRNAs; and

B) an RNA-guided nuclease (RGN) polypeptide, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

318. The system of embodiment 317, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is a positively charged amino acid residue.

319. The system of embodiment 317 or 318, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is an R.

320. The system of embodiment 319, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

321. The system of embodiment 320, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

I the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2;

(b) the amino acid sequence set forth as SEQ ID NO: 3;

(c) the amino acid sequence set forth as SEQ ID NO: 4;

(d) the amino acid sequence set forth as SEQ ID NO: 5;

(e) the amino acid sequence set forth as SEQ ID NO: 6;

(f) the amino acid sequence set forth as SEQ ID NO: 7;

(g) the amino acid sequence set forth as SEQ ID NO: 8;

(h) the amino acid sequence set forth as SEQ ID NO: 9;

(i) the amino acid sequence set forth as SEQ ID NO: 10;

(j) the amino acid sequence set forth as SEQ ID NO: 11;

(k) the amino acid sequence set forth as SEQ ID NO: 12;

(l) the amino acid sequence set forth as SEQ ID NO: 13;

(m) the amino acid sequence set forth as SEQ ID NO: 14;

(n) the amino acid sequence set forth as SEQ ID NO: 15; and

(o) the amino acid sequence set forth as SEQ ID NO: 16.

322. A method for cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule, comprising contacting the nucleic acid molecule with:

a) an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1; and

b) one or more guide RNAs (gRNAs) capable of targeting the RGN of (a) to the target polynucleotide sequencethereby cleaving and/or modifying said target polynucleotide sequence to generate a modified target polynucleotide sequence.

323. The method of embodiment 322, wherein the one or more gRNAs is capable of hybridizing to the target polynucleotide sequence and binding to said RGN polypeptide.

324. The method of embodiment 322 or 323, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is a positively charged amino acid residue.

325. The method of any one of embodiments 322-324, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at amino acid position 778 and/or 969 of said amino acid sequence is an R.

326. The method of embodiment 325, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

327. The method of embodiment 326, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(c) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(h) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(o) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 2;

(b) the amino acid sequence set forth as SEQ ID NO: 3;

(c) the amino acid sequence set forth as SEQ ID NO: 4;

(d) the amino acid sequence set forth as SEQ ID NO: 5;

(e) the amino acid sequence set forth as SEQ ID NO: 6;

(f) the amino acid sequence set forth as SEQ ID NO: 7;

(g) the amino acid sequence set forth as SEQ ID NO: 8;

(h) the amino acid sequence set forth as SEQ ID NO: 9;

(i) the amino acid sequence set forth as SEQ ID NO: 10;

(j) the amino acid sequence set forth as SEQ ID NO: 11;

(k) the amino acid sequence set forth as SEQ ID NO: 12;

(l) the amino acid sequence set forth as SEQ ID NO: 13;

(m) the amino acid sequence set forth as SEQ ID NO: 14;

(n) the amino acid sequence set forth as SEQ ID NO: 15; and

(o) the amino acid sequence set forth as SEQ ID NO: 16.

328. One or more polynucleotides encoding a polymerase editor (PE) comprising a polymerase and an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

329. The one or more polynucleotides of embodiment 328, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

330. The one or more polynucleotides embodiment 328 or 329, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

331. The one or more polynucleotides of embodiment 330, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

332. The one or more polynucleotides of embodiment 331, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

333. The one or more polynucleotides of any one of embodiments 328-332, wherein said polymerase is a DNA polymerase.

334. The one or more polynucleotides of any one of embodiments 328-332, wherein said polymerase is a reverse transcriptase.

335. The one or more polynucleotides of embodiment 334, wherein said reverse transcriptase lacks an RNAse H domain.

336. The one or more polynucleotides of embodiment 334 or 335, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

337. The one or more polynucleotides of any one of embodiments 334-336, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

338. The one or more polynucleotides of any one of embodiments 328-337, wherein the one or more polynucleotides encoding the PE are codon optimized for expression in a eukaryotic cell.

339. The one or more polynucleotides of embodiment 338, wherein the eukaryotic cell is a mammalian cell.

340. The one or more polynucleotides of any one of embodiments 328-339, comprising at least a first and a second polynucleotide, wherein the first polynucleotide comprises a nucleotide sequence encoding said polymerase and the second polynucleotide comprises a nucleotide sequence encoding said RGN polypeptide.

341. The one or more polynucleotides of any one of embodiments 328-339, wherein a nucleotide sequence encoding said polymerase and a nucleotide sequence encoding said RGN polypeptide are comprised within a single polynucleotide.

342. The one or more polynucleotides of embodiment 341, wherein said nucleotide sequence encoding said polymerase and said nucleotide sequence encoding said RGN polypeptide are translated as one polypeptide.

343. The one or more polynucleotides of embodiment 340 or 341, wherein said nucleotide sequence encoding said polymerase and said nucleotide sequence encoding said RGN polypeptide are translated as two separate polypeptides.

344. The one or more polynucleotides of any one of embodiments 328-343, wherein at least one heterologous promoter is operably linked to said one or more polynucleotides.

345. The one or more polynucleotides of embodiment 341 or 342, wherein said polymerase is operably fused to the N-terminus of said RGN polypeptide.

346. The one or more polynucleotides of embodiment 341 or 342, wherein said polymerase is operably fused to the C-terminus of said RGN polypeptide.

347. The one or more polynucleotides of any one of embodiments 328-346, wherein said PE comprises one or more nuclear localization signal (NLS).

348. The one or more polynucleotides of embodiment 347, wherein said one or more NLS is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said polymerase or said RGN polypeptide.

349. The one or more polynucleotides of embodiment 347 or 348, wherein said one or more NLS is selected from the group consisting of SEQ ID NOs: 36, 37, 234, and 235.

350. The one or more polynucleotides of any one of embodiments 328-349, wherein said PE further comprises one or more peptide linker.

351. The one or more polynucleotides of embodiment 350, wherein said one or more peptide linker comprises at least one NLS.

352. The one or more polynucleotides of embodiment 351, wherein said one or more peptide linker comprises two NLSs.

353. The one or more polynucleotides of any one of embodiments 350-352, wherein said one or more peptide linker is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said polymerase or said RGN polypeptide.

354. The one or more polynucleotides of any one of embodiments 350-353, wherein said one or more peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1.

355. The one or more polynucleotides of any one of embodiments 350-354, wherein said one or more peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241).

356. The one or more polynucleotides of any one of embodiments 350-355, wherein said one or more peptide linker has the sequence of any one of SEQ ID NOs: 236-241.

357. The one or more polynucleotides of embodiment 341 or 342, wherein said polymerase is operably fused to an internal location of said RGN polypeptide.

358. The one or more polynucleotides of embodiment 357, wherein said polymerase is operably fused within a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide, or wherein said polymerase is operably fused between a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide and another domain N-terminal or C-terminal to said linker domain 2, said wedge domain, said RuvC domain, said HNH domain, said Rec-2 domain, or said PAM-interacting domain.

359. The one or more polynucleotides of embodiment 358, wherein said RuvC domain is a RuvCIII domain.

360. The one or more polynucleotides of embodiment 359, wherein said polymerase is operably fused within a linker domain 2, a wedge domain, or a RuvCIII domain of said RGN polypeptide.

361. The one or more polynucleotides of embodiment 357, wherein said polymerase is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

362. The one or more polynucleotides of any one of embodiments 328-361, wherein said RGN polypeptide is capable of binding a target sequence in a target polynucleotide in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA), wherein said target sequence comprises a target strand and a non-target strand, and wherein said gRNA is capable of hybridizing to the target strand of the target sequence.

363. The one or more polynucleotides of embodiment 362, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target sequence.

364. The one or more polynucleotides of embodiment 363, wherein the PAM comprises a consensus nucleotide sequence set forth as NNGG.

365. The one or more polynucleotides of any one of embodiments 362-364, wherein said RGN polypeptide is capable of cleaving said target polynucleotide upon binding.

366. The one or more polynucleotides of embodiment 365, wherein said RGN polypeptide is capable of generating a double-stranded break.

367. The one or more polynucleotides of embodiment 365, wherein said RGN polypeptide is capable of generating a single-stranded break.

368. The one or more polynucleotides of any one of embodiments 328-364, wherein said RGN polypeptide comprises an HNH domain with at least one mutation that reduces or eliminates its nuclease activity.

369. The one or more polynucleotides of any one of embodiments 328-364, wherein said RGN polypeptide comprises an HNH domain with at least two mutations that reduces or eliminates its nuclease activity.

370. The one or more polynucleotides of any one of embodiments 328-364, wherein said RGN polypeptide does not comprise an HNH domain.

371. The one or more polynucleotides of embodiment 370, wherein said HNH domain of said RGN polypeptide has been replaced with said polymerase.

372. The one or more polynucleotides of any one of embodiments 328-364, wherein said RGN polypeptide is nuclease inactive.

373. The one or more polynucleotides of any one of embodiments 328-364, wherein said RGN polypeptide comprises an RGN nickase.

374. The one or more polynucleotides of embodiment 373, wherein said RGN nickase comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

375. The one or more polynucleotides of embodiment 374, wherein said RGN nickase comprises an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

376. The one or more polynucleotides of embodiment 374 or 375, wherein said RGN nickase comprises the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

377. The one or more polynucleotides of any one of embodiments 328-376, wherein said PE comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 135, and 162-181.

378. The one or more polynucleotides of any one of embodiments 328-376, wherein said PE has an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 135, and 162-181.

379. The one or more polynucleotides of any one of embodiments 328-378, wherein at least one of said one or more polynucleotides is an RNA polynucleotide.

380. The one or more polynucleotides of embodiment 379, wherein said RNA polynucleotide is an mRNA.

381. The one or more polynucleotides of embodiment 379, wherein said RNA polynucleotide is a circRNA.

382. One or more vectors comprising the one or more polynucleotides of any one of embodiments 328-378.

383. The one or more vectors of embodiment 382, wherein said one or more vectors further comprise at least one nucleotide sequence encoding a polymerase editing guide RNA (PEgRNA), wherein said PEgRNA comprises an extension arm, wherein said extension arm comprises a primer binding site and a DNA synthesis template sequence, and wherein said PEgRNA is capable of binding to said RGN polypeptide of said PE.

384. The one or more vectors of embodiment 383, wherein said extension arm is at the 3' end of said PEgRNA.

385. The one or more vectors of embodiment 383 or 384, wherein said DNA synthesis template sequence is 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 34, 37, 38, 39, 40, 42, or 46 nucleotides in length.

386. The one or more vectors of any one of embodiments 383-385, wherein said DNA synthesis template sequence comprises an RT template (RTT) sequence.

387. The one or more vectors of any one of embodiments 383-386, wherein said primer binding site is 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 nucleotides in length.

388. The one or more vectors of any one of embodiments 383-387, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence of any one of SEQ ID NOs: 33, 244, and 245, or that differs from any one of SEQ ID NOs: 33, 244, and 245 by 1 to 5 nucleotides.

389. The one or more vectors of embodiment 388, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence that differs from any one of SEQ ID NOs: 33, 244, and 245 by 5 nucleotides, by 4 nucleotides, by 3 nucleotides, by 2 nucleotides, or by 1 nucleotide.

390. The one or more vectors of embodiment 388, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising the nucleotide sequence set forth as any one of SEQ ID NOs: 33, 244, and 245.

391. The one or more vectors of any one of embodiments 383-390, wherein the PEgRNA comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

392. The one or more vectors of any one of embodiments 383-391, wherein the PEgRNA comprises a tracrRNA comprising a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

393. The one or more vectors of any one of embodiments 382-392, wherein said one or more vectors further comprise a nucleotide sequence encoding a nicking guide RNA.

394. The one or more vectors of any one of embodiments 382-393, wherein said one or more vectors further comprise a nucleotide sequence encoding a dominant negative MLH1.

395. The one or more vectors of any one of embodiments 382-384, wherein the one or more vectors are adeno-associated viral (AAV) vectors.

396. A cell comprising the one or more polynucleotides of any one of embodiments 328-381 or the one or more vectors of any one of embodiments 382-395.

397. The cell of embodiment 396, wherein the cell is a prokaryotic cell.

398. The cell of embodiment 396, wherein the cell is a eukaryotic cell.

399. The cell of embodiment 398, wherein the eukaryotic cell is a mammalian cell.

400. The cell of embodiment 399, wherein the mammalian cell is a human cell.

401. The cell of embodiment 400, wherein the human cell is an immune cell.

402. The cell of embodiment 400, wherein the human cell is a stem cell.

403. The cell of embodiment 402, wherein the stem cell is an induced pluripotent stem cell.

404. The cell of embodiment 398, wherein the eukaryotic cell is an insect or avian cell.

405. The cell of embodiment 398, wherein the eukaryotic cell is a fungal cell.

406. The cell of embodiment 398, wherein the eukaryotic cell is a plant cell.

407. A plant or plant part comprising the plant cell of embodiment 406.

408. A method for making a PE comprising culturing the cell of embodiment 396 under conditions in which the PE is expressed.

409. The method of embodiment 408, further comprising purifying said PE.

410. The method of embodiment 408, wherein said cell further expresses one or more guide RNAs capable of binding to said RGN polypeptide of said PE to form a ribonucleoprotein complex.

411. The method of embodiment 410, further comprising purifying said ribonucleoprotein complex.

412. A polymerase editor (PE) comprising a polymerase and an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

413. The PE of embodiment 412, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

414. The PE of embodiment 412 or 413, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

415. The PE of embodiment 414, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

416. The PE of embodiment 415, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

417. The PE of any one of embodiments 412-416, wherein said polymerase is a DNA polymerase.

418. The PE of any one of embodiments 412-416, wherein said polymerase is a reverse transcriptase.

419. The PE of embodiment 418, wherein said reverse transcriptase lacks an RNAse H domain.

420. The PE of embodiment 418 or 419, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

421. The PE of any one of embodiments 418-420, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

422. The PE of any one of embodiments 418-421, wherein said polymerase and said RGN polypeptide are translated as two separate polypeptides.

423. The PE of any one of embodiments 418-421, wherein said polymerase and said RGN polypeptide are translated as one polypeptide.

424. The PE of embodiment 423, wherein said polymerase is operably fused to the N-terminus of said RGN polypeptide.

425. The PE of embodiment 423, wherein said polymerase is operably fused to the C-terminus of said RGN polypeptide.

426. The PE of any one of embodiments 412-425, wherein said PE comprises one or more nuclear localization signal (NLS).

427. The PE of embodiment 426, wherein said one or more NLS is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said polymerase or said RGN polypeptide.

428. The PE of embodiment 426 or 427, wherein said one or more NLS is selected from the group consisting of SEQ ID NOs: 36, 37, 234, and 235.

429. The PE of any one of embodiments 412-428, wherein said PE further comprises one or more peptide linker.

430. The PE of embodiment 429, wherein said one or more peptide linker comprises at least one NLS.

431. The PE of embodiment 430, wherein said one or more peptide linker comprises two NLSs.

432. The PE of any one of embodiments 429-431, wherein said one or more peptide linker is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said polymerase or said RGN polypeptide.

433. The PE of any one of embodiments 429-432, wherein said one or more peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1.

434. The PE of any one of embodiments 429-433, wherein said one or more peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241).

435. The PE of any one of embodiments 429-434, wherein said one or more peptide linker has the sequence of any one of SEQ ID NOs: 236-241.

436. The PE of embodiment 423, wherein said polymerase is operably fused to an internal location of said RGN polypeptide.

437. The PE of embodiment 436, wherein said polymerase is operably fused within a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide, or wherein said polymerase is operably fused between a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said

RGN polypeptide and another domain N-terminal or C-terminal to said linker domain 2, said wedge domain, said RuvC domain, said HNH domain, said Rec-2 domain, or said PAM-interacting domain.

438. The PE of embodiment 437, wherein said RuvC domain is a RuvCIII domain.

439. The PE of embodiment 438, wherein said polymerase is operably fused within a linker domain 2, a wedge domain, or a RuvCIII domain of said RGN polypeptide.

440. The PE of embodiment 436, wherein said polymerase is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

441. The PE of any one of embodiments 412-440, wherein said RGN polypeptide is capable of binding a target sequence in a target polynucleotide in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA), wherein said target sequence comprises a target strand and a non-target strand, and wherein said gRNA is capable of hybridizing to the target strand of the target sequence.

442. The PE of embodiment 441, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target sequence.

443. The PE of embodiment 442, wherein the PAM comprises a consensus nucleotide sequence set forth as NNGG.

444. The PE of any one of embodiments 441-443, wherein said RGN polypeptide is capable of cleaving said target polynucleotide upon binding.

445. The PE of embodiment 444, wherein said RGN polypeptide is capable of generating a double-stranded break.

446. The PE of embodiment 444, wherein said RGN polypeptide is capable of generating a single-stranded break.

447. The PE of any one of embodiments 412-443, wherein said RGN polypeptide comprises an HNH domain with at least one mutation that reduces or eliminates its nuclease activity.

448. The PE of any one of embodiments 412-443, wherein said RGN polypeptide comprises an HNH domain with at least two mutations that reduces or eliminates its nuclease activity.

449. The PE of any one of embodiments 412-443, wherein said RGN polypeptide does not comprise an HNH domain.

450. The PE of embodiment 449, wherein said HNH domain of said RGN polypeptide has been replaced with said polymerase.

451. The PE of any one of embodiments 412-443, wherein said RGN polypeptide is nuclease inactive.

452. The PE of any one of embodiments 412-443, wherein said RGN polypeptide comprises an RGN nickase.

453. The PE of embodiment 452, wherein said RGN nickase comprises a D16A and/or a H611A mutation(s).

454. The PE of embodiment 452 or 453, wherein said RGN nickase comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

455. The PE of embodiment 454, wherein said RGN nickase comprises an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

456. The PE of embodiment 454 or 455, wherein said RGN nickase comprises the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

457. The PE of any one of embodiments 412-456, wherein said PE comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 135, and 162-181.

458. The PE of any one of embodiments 412-457, wherein said PE has an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 135, and 162-181.

459. A ribonucleoprotein (RNP) complex comprising the PE of any one of embodiments 412-458 and a PEgRNA bound to the RGN polypeptide, wherein said PEgRNA comprises a primer binding site (PBS) and a DNA synthesis template.

460. The RNP complex of embodiment 459, wherein the PEgRNA comprises a spacer that hybridizes to a eukaryotic target sequence.

461. The RNP complex of embodiment 460, wherein the eukaryotic target sequence comprises a mammalian target sequence.

462. A PE system for modifying one or more target sequences in a target polynucleotide, said system comprising:

a) one or more polymerase editing guide RNAs (PEgRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more PEgRNAs, wherein the one or more PEgRNAs comprise an extension arm, wherein the extension arm comprises a primer binding site (PBS) and a DNA synthesis template sequence; and

b) a PE of any one of embodiments 412-458;

wherein said one or more PEgRNAs are capable of binding to said RGN polypeptide of said PE.

463. The PE system of embodiment 462, wherein each of the one or more PEgRNAs is capable of hybridizing to the target strand of said target sequence and forming a complex with the RGN polypeptide to direct said RGN polypeptide to bind to said target sequence.

464. The PE system of embodiment 462 or 463, wherein the spacer hybridizes to a eukaryotic target sequence.

465. The PE system of embodiment 464, wherein the eukaryotic target sequence comprises a mammalian target sequence.

466. The PE system of any one of embodiments 462-465, wherein said extension arm is at the 3' end of said PEgRNA.

467. The PE system of any one of embodiments 462-466, wherein said DNA synthesis template sequence is 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 34, 37, 38, 39, 40, 42, or 46 nucleotides in length.

468. The PE system of any one of embodiments 462-467, wherein said DNA synthesis template sequence comprises an RT template sequence (RTT).

469. The PE system of any one of embodiments 462-468, wherein said primer binding site is 9, 10, 11, 12, 13, 14, 15, 16, 17, or 18 nucleotides in length.

470. The PE system of any one of embodiments 462-469, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence of any one of SEQ ID NOs: 33, 244, and 245, or that differs from any one of SEQ ID NOs: 33, 244, and 245 by 1 to 5 nucleotides.

471. The PE system of embodiment 470, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence that differs from any one of SEQ ID NOs: 33, 244, and 245 by 5 nucleotides, by 4 nucleotides, by 3 nucleotides, by 2 nucleotides, or by 1 nucleotide.

472. The PE system of embodiment 470, wherein the PEgRNA comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising the nucleotide sequence set forth as any one of SEQ ID NOs: 33, 244, and 245.

473. The PE system of any one of embodiments 462-472, wherein the PEgRNA comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

474. The PE system of any one of embodiments 462-473, wherein the PEgRNA comprises a tracrRNA comprising a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

475. The PE system of any one of embodiments 462-474, wherein said polynucleotide encoding said one or more PEgRNAs and said polynucleotide encoding said PE are on a single vector.

476. The PE system of any one of embodiments 462-474, wherein said polynucleotide encoding said one or more PEgRNAs and said polynucleotide encoding said PE are on separate vectors.

477. The PE system of any one of embodiments 462-474, wherein said polynucleotide encoding said polymerase and said polynucleotide encoding said RGN polypeptide are on separate vectors.

478. The PE system of any one of embodiments 462-474, wherein said polynucleotide encoding said one or more PEgRNAs, said polynucleotide encoding said polymerase, and said polynucleotide encoding said RGN polypeptide are all on separate vectors.

479. The PE system of any one of embodiments 462-474, wherein said polynucleotide encoding said PE is an RNA polynucleotide.

480. The PE system of embodiment 479, wherein said RNA polynucleotide is an mRNA.

481. The PE system of embodiment 479, wherein said RNA polynucleotide is a circRNA.

482. The PE system of any one of embodiments 462-481, wherein said PE system further comprises a nicking guide RNA or a polynucleotide comprising a nucleotide sequence encoding a nicking guide RNA.

483. The PE system of any one of embodiments 462-482, wherein said PE system further comprises a dominant negative MLH1 or a polynucleotide comprising a nucleotide sequence encoding a dominant negative MLH1.

484. A cell comprising the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, or the PE system of any one of embodiments 462-483.

485. The cell of embodiment 484, wherein the cell is a prokaryotic cell.

486. The cell of embodiment 484, wherein the cell is a eukaryotic cell.

487. The cell of embodiment 486, wherein the eukaryotic cell is a mammalian cell.

488. The cell of embodiment 487, wherein the mammalian cell is a human cell.

489. The cell of embodiment 488, wherein the human cell is an immune cell.

490. The cell of embodiment 488, wherein the human cell is a stem cell.

491. The cell of embodiment 490, wherein the stem cell is an induced pluripotent stem cell.

492. The cell of embodiment 486, wherein the eukaryotic cell is an insect or avian cell.

493. The cell of embodiment 486, wherein the eukaryotic cell is a fungal cell.

494. The cell of embodiment 486, wherein the eukaryotic cell is a plant cell.

495. A plant or plant part comprising the plant cell of embodiment 494.

496. A pharmaceutical composition comprising the one or more polynucleotides of any one of embodiments 328-381, the one or more vectors of any one of embodiments 382-395, the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, the cell of any one of embodiments 398-403 and 486-491, or the PE system of any one of embodiments 462-483, and a pharmaceutically acceptable carrier.

497. A method for modifying a target polynucleotide comprising a target sequence, said method comprising delivering a PE system according to any one of embodiments 462-483 to said target sequence or a cell comprising the target sequence, wherein said method generates a modified target polynucleotide, and wherein components of said PE system are delivered simultaneously or sequentially to said target sequence or said cell comprising the target sequence.

498. The method of embodiment 497, wherein said modified target polynucleotide comprises insertion of heterologous DNA into the target polynucleotide.

499. The method of embodiment 497, wherein said modified target polynucleotide comprises deletion of at least one nucleotide from the target polynucleotide.

500. The method of embodiment 497, wherein said modified target polynucleotide comprises mutation of at least one nucleotide in the target polynucleotide.

501. A method for modifying one or more target sequences in a target polynucleotide, said method comprising:

a) assembling a ribonucleoprotein (RNP) complex by combining:

i) a PEgRNA comprising an extension arm, wherein said extension arm comprises a primer binding site (PBS) and a DNA synthesis template; and

ii) a PE of any one of embodiments 412-458;

under conditions suitable for formation of the RNP complex; and

b) contacting said target polynucleotide or a cell comprising said target polynucleotide with the assembled RNP complex;

thereby modifying said one or more target sequences in a target polynucleotide.

502. The method of any one of embodiments 497-501, wherein said method is performed *in vitro*, *in vivo*, or *ex vivo*.

503. The method of any one of embodiments 497-502, wherein said target polynucleotide is within a cell.

504. The method of embodiment 503, wherein the cell is a eukaryotic cell.

505. The method of embodiment 504, wherein the eukaryotic cell is a mammalian cell.

506. A cell produced according to the method of embodiment 503, wherein said target polynucleotide has been modified at said target sequence.

507. The cell of embodiment 506, wherein the cell is a prokaryotic cell.

508. The cell of embodiment 506, wherein the cell is a eukaryotic cell.

509. The cell of embodiment 508, wherein the eukaryotic cell is a mammalian cell.

510. The cell of embodiment 509, wherein the mammalian cell is a human cell.

511. The cell of embodiment 510, wherein the human cell is an immune cell.

512. The cell of embodiment 510, wherein the human cell is a stem cell.

513. The cell of embodiment 512, wherein the stem cell is an induced pluripotent stem cell.

514. The cell of embodiment 508, wherein the eukaryotic cell is an insect or avian cell.

515. The cell of embodiment 508, wherein the eukaryotic cell is a fungal cell.

516. The cell of embodimen508, wherein the eukaryotic cell is a plant cell.

517. A plant or plant part comprising the plant cell of embodiment 516.

518. A pharmaceutical composition comprising the cell of any one of embodiments 508-513 and a pharmaceutically acceptable carrier.

519. A method for treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

administering to the subject the one or more polynucleotides of any one of embodiments 328-381, the one or more vectors of any one of embodiments 382-395, the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, the PE system of any one of embodiments 462-483, the cell of any one of embodiments 398-403, 486-491, and 508-513, or the pharmaceutical composition of embodiment 496 or 518.

520. The method of embodiment 519, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

521. Use of the one or more polynucleotides of any one of embodiments 328-381, the one or more vectors of any one of embodiments 382-395, the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, the PE system of any one of embodiments 462-483, the cell of any one of embodiments 398-403, 486-491, and 508-513, or the pharmaceutical composition of embodiment 496 or 518 for the treatment of a disease, disorder, or condition in a subject having or at risk of developing said disease, disorder, or condition.

522. The use of embodiment 521, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

523. Use of the one or more polynucleotides of any one of embodiments 328-381, the one or more vectors of any one of embodiments 382-395, the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, the PE system of any one of embodiments 462-483, the cell of any one of embodiments 398-403, 486-491, and 508-513, or the pharmaceutical composition of embodiment 496 or 518 in the manufacture of a medicament useful for treating a disease, disorder, or condition.

524. The use of embodiment 523, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said medicament corrects said mutation.

525. The one or more polynucleotides of any one of embodiments 328-381, the one or more vectors of any one of embodiments 382-395, the PE of any one of embodiments 412-458, the RNP complex of any one of embodiments 459-461, the PE system of any one of embodiments 462-483, the cell of any one of embodiments 398-403, 486-491, and 508-513, or the pharmaceutical composition of embodiment 496 or 518 for use in treating a disease, disorder, or condition.

526. The one or more polynucleotides, the one or more vectors, the PE, the RNP complex, the PE system, the cell, or the pharmaceutical composition of embodiment 525, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said one or more polynucleotides, one or more vectors, PE, RNP complex, PE system, cell, or pharmaceutical composition corrects said mutation.

527. A fusion polypeptide comprising:

(a) an RGN polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533,

541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1; and

(b) a heterologous polypeptide.

528. The fusion polypeptide of embodiment 527, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

529. The fusion polypeptide of embodiment 527 or 528, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

530. The fusion polypeptide of embodiment 529, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

531. The fusion polypeptide of embodiment 530, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

532. The fusion polypeptide of any one of embodiments 527-531, wherein said RGN polypeptide is an isolated RGN polypeptide.

533. The fusion polypeptide of any one of embodiments 527-532, wherein said RGN polypeptide is capable of binding a target polynucleotide sequence of a DNA molecule in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target polynucleotide sequence.

534. The fusion polypeptide of embodiment 533, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target polynucleotide sequence.

535. The fusion polypeptide of embodiment 534, wherein said RGN polypeptide recognizes a PAM having a consensus nucleotide sequence set forth as NNGG.

536. The fusion polypeptide of any one of embodiments 527-535, wherein the RGN polypeptide comprises a PAM-interacting domain comprising the amino acid sequence set forth as SEQ ID NO: 253.

537. The fusion polypeptide of any one of embodiments 527-536, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

538. The fusion polypeptide of any one of embodiments 527-537, wherein the RGN polypeptide comprises one or more nuclear localization signals (NLS).

539. The fusion polypeptide of embodiment 538, wherein said one or more NLS is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said RGN polypeptide or said heterologous polypeptide.

540. The fusion polypeptide of embodiment 538 or 539, wherein said one or more NLS is selected from the group consisting of SEQ ID NOs: 36, 37, 234, and 235.

541. The fusion polypeptide of any one of embodiments 527-540, wherein said fusion polypeptide further comprises one or more peptide linker.

542. The fusion polypeptide of embodiment 541, wherein said one or more peptide linker comprises at least one NLS.

543. The fusion polypeptide of embodiment 541, wherein said one or more peptide linker comprises two NLSs.

544. The fusion polypeptide of any one of embodiments 541-543, wherein said one or more peptide linker is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said RGN polypeptide or said heterologous polypeptide.

545. The fusion polypeptide of any one of embodiments 541-544, wherein said one or more peptide linker has a formula of $–(SGGS)_x\text{-}NLS_m\text{-}(SGGS)_y\text{-}NLS_n\text{-}(SGGS)_z$-, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1.

546. The fusion polypeptide of any one of embodiments 541-545, wherein said one or more peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241).

547. The fusion polypeptide of any one of embodiments 541-546, wherein said one or more peptide linker has the sequence of any one of SEQ ID NOs: 236-241.

548. The fusion polypeptide of any one of embodiments 527-547, wherein said RGN polypeptide is nuclease inactive or a nickase.

549. The fusion polypeptide of embodiment 548, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

550. The fusion polypeptide of embodiment 548 or 549, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

551. The fusion polypeptide of any one of embodiments 548-550, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

552. The fusion polypeptide of any one of embodiments 527-547, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence upon binding.

553. The fusion polypeptide of embodiment 552, wherein cleavage by said RGN polypeptide generates a double-stranded break.

554. The fusion polypeptide of embodiment 552, wherein cleavage by said RGN polypeptide generates a single-stranded break.

555. The fusion polypeptide of any one of embodiments 527-554, wherein said heterologous polypeptide is a polymerase editing polypeptide.

556. The fusion polypeptide of embodiment 555, wherein said polymerase editing polypeptide comprises a DNA polymerase.

557. The fusion polypeptide of embodiment 555, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

558. The fusion polypeptide of embodiment 557, wherein said reverse transcriptase lacks an RNAse H domain.

559. The fusion polypeptide of embodiment 557 or 558, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

560. The fusion polypeptide of any one of embodiments 557-559, wherein said reverse transcriptase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

561. The fusion polypeptide of any one of embodiments 527-554, wherein the heterologous polypeptide is a base-editing polypeptide.

562. The fusion polypeptide of embodiment 561, wherein the base-editing polypeptide is a deaminase.

563. The fusion polypeptide of embodiment 562, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

564. The fusion polypeptide of embodiment 562 or 563, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

565. The fusion polypeptide of any one of embodiments 562-564, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

566. The fusion polypeptide of any one of embodiments 527-554, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

567. The fusion polypeptide of embodiment 566, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

568. A polynucleotide comprising a nucleotide sequence encoding the fusion polypeptide of any one of embodiments 527-567.

569. The polynucleotide of embodiment 568, wherein the nucleotide sequence encoding the fusion polypeptide is codon optimized for expression in a eukaryotic cell.

570. The polynucleotide of embodiment 569, wherein the eukaryotic cell is a mammalian cell.

571. The polynucleotide of any one of embodiments 568-570, wherein the nucleotide sequence encoding the fusion polypeptide is operably linked to a promoter.

572. A cell comprising the fusion polypeptide of any one of embodiments 527-567 or the polynucleotide of any one of embodiments 568-571.

573. The cell of embodiment 572, wherein the cell is a prokaryotic cell.

574. The cell of embodiment 572, wherein the cell is a eukaryotic cell.

575. The cell of embodiment 574, wherein the eukaryotic cell is a mammalian cell.

576. The cell of embodiment 575, wherein the mammalian cell is a human cell.

577. The cell of embodiment 576, wherein the human cell is an immune cell.

578. The cell of embodiment 576, wherein the human cell is a stem cell.

579. The cell of embodiment 578, wherein the stem cell is an induced pluripotent stem cell.

580. The cell of embodiment 574, wherein the eukaryotic cell is an insect or avian cell.

581. The cell of embodiment 574, wherein the eukaryotic cell is a fungal cell.

582. The cell of embodiment 574, wherein the eukaryotic cell is a plant cell.

583. A plant or plant part comprising the plant cell of embodiment 582.

584. One or more polynucleotides encoding a base editor comprising a base editing polypeptide and an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

585. The one or more polynucleotides of embodiment 584, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745,

774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

586. The one or more polynucleotides embodiment 584 or 585, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

587. The one or more polynucleotides of embodiment 586, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

588. The one or more polynucleotides of embodiment 587, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

589. The one or more polynucleotides of any one of embodiments 584-588, wherein said base editing polypeptide is a deaminase.

590. The one or more polynucleotides of embodiment 589, wherein said deaminase is a cytosine deaminase or an adenine deaminase.

591. The one or more polynucleotides of embodiment 589 or 590, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

592. The one or more polynucleotides of any one of embodiments 589-591, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

593. The one or more polynucleotides of any one of embodiments 584-592, wherein the one or more polynucleotides encoding the base editor are codon optimized for expression in a eukaryotic cell.

594. The one or more polynucleotides of embodiment 593, wherein the eukaryotic cell is a mammalian cell.

595. The one or more polynucleotides of any one of embodiments 584-594, comprising at least a first and a second polynucleotide, wherein the first polynucleotide comprises a nucleotide sequence encoding said base editing polypeptide and the second polynucleotide comprises a nucleotide sequence encoding said RGN polypeptide.

596. The one or more polynucleotides of any one of embodiments 584-594, wherein a nucleotide sequence encoding said base editing polypeptide and a nucleotide sequence encoding said RGN polypeptide are comprised within a single polynucleotide.

597. The one or more polynucleotides of embodiment 596, wherein said nucleotide sequence encoding said base editing polypeptide and said nucleotide sequence encoding said RGN polypeptide are translated as one polypeptide.

598. The one or more polynucleotides of embodiment 595 or 596, wherein said nucleotide sequence encoding said base editing polypeptide and said nucleotide sequence encoding said RGN polypeptide are translated as two separate polypeptides.

599. The one or more polynucleotides of any one of embodiments 584-598, wherein at least one heterologous promoter is operably linked to said one or more polynucleotides.

600. The one or more polynucleotides of embodiment 596 or 597, wherein said base editing polypeptide is operably fused to the N-terminus of said RGN polypeptide.

601. The one or more polynucleotides of embodiment 596 or 597, wherein said base editing polypeptide is operably fused to the C-terminus of said RGN polypeptide.

602. The one or more polynucleotides of any one of embodiments 584-601, wherein said base editor comprises one or more nuclear localization signal (NLS).

603. The one or more polynucleotides of embodiment 602, wherein said one or more NLS is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said base editing polypeptide or said RGN polypeptide.

604. The one or more polynucleotides of embodiment 602 or 603, wherein said one or more NLS is selected from the group consisting of SEQ ID NOs: 36, 37, 234, and 235.

605. The one or more polynucleotides of any one of embodiments 584-604, wherein said base editor further comprises one or more peptide linker.

606. The one or more polynucleotides of embodiment 605, wherein said one or more peptide linker comprises at least one NLS.

607. The one or more polynucleotides of embodiment 606, wherein said one or more peptide linker comprises two NLSs.

608. The one or more polynucleotides of any one of embodiments 605-607, wherein said one or more peptide linker is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said base editing polypeptide or said RGN polypeptide.

609. The one or more polynucleotides of any one of embodiments 605-608, wherein said one or more peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1.

610. The one or more polynucleotides of any one of embodiments 605-609, wherein said one or more peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241).

611. The one or more polynucleotides of any one of embodiments 605-610, wherein said one or more peptide linker has the sequence of any one of SEQ ID NOs: 236-241.

612. The one or more polynucleotides of embodiment 596 or 597, wherein said base editing polypeptide is operably fused to an internal location of said RGN polypeptide.

613. The one or more polynucleotides of embodiment 612, wherein said base editing polypeptide is operably fused within a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide, or wherein said base editing polypeptide is operably fused between a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide and another domain N-terminal or C-terminal to said linker domain 2, said wedge domain, said RuvC domain, said HNH domain, said Rec-2 domain, or said PAM-interacting domain.

614. The one or more polynucleotides of embodiment 613, wherein said RuvC domain is a RuvCIII domain.

615. The one or more polynucleotides of embodiment 614, wherein said base editing polypeptide is operably fused within a linker domain 2, a wedge domain, or a RuvCIII domain of said RGN polypeptide.

616. The one or more polynucleotides of embodiment 612, wherein said base editing polypeptide is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

617. The one or more polynucleotides of any one of embodiments 584-616, wherein said RGN polypeptide is capable of binding a target sequence in a target polynucleotide in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA), wherein said target sequence comprises a target strand and a non-target strand, and wherein said gRNA is capable of hybridizing to the target strand of the target sequence.

618. The one or more polynucleotides of embodiment 617, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target sequence.

619. The one or more polynucleotides of embodiment 618, wherein the PAM comprises a consensus nucleotide sequence set forth as NNGG.

620. The one or more polynucleotides of any one of embodiments 617-619, wherein said RGN polypeptide is capable of cleaving said target polynucleotide upon binding.

621. The one or more polynucleotides of embodiment 620, wherein said RGN polypeptide is capable of generating a double-stranded break.

622. The one or more polynucleotides of embodiment 620, wherein said RGN polypeptide is capable of generating a single-stranded break.

623. The one or more polynucleotides of any one of embodiments 584-619, wherein said RGN polypeptide comprises an HNH domain with at least one mutation that reduces or eliminates its nuclease activity.

624. The one or more polynucleotides of any one of embodiments 584-619, wherein said RGN polypeptide comprises an HNH domain with at least two mutations that reduces or eliminates its nuclease activity.

625. The one or more polynucleotides of any one of embodiments 584-619, wherein said RGN polypeptide does not comprise an HNH domain.

626. The one or more polynucleotides of embodiment 625, wherein said HNH domain of said RGN polypeptide has been replaced with said base editing polypeptide.

627. The one or more polynucleotides of any one of embodiments 584-619, wherein said RGN polypeptide is nuclease inactive.

628. The one or more polynucleotides of any one of embodiments 584-619, wherein said RGN polypeptide comprises an RGN nickase.

629. The one or more polynucleotides of embodiment 628, wherein said RGN nickase comprises a D16A and/or a H611A mutation(s).

630. The one or more polynucleotides of embodiment 628 or 629, wherein said RGN nickase comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

631. The one or more polynucleotides of embodiment 630, wherein said RGN nickase comprises an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

632. The one or more polynucleotides of embodiment 630 or 631, wherein said RGN nickase comprises the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

633. The one or more polynucleotides of any one of embodiments 584-632, wherein said base editor comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 258-270.

634. The one or more polynucleotides of any one of embodiments 584-633, wherein said base editor has an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 258-270.

635. The one or more polynucleotides of any one of embodiments 584-634, wherein at least one of said one or more polynucleotides is an RNA polynucleotide.

636. The one or more polynucleotides of embodiment 635, wherein said RNA polynucleotide is an mRNA.

637. The one or more polynucleotides of embodiment 635, wherein said RNA polynucleotide is a circRNA.

638. One or more vectors comprising the one or more polynucleotides of any one of embodiments 584-634.

639. The one or more vectors of embodiment 638, wherein said one or more vectors further comprise at least one nucleotide sequence encoding a guide RNA.

640. The one or more vectors of embodiment 639, wherein the guide RNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

641. The one or more vectors of embodiment 640, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 5 nucleotides.

642. The one or more vectors of embodiment 640, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 4 nucleotides.

643. The one or more vectors of embodiment 640, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 3 nucleotides.

644. The one or more vectors of embodiment 640, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 2 nucleotides.

645. The one or more vectors of embodiment 640, wherein the CRISPR repeat differs from SEQ ID NO: 33, 244, or 245 by 1 nucleotide.

646. The one or more vectors of embodiment 640, wherein the CRISPR repeat comprises the nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245.

647. The one or more vectors of any one of embodiments 640-646, wherein the guide RNA comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

648. The one or more vectors of embodiment 647, wherein the tracrRNA comprises a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, or 100% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

649. The one or more vectors of any one of embodiments 639-648, where said guide RNA is a single guide RNA.

650. The one or more vectors of any one of embodiments 639-649, wherein the one or more vectors are adeno-associated viral (AAV) vectors.

651. A cell comprising the one or more polynucleotides of any one of embodiments 584-637 or the one or more vectors of any one of embodiments 638-650.

652. The cell of embodiment 651, wherein the cell is a prokaryotic cell.

653. The cell of embodiment 651, wherein the cell is a eukaryotic cell.

654. The cell of embodiment 653, wherein the eukaryotic cell is a mammalian cell.

655. The cell of embodiment 654, wherein the mammalian cell is a human cell.

656. The cell of embodiment 655, wherein the human cell is an immune cell.

657. The cell of embodiment 655, wherein the human cell is a stem cell.

658. The cell of embodiment 657, wherein the stem cell is an induced pluripotent stem cell.

659. The cell of embodiment 653, wherein the eukaryotic cell is an insect or avian cell.

660. The cell of embodiment 653, wherein the eukaryotic cell is a fungal cell.

661. The cell of embodiment 653, wherein the eukaryotic cell is a plant cell.

662. A plant or plant part comprising the plant cell of embodiment 661.

663. A method for making a base editor comprising culturing the cell of embodiment 651 under conditions in which the base editor is expressed.

664. The method of embodiment 663, further comprising purifying said base editor.

665. The method of embodiment 663, wherein said cell further expresses one or more guide RNAs capable of binding to said RGN polypeptide of said base editor to form a ribonucleoprotein complex.

666. The method of embodiment 665, further comprising purifying said ribonucleoprotein complex.

667. A base editor comprising a base editing polypeptide and an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

668. The base editor of embodiment 667, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

669. The base editor of embodiment 667 or 668, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780,

795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

670. The base editor of embodiment 669, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

671. The base editor of embodiment 670, wherein said RGN polypeptide comprises an amino acid sequence selected from:

(I) an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R,

or

(II) an amino acid sequence having 100% sequence identity to the amino acid sequence selected from the group consisting of:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

672. The base editor of any one of embodiments 667-671, wherein said base editing polypeptide is a deaminase.

673. The base editor of embodiment 672, wherein said deaminase is a cytosine deaminase or an adenine deaminase.

674. The base editor of embodiment 672 or 673, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

675. The base editor of any one of embodiments 672-674, wherein the deaminase has at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

676. The base editor of any one of embodiments 667-675, wherein said base editing polypeptide and said RGN polypeptide are translated as two separate polypeptides.

677. The base editor of any one of embodiments 667-675, wherein said base editing polypeptide and said RGN polypeptide are translated as one polypeptide.

678. The base editor of embodiment 677, wherein said base editing polypeptide is operably fused to the N-terminus of said RGN polypeptide.

679. The base editor of embodiment 677, wherein said base editing polypeptide is operably fused to the C-terminus of said RGN polypeptide.

680. The base editor of any one of embodiments 667-679, wherein said base editor comprises one or more nuclear localization signal (NLS).

681. The base editor of embodiment 680, wherein said one or more NLS is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said base editing polypeptide or said RGN polypeptide.

682. The base editor of embodiment 680 or 681, wherein said one or more NLS is selected from the group consisting of SEQ ID NOs: 36, 37, 234, and 235.

683. The base editor of any one of embodiments 667-682, wherein said base editor further comprises one or more peptide linker.

684. The base editor of embodiment 683, wherein said one or more peptide linker comprises at least one NLS.

685. The base editor of embodiment 684, wherein said one or more peptide linker comprises two NLSs.

686. The base editor of any one of embodiments 683-685, wherein said one or more peptide linker is operably fused at the N-terminus, C-terminus, or both the N-terminus and C-terminus of said base editing polypeptide or said RGN polypeptide.

687. The base editor of any one of embodiments 683-686, wherein said one or more peptide linker has a formula of $-(SGGS)_x-NLS_m-(SGGS)_y-NLS_n-(SGGS)_z-$, wherein each of x, y, or z is 0, 1, 2, 3, or 4; and wherein each of m or n is 0 or 1.

688. The base editor of any one of embodiments 683-687, wherein said one or more peptide linker comprises one or more copies of amino acid sequence SGGS (SEQ ID NO: 241).

689. The base editor of any one of embodiments 683-688, wherein said one or more peptide linker has the sequence of any one of SEQ ID NOs: 236-241.

690. The base editor of embodiment 677, wherein said base editing polypeptide is operably fused to an internal location of said RGN polypeptide.

691. The base editor of embodiment 690, wherein said base editing polypeptide is operably fused within a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide, or wherein said base editing polypeptide is operably fused between a linker domain 2, a wedge domain, a RuvC domain, an HNH domain, a Rec-2 domain, or a PAM-interacting domain of said RGN polypeptide and another domain N-terminal or C-terminal to said linker domain 2, said wedge domain, said RuvC domain, said HNH domain, said Rec-2 domain, or said PAM-interacting domain.

692. The base editor of embodiment 691, wherein said RuvC domain is a RuvCIII domain.

693. The base editor of embodiment 692, wherein said base editing polypeptide is operably fused within a linker domain 2, a wedge domain, or a RuvCIII domain of said RGN polypeptide.

694. The base editor of embodiment 690, wherein said base editing polypeptide is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

695. The base editor of any one of embodiments 667-694, wherein said RGN polypeptide is capable of binding a target sequence in a target polynucleotide in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA), wherein said target sequence comprises a target strand and a non-target strand, and wherein said gRNA is capable of hybridizing to the target strand of the target sequence.

696. The base editor of embodiment 695, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target sequence.

697. The base editor of embodiment 696, wherein the PAM comprises a consensus nucleotide sequence set forth as NNGG.

698. The base editor of any one of embodiments 695-697, wherein said RGN polypeptide is capable of cleaving said target polynucleotide upon binding.

699. The base editor of embodiment 698, wherein said RGN polypeptide is capable of generating a double-stranded break.

700. The base editor of embodiment 698, wherein said RGN polypeptide is capable of generating a single-stranded break.

701. The base editor of any one of embodiments 667-697, wherein said RGN polypeptide comprises an HNH domain with at least one mutation that reduces or eliminates its nuclease activity.

702. The base editor of any one of embodiments 667-697, wherein said RGN polypeptide comprises an HNH domain with at least two mutations that reduces or eliminates its nuclease activity.

703. The base editor of any one of embodiments 667-697, wherein said RGN polypeptide does not comprise an HNH domain.

704. The base editor of embodiment 703, wherein said HNH domain of said RGN polypeptide has been replaced with said base editing polypeptide.

705. The base editor of any one of embodiments 667-697, wherein said RGN polypeptide is nuclease inactive.

706. The base editor of any one of embodiments 667-697, wherein said RGN polypeptide comprises an RGN nickase.

707. The base editor of embodiment 706, wherein said RGN nickase comprises a D16A and/or a H611A mutation(s).

708. The base editor of embodiment 706 or 707, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

709. The base editor of embodiment 708, wherein said RGN nickase comprises an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

710. The base editor of embodiment 708 or 709, wherein said RGN nickase comprises the amino acid sequence of any one of SEQ ID NOs: 182-196, and 271-285.

711. The base editor of any one of embodiments 667-710, wherein said base editor comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 258-270.

712. The base editor of any one of embodiments 667-711, wherein said base editor has an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 258-270.

713. A ribonucleoprotein (RNP) complex comprising the base editor of any one of embodiments 667-712 and a guide RNA bound to the RGN polypeptide.

714. The RNP complex of embodiment 713, wherein the guide RNA comprises a spacer that hybridizes to a eukaryotic target sequence.

715. The RNP complex of embodiment 714, wherein the eukaryotic target sequence comprises a mammalian target sequence.

716. A base editing system for modifying one or more target sequences in a target polynucleotide, said system comprising:

a) one or more guide RNAs (gRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more gRNAs; and

b) a base editor of any one of embodiments 667-712;

wherein said one or more gRNAs is capable of binding to said RGN polypeptide of said base editor.

717. The base editing system of embodiment 716, wherein each of the one or more gRNAs is capable of hybridizing to the target strand of said target sequence and forming a complex with the RGN polypeptide to direct said RGN polypeptide to bind to said target sequence.

718. The base editing system of embodiment 716 or 717, wherein the spacer hybridizes to a eukaryotic target sequence.

719. The base editing system of embodiment 718, wherein the eukaryotic target sequence comprises a mammalian target sequence.

720. The base editing system of any one of embodiments 716-719, wherein the one or more gRNAs comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence of any one of SEQ ID NOs: 33, 244, and 245, or that differs from any one of SEQ ID NOs: 33, 244, and 245 by 1 to 5 nucleotides.

721. The base editing system of embodiment 720, wherein the one or more gRNAs comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising a nucleotide sequence that differs from any one of SEQ ID NOs: 33, 244, and 245 by 5 nucleotides, by 4 nucleotides, by 3 nucleotides, by 2 nucleotides, or by 1 nucleotide.

722. The base editing system of embodiment 720, wherein the one or more gRNAs comprises a CRISPR RNA comprising a CRISPR RNA (crRNA) repeat comprising the nucleotide sequence set forth as any one of SEQ ID NOs: 33, 244, and 245.

723. The base editing system of any one of embodiments 716-722, wherein the one or more gRNAs comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

724. The base editing system of any one of embodiments 716-723, wherein the one or more gRNAs comprises a tracrRNA comprising a nucleotide sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% sequence identity to any one of SEQ ID NOs: 34, and 246-248.

725. The base editing system of any one of embodiments 716-724, wherein said polynucleotide encoding said one or more gRNAs and said polynucleotide encoding said base editor are on a single vector.

726. The base editing system of any one of embodiments 716-724, wherein said polynucleotide encoding said one or more gRNAs and said polynucleotide encoding said base editor are on separate vectors.

727. The base editing system of any one of embodiments 716-726, wherein said polynucleotide encoding said base editing polypeptide and said polynucleotide encoding said RGN polypeptide are on separate vectors.

728. The base editing system of any one of embodiments 716-724, wherein said polynucleotide encoding said one or more gRNAs, said polynucleotide encoding said base editing polypeptide, and said polynucleotide encoding said RGN polypeptide are all on separate vectors.

729. The base editing system of any one of embodiments 716-724, wherein said polynucleotide encoding said base editor is an RNA polynucleotide.

730. The base editing system of embodiment 729, wherein said RNA polynucleotide is an mRNA.

731. The base editing system of embodiment 729, wherein said RNA polynucleotide is a circRNA.

732. A cell comprising the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, or the base editing system of any one of embodiments 716-731.

733. The cell of embodiment 732, wherein the cell is a prokaryotic cell.

734. The cell of embodiment 732, wherein the cell is a eukaryotic cell.

735. The cell of embodiment 734, wherein the eukaryotic cell is a mammalian cell.

736. The cell of embodiment 735, wherein the mammalian cell is a human cell.

737. The cell of embodiment 736, wherein the human cell is an immune cell.

738. The cell of embodiment 736, wherein the human cell is a stem cell.

739. The cell of embodiment 738, wherein the stem cell is an induced pluripotent stem cell.

740. The cell of embodiment 734, wherein the eukaryotic cell is an insect or avian cell.

741. The cell of embodiment 734, wherein the eukaryotic cell is a fungal cell.

742. The cell of embodiment 734, wherein the eukaryotic cell is a plant cell.

743. A plant or plant part comprising the plant cell of embodiment 743.

744. A pharmaceutical composition comprising the one or more polynucleotides of any one of embodiments 584-637, the one or more vectors of any one of embodiments 638-650, the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, the cell of any one of embodiments 653-658 and 734-739, or the base editing system of any one of embodiments 716-731, and a pharmaceutically acceptable carrier.

745. A method for modifying a target polynucleotide comprising a target sequence, said method comprising delivering a base editing system according to any one of embodiments 716-731 to said target sequence or a cell comprising the target sequence, wherein said method generates a modified target polynucleotide, and wherein components of said base editing system are delivered simultaneously or sequentially to said target sequence or said cell comprising the target sequence.

746. The method of embodiment 745, wherein said modified target polynucleotide comprises mutation of at least one nucleotide in the target polynucleotide.

747. A method for modifying one or more target sequences in a target polynucleotide, said method comprising:

a) assembling a ribonucleoprotein (RNP) complex by combining:

i) a guide RNA; and

ii) a base editor of any one of embodiments 667-712;

under conditions suitable for formation of the RNP complex; and

b) contacting said target polynucleotide or a cell comprising said target polynucleotide with the assembled RNP complex;

thereby modifying said one or more target sequences in a target polynucleotide.

748. The method of any one of embodiments 745-747, wherein said method is performed *in vitro*, *in vivo*, or *ex vivo*.

749. The method of any one of embodiments 745-748, wherein said target polynucleotide is within a cell.

750. The method of embodiment 749, wherein the cell is a eukaryotic cell.

751. The method of embodiment 750, wherein the eukaryotic cell is a mammalian cell.

752. A cell produced according to the method of embodiment 749, wherein said target polynucleotide has been modified at said target sequence.

753. The cell of embodiment 752, wherein the cell is a prokaryotic cell.

754. The cell of embodiment 752, wherein the cell is a eukaryotic cell.

755. The cell of embodiment 754, wherein the eukaryotic cell is a mammalian cell.

756. The cell of embodiment 755, wherein the mammalian cell is a human cell.

757. The cell of embodiment 756, wherein the human cell is an immune cell.

758. The cell of embodiment 756, wherein the human cell is a stem cell.

759. The cell of embodiment 758, wherein the stem cell is an induced pluripotent stem cell.

760. The cell of embodiment 754, wherein the eukaryotic cell is an insect or avian cell.

761. The cell of embodiment 754, wherein the eukaryotic cell is a fungal cell.

762. The cell of embodiment 754, wherein the eukaryotic cell is a plant cell.

763. A plant or plant part comprising the plant cell of embodiment 762.

764. A pharmaceutical composition comprising the cell of any one of embodiments 754-759 and a pharmaceutically acceptable carrier.

765. A method for treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

administering to the subject the one or more polynucleotides of any one of embodiments 584-637, the one or more vectors of any one of embodiments 638-650, the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, the cell of any one of embodiments 653-658, 734-739, and 754-759, the base editing system of any one of embodiments 716-731, or the pharmaceutical composition of embodiment 744 or 764.

766. The method of embodiment 765, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

767. Use of the one or more polynucleotides of any one of embodiments 584-637, the one or more vectors of any one of embodiments 638-650, the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, the cell of any one of embodiments 653-658, 734-739, and 754-759, the base editing system of any one of embodiments 716-731, or the pharmaceutical composition of embodiment 744 or 764 for the treatment of a disease, disorder, or condition in a subject having or at risk of developing said disease, disorder, or condition.

768. The use of embodiment 767, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

769. Use of the one or more polynucleotides of any one of embodiments 584-637, the one or more vectors of any one of embodiments 638-650, the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, the cell of any one of embodiments 653-658, 734-739, and 754-759, the base editing system of any one of embodiments 716-731, or the pharmaceutical composition of embodiment 744 or 764 in the manufacture of a medicament useful for treating a disease, disorder, or condition.

770. The use of embodiment 769, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said medicament corrects said mutation.

771. The one or more polynucleotides of any one of embodiments 584-637, the one or more vectors of any one of embodiments 638-650, the base editor of any one of embodiments 667-712, the RNP complex of any one of embodiments 713-715, the cell of any one of embodiments 653-658, 734-739, and 754-759, the base editing system of any one of embodiments 716-731, or the pharmaceutical composition of embodiment 744 or 764 for use in treating a disease, disorder, or condition.

772. The one or more polynucleotides, the one or more vectors, the base editor, the RNP complex, the base editing system, the cell, or the pharmaceutical composition of embodiment 771, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said one or more polynucleotides, one or more vectors, base editor, RNP complex, base editing system, cell, or pharmaceutical composition corrects said mutation.

773. The RNP complex of embodiment 105 or the system of any one of embodiments 106-156 for use in binding a target polynucleotide sequence of a nucleic acid molecule.

774. The RNP complex of embodiment 105 or the system of any one of embodiments 106-156 for use in cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule.

775. The RNP complex of any one of embodiments 459-461 or the PE system of any one of embodiments 462-483 for use in modifying one or more target sequences in a target polynucleotide.

776. The RNP complex of any one of embodiments 713-715 or the base editing system of any one of embodiments 716-731 for use in modifying one or more target sequences in a target polynucleotide.

The following examples are offered by way of illustration and not by way of limitation.

EXAMPLES

Example 1. Identification of engineered variant LPG10145 RNA-guided nucleases (RGNs) with increased editing activity as compared to the wild-type LPG10145 RGN

The initial screen for engineered variant LPG10145 RGNs tested 1 to 2 guide RNAs and gene editing was assessed by flow cytometry or next generation sequencing (NGS).

Each construct encoding a variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1 in some analyses.

FIG. 1 shows percent gene editing efficiency for 581 constructs in an initial screen. Tracking of Indels by DEcomposition (TIDE) analysis was performed to determine % gene editing efficiency 2 days post-transfection.

123 variant LPG10145 nucleases having an increase in gene editing activity $\geq$ 15% as compared to LPG10145 were obtained from the initial screen (FIG. 2). The 123 engineered variant LPG10145 nucleases

were subjected to NGS for confirmation of editing activity, which narrowed the 123 hits to 71 variants. The 71 variants were further tested with additional guide RNAs.

FIG. 3 shows gene editing activity (% Indel) for variant LPG10145 nucleases tested with 3 additional guide RNAs (for a total of 4 guide RNAs tested: guides A, B, C, D). The guides target 4 different genes. Percent Indel (insertions/deletions) was determined by NGS 2 days post-transfection.

FIG. 4 shows confirmation of gene editing activity by NGS of variant LPG10145 nucleases having an increase in gene editing activity from the initial screen. Twenty R variants increased gene editing > 20%. A total of 4 guide RNAs were tested.

The top single variant LPG10145 nucleases with increased activity were ranked by statistical analysis (FIG. 5). FIG. 6 shows a structural alignment of LPG10145 from *S. thermophilus* (6M0W) with a guide RNA. Many mutations, but not all, are at the interface with DNA/RNA.

Example 2. Combinatorial variant LPG10145 RGNs have increased editing activity as compared to the single variant LPG10145 RGNs

Six variant LPG10145 nucleases were selected based on statistical analysis and structural modeling for a combinatorial library (FIG. 7). The single variants (1x variant), double variant combinations (2x variants), triple variant combinations (3x variants), quadruple variant combinations (4x variants), quintuple variant combinations (5x variant), and the sextuple combination (6x variant) are shown in FIG. 7. Sixty-three constructs encoding the 63 variant/combinatorial variant were generated and were tested with multiple different guide RNAs.

Each construct encoding a variant/combinatorial variant LPG10145 nuclease was delivered, along with a plasmid encoding a guide RNA, to HEK293T cells by plasmid lipofection. Each guide RNA was tested in duplicate (n=2). Percent Indel (insertions/deletions) was determined by NGS 2 days post-transfection.

Combinatorial variant LPG10145 nucleases were tested with seven different guide RNAs (shown as A, B1, B2, C1, C2, D1, and D2 in FIG. 8) to assess gene editing. Six of the seven guide RNAs tested showed significant increase in editing with the combinatorial variant LPG10145 RGNs (FIG. 8). Higher editing guide RNAs see less of an effect.

The highest gene editing was obtained with 3x, 4x, and 5x variants (FIG. 9). E778R and E969R were the common variants in the high editing populations. The gene editing efficiency of wild-type LPG10145 nuclease was normalized to 1. FIG. 10 shows the top 15 combinatorial variant LPG10145 nucleases with the highest significant editing.

Example 3: Off target analysis of variant LPG10145 RGNs

To assess the specificity of the variant LPG10145 RGNs, off target editing is determined at potential sites identified via bioinformatics. Potential off target sites for wild-type LPG10145 have been identified by

targets with less than five mismatches in the target sequence and at least one residue match in the PAM sequence (as described in International Publ. No. WO 2023/139557, filed January 23, 2023, which is incorporated herein in its entirety).

Ribonucleoprotein (RNP) delivery to mammalian cells and amplicon sequencing is used to test the specificity and off target editing of variant LPG10145 RGNs. Primers are used to amplify potential off target sites with sequence similarity to the on target site to look for off target editing. Off-target analysis is performed as described in International Publ. No. WO 2023/139557, filed January 23, 2023, which is incorporated herein in its entirety.

Example 4: Base editing using variant LPG10145 RGNs

To determine if the variant LPG10145 RGNs were able to perform adenine base editing in mammalian cells, each variant LPG10145 RGN was mutated to have nickase activity. A deaminase was operably fused to each variant LPG10145 RGN nickase to produce a fusion protein. A nickase variant of an RGN is referred to herein as "nRGN". Deaminase and variant LPG10145 nRGN nucleotide sequences codon optimized for mammalian expression were synthesized as fusion proteins with an N-terminal nuclear localization tag (NLS) and cloned into an expression plasmid. Each fusion protein comprised at least one NLS and had a detectable label/purification tag. Expression plasmids comprising an expression cassette encoding for a guide RNA were also produced. The plasmid comprising an expression cassette comprising a coding sequence for a variant LPG10145 nRGN-deaminase fusion protein and a plasmid comprising an expression cassette encoding a guide RNA were co-transfected into HEK293FT cells using, for example, Lipofectamine 2000 reagent (Life Technologies). Cells were then incubated. Following incubation, genomic DNA was extracted and the genomic region flanking the targeted genomic site was PCR amplified using primers. PCR products were purified and underwent NGS (e.g., on Illumina MiSeq). Typically, 100,000 of 250 bp paired-end reads (2 x 100,000 reads) were generated per amplicon. The reads were analyzed using CRISPResso (Pinello, et al. 2016 *Nature Biotech*, 34:695-697) to calculate the rates of editing. Output alignments were analyzed for INDEL formation or introduction of specific adenine mutations. Base editing is described in WO 2022/056254, which is incorporated herein in its entirety.

The following describes specific experiments testing WT and variant LPG10145 RGNs operably fused to a deaminase (i.e. the fusion protein is a base editor) in base editing. The deaminase was LPG50274, described in International Publ. No. WO 2024/095245, which is herein incorporated by reference in its entirety. HEK293T cells were transfected with two plasmids containing the base editor and the guide RNA in the following quantities: 160ng base editor fusion protein and 40ng guide RNA using lipofectamine 3000. The editor was designed as an inlaid base editor with the deaminase inserted after positions 910 in the HNH inactivated LPG10145 protein, which also has a D16A mutation. An SV40 N-terminal NLS and C-terminal nucleoplasmin NLS were added to the base editor. Two different guide RNA plasmids were tested in separate reactions, one targeting B2M and the other targeting TRAC. 15,000 HEK293T cells were seeded

the day prior to transfection. Two days after transfection, genomic DNA from the cells were harvested, the loci of interest amplified and sent for NGS readout.

Purified PCR amplicons were analyzed by NGS to evaluate the editing efficiency. Allele sequences and associated read counts were determined by CRISPResso. The base editing rate (% ABE) was determined by calculating the frequency of an A to G change at each position in the target region. The maximum frequency was selected for each sample and the average of two replicates are shown in Table 1.

The two guides tested were selected because they represent a poor performing guide (TRAC) and a higher performing guide (B2M). All tested engineered LPG10145 variants were functional as adenine base editors. The protein variants have a target-specific effect on editing. Some mutations improve editing at one target.

Table 1: Base editing activity with engineered LPG10145 variants

Nuclease Protein	Base Editor Protein	Additional Mutations	Avg. % ABE TRAC	Avg. % ABE B2M
LPG10145	LPG20062		1.2175	29.22
LPG10247 + D16A mutation	LPG20547	E778R_E969R	1.37	12.93
LPG10249 + D16A mutation	LPG20548	E778R_S55R_E969R	1.22	23.485
LPG10250 + D16A mutation	LPG20549	E778R_Q822R_S647R_E969R	0.435	10.89
LPG10251 + D16A mutation	LPG20550	E778R_Q822R_S55R_E969R	0.615	14.57
LPG10252 + D16A mutation	LPG20551	E778R_G856R_S647R_E969R	0.555	15.5
LPG10253 + D16A mutation	LPG20552	E778R_G856R_S55R_E969R	0.725	19.375
LPG10254 + D16A mutation	LPG20553	E778R_S647R_S55R_E969R	0.71	20.515
LPG10255 + D16A mutation	LPG20554	E778R_Q822R_S647R_S55R_E969R	0.265	8.135
LPG10256 + D16A mutation	LPG20555	E778R_G856R	1.365	22.82
LPG10257 + D16A mutation	LPG20556	E778R_Q822R_G856R	1.02	22.785
LPG10260 + D16A mutation	LPG20557	E778R_Q822R_S647R_S55R	0.19	8.37
LPG10261 + D16A mutation	LPG20558	E778R_Q822R_G856R_S647R_S55R	0.89	21.555

Example 5. Variant LPG10145 RGNs increase gene editing efficiency in reverse transcriptase (RT) mediated gene editing

The utility of variant LPG10145 RGNs in reverse transcriptase (RT) mediated gene editing (Anzalone, AV *et al. Nature* 2019, Nelson JW *et al. Nat Biotechnol* 2021, Chen PJ, *et al. Cell* 2021) was evaluated. RT mediated gene editing uses a reverse transcriptase fused to an RNA guided nickase that selectively nicks the non-target strand. A primer binding site (PBS) and an RT template are fused to the guide RNA. The PBS hybridizes with the nicked genomic DNA, allowing the genomic DNA to serve as a primer for reverse transcription of the RT template, which templates the desired changes. After reverse transcription of the RT template proceeds, a "flap" of DNA containing the edits is produced. Upon repair it is stably incorporated into the genome.

The 3' end of the tracrRNA is extended to include an RT template region that provides a template (with any desired edits) for the fused RT and a PBS. Alternatively, the RT template and PBS can be fused to the 5' end of the crRNA. The RT template is located on the 5' end of the PBS in both cases.

Variant LPG10145 RGNs fused to an RT were evaluated in RT mediated gene editing in cell-based assays, with RT templates directing changes in one or several genomic base pairs. The RT mediated gene editing composition was tested by sequencing the targeted genomic site using targeted amplicon sequencing. Editing can be evaluated by, e.g., reversion of a stop codon that is introduced into a green fluorescent protein gene (or other fluorescent protein)), by knocking out an endogenous gene, or by other means.

The following describes specific experiments testing WT and variant LPG10145 RGNs operably fused to an RT (an RT editor, RTE) in RT mediated gene editing. RT editors were generated by operably fusing a full-length Moloney Murine Leukemia Virus reverse transcriptase (MLV-RT) or a truncated MLV-RT to the N-terminus of HNH-catalytically-inactivated nickase of WT or variant LPG10145 RGN. The truncated MLV-RT (truncated to D497) lacks the RNase H domain and allows generation of smaller RT editors. An Xten 32 linker was used to fuse the RT to the nickase. Each construct contains an N terminal SV40 nuclear localization signal (NLS) and a C terminal nucleoplasmin NLS.

RT editors were also generated where the RT was operably fused at an internal location of a WT or variant LPG10145 RGN. Insertion of the RT within an RGN is referred to as an 'inlaid' architecture. For inlaid architectures, initial inlaid positions were identified from the ABE (adenine base editor) inlaid position described in International Publication No. WO 2024/095245, which is herein incorporated by reference in its entirety. For this approach, inlaid positions were selected that showed any levels of editing for ABE, and editor constructs were designed where the MLV-RT full length or MLV-RT RNAse H del was inlaid at each site. For each inlaid construct, flexible linker sequences of 4, 8, 16, or 32AA were used on each side of the MLV-RT.

Each architecture construct was tested against the previously identified positive control RTSGN with or without a nicking guide RNA. RT guides (RTSGNs) were designed to edit the B2M gene, with each RTSGN substituting a stop codon in exon 1 of B2M. Silent mutations were designed into some of the RT templates (RTTs) within the guide RNA to increase RT editing outcomes. Different PBS and RTT lengths were also tested in experiments.

Components needed for RT editing (RTE, RT guide, and nicking guide if used) were delivered by Lipofection 3000 to HEK293T cells that were seeded at 10k cells/well one day prior in 96-well tissue culture treated plates. Concentrations used were 150 ng of RTE plasmid and 50 ng of RTSGN plasmid, with 25-50 ng of nicking guide plasmid added to some RT reactions. Lipofected cells were grown for three days, and then the genomic DNA harvested and amplicons were generated using the appropriate primers. Purified PCR amplicons were analyzed by NGS to evaluate the editing efficiency using the primer set B2M_ex1RTE (SEQ ID NOs: 127 and 128). Allele sequences and associated read counts were determined by CRISPResso. The editing rate (RT Edit %) was determined by calculating the percent of read counts assigned to alleles containing (i) the stop codon substitution and (ii) no additional edits.

LPG10145-based RT editors were tested with the following RTSGNs: RTSGN1006 (SEQ ID NO: 129), RTSGN1017 (SEQ ID NO: 130), RTSGN1018 (SEQ ID NO: 131), and RTSGN1024 (SEQ ID NO: 132). The LPG10145-based RT editors were tested with nicking guides SGN009161 and SGN009162 (SEQ ID NOs: 133 and 134, respectively). SGN009162 nicks 96bp downstream of RTSGN1018 and 102bp downstream of RTSGN1006. SGN009161 nicks 30bp upstream of RTSGN1006 and 36bp upstream of RTSGN1018.

LPG10145-based RT editors showed similar or improved RT editing outcomes when the RNase H domain was removed from MLV-RT. Therefore, this domain can be removed without detrimental effect and the total size of the RT editor can be reduced for better delivery (Table 2).

Table 2. LPG10145-based RT editing with MLV-RT full-length or MLV-RT RH deletion

Nuclease	RT Used	Fusion Construct Name	Fusion Construct SEQ ID NO	RTSGN	Avg. RT Editing
E778R, S647R, S55R, E969R variant LPG10145 + H611A nickase mutation	Full length MLV-RT	LPG20300	135	RTSGN1006	1.334414
				RTSGN1017	3.31284515
				RTSGN1024	2.2660375
LPG10145 H611A nickase	MLV-RT with Rnase H domain deletion	LPG20164	136	RTSGN1006	1.8747375
				RTSGN1017	4.5383375
				RTSGN1024	2.502764

The S785 and P910 inlaid configurations of full length MLV-RT within LPG10145 nickase (H611A) exhibited functional RT editing at levels similar to the N-terminal fusion. Removal of the Rnase H domain further increased polymerase editing levels as an inlaid architecture (Table 3).

Table 3. LPG10145-based inlaid RT editing

MLV-RT Length	Inlaid Position	Linker Length	Fusion Construct ID	Fusion Construct SEQ ID NO	RTSGN1006 (% RT edit)	RTSGN1018 (% RT edit)
Full length	N-terminal fusion	34	LPG20170	137	1.203	3.099
	R666	4	LPG20171	138	0.067	0.125
		8	LPG20172	139	0.093	0.130
		16	LPG20173	140	0.143	0.104
		32	LPG20174	141	0.115	0.087
	S785	4	LPG20175	142	2.495	4.689
		8	LPG20176	143	3.098	4.981
		16	LPG20177	144	1.605	2.405
		32	LPG20178	145	1.696	2.657
	P910	4	LPG20179	146	0.291	0.300
		8	LPG20180	147	0.302	0.448
		16	LPG20181	148	0.540	0.846
		32	LPG20182	149	0.824	1.383
-Rnase H	R666	4	LPG20184	150	0.129	0.240
		8	LPG20185	151	0.198	0.231
		16	LPG20186	152	0.259	0.305
		32	LPG20187	153	0.350	0.302
	S785	4	LPG20188	154	4.415	8.373
		8	LPG20189	155	4.927	8.859
		16	LPG20190	156	4.715	9.463
		32	LPG20191	157	4.612	8.572
	P910	4	LPG20192	158	1.049	1.625
		8	LPG20193	159	1.444	2.026
		16	LPG20194	160	1.285	1.468
		32	LPG20195	161	1.509	2.109

LPG10145 variants exhibit similar, and in some cases, enhanced performance in comparison with the parental construct. Inlaid variants exhibit increased RT editing in comparison with N-terminal fusions (Table 4).

Table 4. RT editing for engineered LPG10145 variants

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
N-terminal fusion	LPG20170	137	H611A (SEQ ID NO: 125)	RTSGN1005 + SGN009162	0.16

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
				RTSGN1005	0.49
				RTSGN1006 + SGN009162	0.41
				RTSGN1006	0.44
				RTSGN1017 + SGN009162	0.73
				RTSGN1017	1.20
				RTSGN1018 + SGN009162	0.60
				RTSGN1018	1.32
	LPG20300	135	E778R, S647R, S55R, E969R, H611A (SEQ ID NO: 182)	RTSGN1005 + SGN009162	0.36
				RTSGN1005	0.80
				RTSGN1006 + SGN009162	0.29
				RTSGN1006	0.44
				RTSGN1017 + SGN009162	0.51
				RTSGN1017	1.26
				RTSGN1018 + SGN009162	0.43
				RTSGN1018	1.19
	LPG20301	162	E778R, G856R, S647R, E969R, H611A (SEQ ID NO: 183)	RTSGN1005 + SGN009162	0.27
				RTSGN1005	0.39
				RTSGN1006 + SGN009162	0.10
				RTSGN1006	0.27
				RTSGN1017 + SGN009162	0.67
				RTSGN1017	0.97
				RTSGN1018 + SGN009162	0.77
				RTSGN1018	1.11
	LPG20302	163	E778R, G856R, S55R, E969R, H611A (SEQ ID NO: 184)	RTSGN1005 + SGN009162	0.36
				RTSGN1005	0.51
				RTSGN1006 + SGN009162	0.25
				RTSGN1006	0.41
				RTSGN1017 + SGN009162	0.66
				RTSGN1017	1.26
				RTSGN1018 + SGN009162	0.97

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
				RTSGN1018	1.53
	LPG20303	164	E778R, G856R, E969R, H611A (SEQ ID NO: 185)	RTSGN1005 + SGN009162	0.34
				RTSGN1005	0.45
				RTSGN1006 + SGN009162	0.12
				RTSGN1006	0.47
				RTSGN1017 + SGN009162	0.54
				RTSGN1017	1.47
				RTSGN1018 + SGN009162	0.49
				RTSGN1018	1.37
	LPG20304	165	E778R, S55R, E969R, H611A (SEQ ID NO: 186)	RTSGN1005 + SGN009162	0.31
				RTSGN1005	0.38
				RTSGN1006 + SGN009162	0.12
				RTSGN1006	0.41
				RTSGN1017 + SGN009162	0.60
				RTSGN1017	1.22
				RTSGN1018 + SGN009162	1.43
				RTSGN1018	1.65
	LPG20305	166	E778R, E969R, H611A (SEQ ID NO: 187)	RTSGN1005 + SGN009162	0.30
				RTSGN1005	0.49
				RTSGN1006 + SGN009162	0.11
				RTSGN1006	1.31
				RTSGN1017 + SGN009162	0.82
				RTSGN1017	1.50
				RTSGN1018 + SGN009162	0.76
				RTSGN1018	1.85
	LPG20306	167	G856R, E778R, H611A (SEQ ID NO: 188)	RTSGN1005 + SGN009162	0.41
				RTSGN1005	0.72
				RTSGN1006 + SGN009162	0.22
				RTSGN1006	0.45
				RTSGN1017 + SGN009162	0.45
				RTSGN1017	1.35

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
				RTSGN1018 + SGN009162	1.20
				RTSGN1018	1.26
	LPG20307	168	Q822R, G856R, E778R, H611A (SEQ ID NO: 189)	RTSGN1005 + SGN009162	0.44
				RTSGN1005	0.77
				RTSGN1006 + SGN009162	0.30
				RTSGN1006	0.54
				RTSGN1017 + SGN009162	0.98
				RTSGN1017	1.96
				RTSGN1018 + SGN009162	0.57
				RTSGN1018	1.17
	LPG20308	169	Q822R, S55R, E778R, H611A (SEQ ID NO: 190)	RTSGN1005 + SGN009162	0.31
				RTSGN1005	0.57
				RTSGN1006 + SGN009162	0.14
				RTSGN1006	0.31
				RTSGN1017 + SGN009162	0.30
				RTSGN1017	0.61
				RTSGN1018 + SGN009162	0.25
				RTSGN1018	0.29
	LPG20309	170	E778R, Q822R, S55R, E969R, H611A (SEQ ID NO: 191)	RTSGN1005 + SGN009162	0.46
				RTSGN1005	1.09
				RTSGN1006 + SGN009162	0.30
				RTSGN1006	0.47
				RTSGN1017 + SGN009162	0.77
				RTSGN1017	1.16
				RTSGN1018 + SGN009162	0.98
				RTSGN1018	1.30
	LPG20310	171	Q822R, S647R, S55R, E778R, H611A (SEQ ID NO: 192)	RTSGN1005 + SGN009162	0.29
				RTSGN1005	0.65
				RTSGN1006 + SGN009162	0.15
				RTSGN1006	0.36

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
				RTSGN1017 + SGN009162	0.38
				RTSGN1017	0.67
				RTSGN1018 + SGN009162	0.30
				RTSGN1018	0.86
	LPG20311	172	E778R, Q822R, S647R, S55R, E969R, H611A (SEQ ID NO: 193)	RTSGN1005 + SGN009162	1.70
				RTSGN1005	0.72
				RTSGN1006 + SGN009162	0.22
				RTSGN1006	0.55
				RTSGN1017 + SGN009162	0.75
				RTSGN1017	1.38
				RTSGN1018 + SGN009162	0.34
				RTSGN1018	1.13
	LPG20312	173	E778R, Q822R, S647R, E969R, H611A (SEQ ID NO: 194)	RTSGN1005 + SGN009162	0.32
				RTSGN1005	0.45
				RTSGN1006 + SGN009162	0.18
				RTSGN1006	0.33
				RTSGN1017 + SGN009162	0.94
				RTSGN1017	1.48
				RTSGN1018 + SGN009162	0.96
				RTSGN1018	1.27
	LPG20313	174	Q822R, S647R, E778R, H611A (SEQ ID NO: 195)	RTSGN1005 + SGN009162	0.28
				RTSGN1005	0.65
				RTSGN1006 + SGN009162	0.30
				RTSGN1006	0.45
				RTSGN1017 + SGN009162	0.33
				RTSGN1017	1.43
				RTSGN1018 + SGN009162	0.59
				RTSGN1018	1.59
	LPG20314	175	Q822R, G856R, S647R, S55R,	RTSGN1005 + SGN009162	0.28
				RTSGN1005	0.63

Inlaid Position	Fusion Construct ID	Fusion Construct SEQ ID NO	LPG10145 variant (SEQ ID NO)	RTSGN	% RT Edit
			E778R, H611A (SEQ ID NO: 196)	RTSGN1006 + SGN009162	0.26
				RTSGN1006	0.35
				RTSGN1017 + SGN009162	0.33
				RTSGN1017	1.04
				RTSGN1018 + SGN009162	0.52
				RTSGN1018	0.98
S785	LPG20190	156	H611A (SEQ ID NO: 125)	RTSGN1006	3.85
				RTSGN1018	5.25
	LPG20328	176	E778R, E969R, H611A (SEQ ID NO: 187)	RTSGN1006	6.42
				RTSGN1018	10.74
	LPG20329	177	G856R, E778R, H611A (SEQ ID NO: 188)	RTSGN1006	7.19
				RTSGN1018	10.31
	LPG20330	178	Q822R, G856R, E778R, H611A (SEQ ID NO: 189)	RTSGN1006	6.07
				RTSGN1018	11.47
	LPG20331	179	E778R, Q822R, S55R, E969R, H611A (SEQ ID NO: 191)	RTSGN1006	3.46
				RTSGN1018	5.39
	LPG20332	180	E778R, Q822R, S647R, S55R, E969R, H611A (SEQ ID NO: 193)	RTSGN1006	5.91
				RTSGN1018	6.08
	LPG20333	181	Q822R, S647R, E778R, H611A (SEQ ID NO: 195)	RTSGN1006	5.38
				RTSGN1018	8.77

Example 6: RT editing with engineered LPG10145 variants

HEK293T cells were engineered to contain a murine Hao1 sequence. HEK293T cells were transfected using lipofectamine 3000 with two plasmids containing the RT editor and the RTSGN in the following quantities: 150ng RT editor fusion protein and 50ng RTSGN (PEgRNA). 10,000 HEK293T cells

were seeded the day prior to transfection. Three days after transfection, genomic DNA from the cells were harvested, the loci of interest amplified and sent for NGS readout.

Purified PCR amplicons were analyzed by NGS to evaluate the editing efficiency. Allele sequences and associated read counts were determined by CRISPResso. The editing rate (% RT Edit) was determined by calculating the percent of read counts assigned to alleles containing the stop codon. Results are shown in Table 7.

All RTSGNs contained the same spacer sequence. The PBS and RTT lengths were varied. The RTSGN parameters are shown in Table 5. The RTSGNs (PEgRNAs) were designed to introduce a stop codon into the murine Hao1 gene sequence. Some RT templates contain an additional silent mutation, noted as "add SM" in Table 5.

The parental LPG10145 (LPG10145 H611A nickase) was tested as an RT editor with the RT inlaid at position 785 in the nickase; the fusion construct LPG20190 contains linkers and NLSes. Three engineered LPG10145 variants were also tested as inlaid RT editors with the RT inserted after position 785 in the nickase. The fusion constructs are described in Table 6.

The parental fusion LPG20190 was replicated 8 times. The engineered LPG10145 variant/inlaid RT fusions were replicated twice. RT editing efficiency was higher with an RT template 23 nucleotides or longer. Variant LPG20328 showed higher editing than the parental LPG10145 nickase-based RT editors with some RTSGN combinations.

Table 5. PEgRNA parameters

PEgRNA Name	Target Length (nt)	PBS Length (nt)	RTT Length (nt)	Additional Edits
RTSGN2779	25	18	29	
RTSGN2778	25	15	29	
RTSGN2777	25	12	29	
RTSGN2776	25	18	29	add SM
RTSGN2775	25	15	29	add SM
RTSGN2774	25	12	29	add SM
RTSGN2773	25	18	26	
RTSGN2772	25	15	26	
RTSGN2771	25	12	26	
RTSGN2770	25	18	23	
RTSGN2769	25	15	23	
RTSGN2768	25	12	23	
RTSGN2767	25	18	20	
RTSGN2766	25	15	20	
RTSGN2765	25	12	20	
RTSGN2764	25	18	23	add SM

RTSGN2763	25	15	23	add SM
RTSGN2762	25	12	23	add SM
RTSGN2761	25	18	20	add SM
RTSGN2760	25	15	20	add SM
RTSGN2759	25	12	20	add SM
RTSGN2758	25	18	17	
RTSGN2757	25	15	17	
RTSGN2756	25	12	17	

Table 6. Description of LPG10145 RGNs tested

Parental Nuclease	**Additional Mutations**	**Inlaid location**	**Fusion Construct Identifier**
LPG10145 nickase H611A	None	785	LPG20190
LPG10145 nickase H611A	E778R, E969R	785	LPG20328
LPG10145 nickase H611A	G856R, E778R	785	LPG20329
LPG10145 nickase H611A	Q822R, G856R, E778R	785	LPG20330

Table 7. Percent RT Editing results with inlaid engineered LPG10145 variants

	LPG20190	LPG20328	LPG20329	LPG20330
RTSGN2756	2.87	0.97	1.72	1.02
RTSGN2757	2.38	0.24	1.84	0.74
RTSGN2758	1.58	1.1	0.5	0.31
RTSGN2759	3.78	1.21	2.33	1.86
RTSGN2760	1.85	1.08	0.73	1
RTSGN2761	1.15	0.79	0.54	0.79
RTSGN2762	6.31	4.71	4.98	4.57
RTSGN2763	4.14	1.93	3.56	1.71
RTSGN2764	2.04	0.72	1.54	0.79
RTSGN2765	11.11	5.69	16.55	10.31
RTSGN2766	6.57	7.07	9.92	7.27
RTSGN2767	4.04	2.89	5.5	6.83
RTSGN2768	8.72	9	7.84	8.04
RTSGN2769	9.6	11.49	11.81	7.64

RTSGN2770	6.91	13.17	9.77	6.46
RTSGN2771	14.61	21.11	13.22	11.05
RTSGN2772	10.04	13.11	10.52	9.65
RTSGN2773	8.77	13.15	8	6.64
RTSGN2774	6.8	13.09	8.36	5.94
RTSGN2775	6.87	14.1	7.06	4.68
RTSGN2776	5.44	9.18	4.77	3.02
RTSGN2777	7.4	7.66	6.37	4.82
RTSGN2778	6.76	13.61	8.69	7.44
RTSGN2779	4.28	5.29	4.23	4.26

5

Table 8: Description of the Sequences of the application

SEQ ID NO	Description	SEQ ID NO	Description	SEQ ID NO	Description
1	LPG10145	101	LPG50179 deaminase	197	RTSGN2779 PEgRNA
2	Q822R, G856R, S647R, S55R_E778R (LPG10261)	102	LPG50180 deaminase	198	RTSGN2778 PEgRNA
3	Q822R, S647R, S55R_E778R (LPG10260)	103	LPG50181 deaminase	199	RTSGN2777 PEgRNA
4	Q822R, S55R_E778R (LPG10259)	104	LPG50182 deaminase	200	RTSGN2776 PEgRNA
5	Q822R, S647R_E778R (LPG10258)	105	APG09980.1 deaminase	201	RTSGN2775 PEgRNA
6	Q822R, G856R_E778R (LPG10257)	106	APG05840.1 deaminase	202	RTSGN2774 PEgRNA
7	G856R_E778R (LPG10256)	107	APG00868.1 deaminase	203	RTSGN2773 PEgRNA
8	E778R, Q822R, S647R, S55R_E969R (LPG10255)	108	APG30125 deaminase	204	RTSGN2772 PEgRNA
9	E778R, S647R, S55R_E969R (LPG10254)	109	APG30126 deaminase	205	RTSGN2771 PEgRNA
10	E778R, G856R, S55R_E969R (LPG10253)	110	APG30127 deaminase	206	RTSGN2770 PEgRNA
11	E778R, G856R, S647R_E969R (LPG10252)	111	APG30128 deaminase	207	RTSGN2769 PEgRNA
12	E778R, Q822R, S55R_E969R (LPG10251)	112	APG30129 deaminase	208	RTSGN2768 PEgRNA
13	E778R, Q822R, S647R_E969R (LPG10250)	113	APG30130 deaminase	209	RTSGN2767 PEgRNA
14	E778R, S55R_E969R (LPG10249)	114	NC_041760 promoter	210	RTSGN2766 PEgRNA
15	E778R, G856R_E969R (LPG10248)	115	NW_004848155 promoter	211	RTSGN2765 PEgRNA
16	E778R_E969R (LPG10247)	116	LR738627 promoter	212	RTSGN2764 PEgRNA

17	LPG10145 coding sequence	117	NC_044556 promoter	213	RTSGN2763 PEgRNA
18	Q822R, G856R, S647R, S55R_E778R (LPG10261) coding sequence	118	XM_030845548 promoter	214	RTSGN2762 PEgRNA
19	Q822R, S647R, S55R_E778R (LPG10260) coding sequence	119	Mini NC_041760 promoter	215	RTSGN2761 PEgRNA
20	Q822R, S55R_E778R (LPG10259) coding sequence	120	Mini NW_004848155 promoter	216	RTSGN2760 PEgRNA
21	Q822R, S647R_E778R (LPG10258) coding sequence	121	Mini LR738627 promoter	217	RTSGN2759 PEgRNA
22	Q822R, G856R_E778R (LPG10257) coding sequence	122	Mini NC_044556 promoter	218	RTSGN2758 PEgRNA
23	G856R_E778R (LPG10256) coding sequence	123	Mini XM_030845548 promoter	219	RTSGN2757 PEgRNA
24	E778R, Q822R, S647R, S55R_E969R (LPG10255) coding sequence	124	LPG10145 D16A nickase	220	RTSGN2756 PEgRNA
25	E778R, S647R, S55R_E969R (LPG10254) coding sequence	125	LPG10145 H611A nickase	221	primer binding site 18 nt length in RTSGNs: 2758, 2761, 2764, 2767, 2770, 2773, 2776, 2779
26	E778R, G856R, S55R_E969R (LPG10253) coding sequence	126	dLPG10145 (D16A H611A)	222	primer binding site 15 nt length in RTSGNs: 2757, 2760, 2763, 2766, 2769, 2772, 2775, 2778
27	E778R, G856R, S647R_E969R (LPG10252)	127	B2M_ex1RTE left primer	223	primer binding site 12 nt length in RTSGNs: 2756, 2759,

	coding sequence				2762, 2765, 2768, 2771, 2774, 2777
28	E778R, Q822R, S55R_E969R (LPG10251) coding sequence	128	B2M_ex1RTE right primer	224	spacer sequence for RTSGN2756-RTSGN2779
29	E778R, Q822R, S647R_E969R (LPG10250) coding sequence	129	RTSGN1006 PEgRNA	225	target sequence for RTSGN2756-RTSGN2779
30	E778R, S55R_E969R (LPG10249) coding sequence	130	RTSGN1017 PEgRNA	226	RT template 29 nt length in RTSGNs: 2777, 2778, 2779
31	E778R, G856R_E969R (LPG10248) coding sequence	131	RTSGN1018 PEgRNA	227	RT template 29 nt length add SM in RTSGNs: 2774, 2775, 2776
32	E778R_E969R (LPG10247) coding sequence	132	RTSGN1024 PEgRNA	228	RT template 26 nt length in RTSGNs: 2771, 2772, 2773
33	LPG10145 crRNA repeat	133	SGN009161 nicking gRNA	229	RT template 23 nt length in RTSGNs: 2768, 2769, 2770
34	LPG10145 tracrRNA	134	SGN009162 nicking gRNA	230	RT template 23 nt length add SM in RTSGNs: 2762, 2763, 2764
35	UGI domain	135	LPG20300 polymerase editor	231	RT template 20 nt length in RTSGNs: 2765, 2766, 2767
36	SV40 Nuclear Localization Signal	136	LPG20164 polymerase editor	232	RT template 20 nt length add SM in RTSGNs: 2759, 2760, 2761
37	Nucleoplasmin Nuclear Localization Signal	137	LPG20170 polymerase editor	233	RT template 17 nt length in RTSGNs: 2756, 2757, 2758

38	Linker nucleotide sequence (for sgRNAs)	138	LPG20171 polymerase editor	234	cMyc NLS
39	CSB	139	LPG20172 polymerase editor	235	5' SV40 NLS
40	USP2	140	LPG20173 polymerase editor	236	L1 linker
41	human RNA polymerase III U6 promoter	141	LPG20174 polymerase editor	237	L2 linker
42	APG00868 deaminase	142	LPG20175 polymerase editor	238	L3 linker
43	APG02316 deaminase	143	LPG20176 polymerase editor	239	Xten32 linker
44	APG05241 deaminase	144	LPG20177 polymerase editor	240	Linker-2
45	APG05840 deaminase	145	LPG20178 polymerase editor	241	Linker-1
46	APG07280 deaminase	146	LPG20179 polymerase editor	242	FLAG tag
47	APG07386 deaminase	147	LPG20180 polymerase editor	243	HiBit Tag
48	APG08360 deaminase	148	LPG20181 polymerase editor	244	LPG10145 crRNA repeat 2
49	APG09260 deaminase	149	LPG20182 polymerase editor	245	LPG10145 crRNA repeat 3
50	APG09688 deaminase	150	LPG20184 polymerase editor	246	LPG10145 tracrRNA 2
51	APG09980 deaminase	151	LPG20185 polymerase editor	247	LPG10145 tracrRNA 3
52	APG09982 deaminase	152	LPG20186 polymerase editor	248	LPG10145 tracrRNA 4
53	APG03724 deaminase	153	LPG20187 polymerase editor	249	LPG10145 guide RNA backbone 1
54	APG09949 deaminase	154	LPG20188 polymerase editor	250	LPG10145 guide RNA backbone 2
55	APG08196 deaminase	155	LPG20189 polymerase editor	251	LPG10145 guide RNA backbone 3

56	APG06333 deaminase	156	LPG20190 polymerase editor	252	LPG10145 guide RNA backbone 4
57	APG06489 deaminase	157	LPG20191 polymerase editor	253	LPG10145 PAM-interacting domain
58	APG08449 deaminase	158	LPG20192 polymerase editor	254	MLV-RT
59	APG05174 deaminase	159	LPG20193 polymerase editor	255	MLV-RT deltaRH
60	APG09102 deaminase	160	LPG20194 polymerase editor	256	Dominant negative MLH1
61	APG05723 deaminase	161	LPG20195 polymerase editor	257	LPG50274 deaminase
62	LPG50140 deaminase	162	LPG20301 polymerase editor	258	LPG20547 base editor
63	LPG50141 deaminase	163	LPG20302 polymerase editor	259	LPG20548 base editor
64	LPG50142 deaminase	164	LPG20303 polymerase editor	260	LPG20549 base editor
65	LPG50143 deaminase	165	LPG20304 polymerase editor	261	LPG20550 base editor
66	LPG50144 deaminase	166	LPG20305 polymerase editor	262	LPG20551 base editor
67	LPG50145 deaminase	167	LPG20306 polymerase editor	263	LPG20552 base editor
68	LPG50146 deaminase	168	LPG20307 polymerase editor	264	LPG20553 base editor
69	LPG50147 deaminase	169	LPG20308 polymerase editor	265	LPG20554 base editor
70	LPG50148 deaminase	170	LPG20309 polymerase editor	266	LPG20555 base editor
71	LPG50149 deaminase	171	LPG20310 polymerase editor	267	LPG20556 base editor
72	LPG50150 deaminase	172	LPG20311 polymerase editor	268	LPG20557 base editor
73	LPG50151 deaminase	173	LPG20312 polymerase editor	269	LPG20558 base editor

74	LPG50152 deaminase	174	LPG20313 polymerase editor	270	LPG20062 base editor
75	LPG50153 deaminase	175	LPG20314 polymerase editor	271	LPG10247 with D16A mutation
76	LPG50154 deaminase	176	LPG20328 polymerase editor	272	LPG10248 with D16A mutation
77	LPG50155 deaminase	177	LPG20329 polymerase editor	273	LPG10249 with D16A mutation
78	LPG50156 deaminase	178	LPG20330 polymerase editor	274	LPG10250 with D16A mutation
79	LPG50157 deaminase	179	LPG20331 polymerase editor	275	LPG10251 with D16A mutation
80	LPG50158 deaminase	180	LPG20332 polymerase editor	276	LPG10252 with D16A mutation
81	LPG50159 deaminase	181	LPG20333 polymerase editor	277	LPG10253 with D16A mutation
82	LPG50160 deaminase	182	LPG10145 variant: E778R, S647R, S55R, E969R, H611A	278	LPG10254 with D16A mutation
83	LPG50161 deaminase	183	LPG10145 variant: E778R, G856R, S647R, E969R, H611A	279	LPG10255 with D16A mutation
84	LPG50162 deaminase	184	LPG10145 variant: E778R, G856R, S55R, E969R, H611A	280	LPG10256 with D16A mutation
85	LPG50163 deaminase	185	LPG10145 variant: E778R, G856R, E969R, H611A	281	LPG10257 with D16A mutation
86	LPG50164 deaminase	186	LPG10145 variant: E778R, S55R, E969R, H611A	282	LPG10258 with D16A mutation
87	LPG50165 deaminase	187	LPG10145 variant: E778R, E969R, H611A	283	LPG10259 with D16A mutation
88	LPG50166 deaminase	188	LPG10145 variant: G856R, E778R, H611A	284	LPG10260 with D16A mutation

89	LPG50167 deaminase	189	LPG10145 variant: Q822R, G856R, E778R, H611A	285	LPG10261 with D16A mutation
90	LPG50168 deaminase	190	LPG10145 variant: Q822R, S55R, E778R, H611A		
91	LPG50169 deaminase	191	LPG10145 variant: E778R, Q822R, S55R, E969R, H611A		
92	LPG50170 deaminase	192	LPG10145 variant: Q822R, S647R, S55R, E778R, H611A		
93	LPG50171 deaminase	193	LPG10145 variant: E778R, Q822R, S647R, S55R, E969R, H611A		
94	LPG50172 deaminase	194	LPG10145 variant: E778R, Q822R, S647R, E969R, H611A		
95	LPG50173 deaminase	195	LPG10145 variant: Q822R, S647R, E778R, H611A		
96	LPG50174 deaminase	196	LPG10145 variant: Q822R, G856R, S647R, S55R, E778R, H611A		
97	LPG50175 deaminase				
98	LPG50176 deaminase				
99	LPG50177 deaminase				
100	LPG50178 deaminase				

THAT WHICH IS CLAIMED:

1. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

2. The nucleic acid molecule of claim 1, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is a positively charged amino acid residue.

3. The nucleic acid molecule of claim 1 or 2, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is an R.

4. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

5. The nucleic acid molecule of claim 4, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

6. The nucleic acid molecule of claim 5, wherein said RGN polypeptide comprises an amino acid sequence having 100% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

7. The nucleic acid molecule of any one of claims 1-6, wherein said polynucleotide encoding an RGN polypeptide is operably linked to a promoter heterologous to said polynucleotide.

8. The nucleic acid molecule of any one of claims 1-7, wherein said RGN polypeptide is nuclease inactive or is a nickase.

9. The nucleic acid molecule of claim 8, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

10. The nucleic acid molecule of claim 8 or 9, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

11. The nucleic acid molecule of any one of claims 1-10, wherein said RGN polypeptide is operably linked to a heterologous polypeptide.

12. The nucleic acid molecule of claim 11, wherein said heterologous polypeptide is operably linked to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

13. The nucleic acid molecule of claim 11 or 12, wherein said heterologous polypeptide is a polymerase editing polypeptide.

14. The nucleic acid molecule of claim 13, wherein said polymerase editing polypeptide comprises a DNA polymerase.

15. The nucleic acid molecule of claim 13, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

16. The nucleic acid molecule of claim 15, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

17. The nucleic acid molecule of claim 11 or 12, wherein the heterologous polypeptide is a base-editing polypeptide.

18. The nucleic acid molecule of claim 17, wherein the base-editing polypeptide is a deaminase.

19. The nucleic acid molecule of claim 18, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

20. The nucleic acid molecule of claim 11 or 12, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

21. The nucleic acid molecule of claim 20, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

22. The nucleic acid molecule of any one of claims 1-21, wherein the polynucleotide encoding the RGN polypeptide is codon optimized for expression in a eukaryotic cell.

23. A vector comprising the nucleic acid molecule of any one of claims 1-22.

24. The vector of claim 23, further comprising at least one nucleotide sequence encoding a guide RNA.

25. The vector of claim 24, wherein the guide RNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

26. The vector of claim 24 or 25, wherein the guide RNA comprises a tracrRNA comprising a nucleotide sequence having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

27. The vector of any one of claims 24-26, where said guide RNA is a single guide RNA.

28. The vector of claim 27, wherein said single guide RNA further comprises an extension comprising an edit template for polymerase editing.

29. The vector of any one of claims 24-26, wherein said guide RNA is a dual-guide RNA.

30. A cell comprising the nucleic acid molecule of any one of claims 1-22 or the vector of any one of claims 23-29.

31. A method for making an RGN polypeptide comprising culturing the cell of claim 30 under conditions in which the RGN polypeptide is expressed.

32. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

33. The RGN polypeptide of claim 32, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

34. The RGN polypeptide of claim 32 or 33, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino

acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is an R.

35. An RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

36. The RGN polypeptide of claim 35, wherein said RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

37. The RGN polypeptide of claim 36, wherein said RGN polypeptide comprises an amino acid sequence having 100% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

38. The RGN polypeptide of any one of claims 32-37, wherein said RGN polypeptide is an isolated RGN polypeptide.

39. The RGN polypeptide of any one of claims 32-38, wherein said RGN polypeptide is capable of binding a target polynucleotide sequence of a DNA molecule in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target polynucleotide sequence.

40. The RGN polypeptide of claim 39, wherein said RGN polypeptide recognizes a protospacer adjacent motif (PAM) that is 3′ of said target polynucleotide sequence.

41. The RGN polypeptide of claim 40, wherein said RGN polypeptide recognizes a PAM having a consensus nucleotide sequence set forth as NNGG.

42. The RGN polypeptide of claim 40 or 41, wherein the RGN polypeptide comprises a PAM-interacting domain comprising the amino acid sequence set forth as SEQ ID NO: 253.

43. The RGN polypeptide of any one of claims 39-42, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence upon binding.

44. The RGN polypeptide of claim 43, wherein cleavage by said RGN polypeptide generates a double-stranded break.

45. The RGN polypeptide of claim 43, wherein cleavage by said RGN polypeptide generates a single-stranded break.

46. The RGN polypeptide of any one of claims 32-42, wherein said RGN polypeptide is nuclease inactive or a nickase.

47. The RGN polypeptide of claim 46, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

48. The RGN polypeptide of claim 46 or 47, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

49. The RGN polypeptide of any one of claims 32-48, wherein said RGN polypeptide is operably fused to a heterologous polypeptide.

50. The RGN polypeptide of claim 49, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

51. The RGN polypeptide of claim 49 or 50, wherein said heterologous polypeptide is a polymerase editing polypeptide.

52. The RGN polypeptide of claim 51, wherein said polymerase editing polypeptide comprises a DNA polymerase.

53. The RGN polypeptide of claim 51, wherein said polymerase editing polypeptide comprises a reverse transcriptase.

54. The RGN polypeptide of claim 56, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

55. The RGN polypeptide of claim 49 or 50, wherein the heterologous polypeptide is a base-editing polypeptide.

56. The RGN polypeptide of claim 55, wherein the base-editing polypeptide is a deaminase.

57. The RGN polypeptide of claim 56, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

58. The RGN polypeptide of claim 56 or 57, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

59. The RGN polypeptide of claim 49 or 50, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

60. The RGN polypeptide of claim 59, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

61. The RGN polypeptide of any one of claims 32-60, wherein the RGN polypeptide comprises one or more nuclear localization signals.

62. A ribonucleoprotein (RNP) complex comprising the RGN polypeptide of any one of claims 32-61 and the guide RNA bound to the RGN polypeptide.

63. A system, said system comprising:

A) one or more guide RNAs (gRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more gRNAs; and

B) an RNA-guided nuclease (RGN) polypeptide, or a polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1.

64. The system of claim 63, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975of said amino acid sequence is a positively charged amino acid residue.

65. The system of claim 63 or 64, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

66. The system of claim 65, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

67. The system of claim 66, wherein the RGN polypeptide comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

68. The system of claim 67, wherein said RGN polypeptide comprises an amino acid sequence having 100% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

69. The system of any one of claims 63-68, wherein the polynucleotide comprising a nucleotide sequence encoding the RGN polypeptide comprises an mRNA comprising a 5' untranslated region (UTR) and/or a 3' UTR, wherein the 5' UTR, the 3' UTR, or both are heterologous to the mRNA.

70. The system of any one of claims 63-68, wherein at least one of said nucleotide sequences encoding the one or more gRNAs and said nucleotide sequence encoding the RGN polypeptide is operably linked to a promoter heterologous to said nucleotide sequence.

71. The system of any one of claims 63-70, wherein said gRNA is a single guide RNA.

72. The system of claim 71, wherein said single guide RNA further comprises an extension comprising an edit template for polymerase editing.

73. The system of any one of claims 63-70, wherein said gRNA is a dual-guide RNA.

74. The system of any one of claims 63-73, wherein said gRNA comprises a CRISPR RNA (crRNA) comprising a CRISPR repeat comprising a nucleotide sequence set forth as SEQ ID NO: 33, 244, or 245, or that differs from SEQ ID NO: 33, 244, or 245 by 1 to 5 nucleotides.

75. The system of any one of claims 63-74, wherein the guide RNA comprises a a tracrRNA having at least 90% sequence identity to SEQ ID NO: 34, 246, 247, or 248.

76. The system of any one of claims 63-75, wherein the one or more gRNAs are capable of hybridizing to a target polynucleotide sequence of a DNA molecule, and wherein the one or more guide RNAs are capable of forming a complex with the RGN polypeptide in order to direct said RGN polypeptide to bind to said target polynucleotide sequence of the DNA molecule.

77. The system of claim 76, wherein said target polynucleotide sequence is a eukaryotic target polynucleotide sequence.

78. The system of claim 76 or 77, wherein the target polynucleotide sequence is within a cell.

79. The system of any one of claims 76-78, wherein the complex comprising the one or more gRNAs and the RGN polypeptide directs cleavage of the target polynucleotide sequence.

80. The system of any one of claims 63-78, wherein said RGN polypeptide is nuclease inactive or is a nickase.

81. The system of claim 80, wherein said RGN polypeptide comprises a D16A and/or a H611A mutation(s).

82. The system of claim 80 or 81, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 182-196, and 271-285.

83. The system of any one of claims 63-82, wherein said RGN polypeptide is operably linked to a heterologous polypeptide.

84. The system of claim 83, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

85. The system of claim 83 or 84, wherein the heterologous polypeptide is a polymerase editing polypeptide.

86. The system of claim 85, wherein the polymerase editing polypeptide comprises a DNA polymerase.

87. The system of claim 85, wherein the polymerase editing polypeptide comprises a reverse transcriptase.

88. The system of claim 87, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

89. The system of claim 83 or 84, wherein the heterologous polypeptide is a base-editing polypeptide.

90. The system of claim 89, wherein the base-editing polypeptide is a deaminase.

91. The system of claim 90, wherein the deaminase is a cytosine deaminase or an adenine deaminase.

92. The system of claim 90 or 91, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

93. The system of claim 83 or 84, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

94. The system of claim 93, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

95. The system of any one of claims 63-94, wherein said system further comprises one or more donor polynucleotides.

96. A cell comprising the RGN polypeptide of any one of claims 32-61, the RNP complex of claim 62, or the system of any one of claims 63-95.

97. A pharmaceutical composition comprising the nucleic acid molecule of any one of claims 1-22, the vector of any one of claims 23-29, the cell of claim 30 or 96, the RGN polypeptide of any one of claims 32-61, the RNP complex of claim 62, or the system of any one of claims 63-95, and a pharmaceutically acceptable carrier.

98. A method for binding a target polynucleotide sequence of a nucleic acid molecule comprising delivering a system according to any one of claims 63-95, to said target polynucleotide sequence or a cell comprising the target polynucleotide sequence.

99. A method for cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule comprising delivering a system according to any one of claims 63-95 to said target polynucleotide sequence or a cell comprising the nucleic acid molecule, wherein cleavage or modification of said target polynucleotide sequence occurs.

100. A method for binding a target polynucleotide sequence of a nucleic acid molecule comprising:

a) assembling a ribonucleoprotein (RNP) complex under conditions suitable for formation of the RNP complex by combining:

i) one or more guide RNAs (gRNAs); and

ii) an RNA-guided nuclease (RGN) polypeptide comprising an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1;

and

b) contacting said target polynucleotide sequence or a cell comprising said target polynucleotide sequence with the assembled RNP complex,

thereby binding said target polynucleotide sequence with said RNP complex.

101. The method of claim 100, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

102. The method of claim 100 or 101, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid

residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

103. The method of claim 102, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

104. The method of claim 103, wherein said RGN polypeptide comprises an amino acid sequence having 100% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

105. The method of any one of claims 100-104, wherein said RGN polypeptide is capable of cleaving said target polynucleotide sequence, thereby allowing for the cleaving and/or modifying of said target polynucleotide sequence.

106. A method for cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule, comprising contacting the nucleic acid molecule with:

a) an RNA-guided nuclease (RGN) polypeptide, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence differs from the corresponding amino acid residue in SEQ ID NO: 1; and

b) one or more guide RNAs (gRNAs) capable of targeting the RGN of (a) to the target polynucleotide sequence,

thereby cleaving and/or modifying said target polynucleotide sequence to generate a modified target polynucleotide sequence.

107. The method of claim 106, wherein the one or more gRNAs is capable of hybridizing to the target polynucleotide sequence and binding to said RGN polypeptide.

108. The method of claim 106 or 107, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is a positively charged amino acid residue.

109. The method of any one of claims 106-108, wherein said RGN polypeptide comprises an amino acid sequence having at least 85% sequence identity to SEQ ID NO: 1, and wherein the amino acid residue at one or more of amino acid positions 52, 55, 86, 472, 533, 541, 643, 647, 653, 745, 774, 778, 780, 795, 822, 843, 856, 871, 872, 900, 911, 913, 954, 958, 968, 969, 973, 974, and 975 of said amino acid sequence is an R.

110. The method of claim 109, wherein said RGN polypeptide comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(ee) the amino acid sequence set forth as SEQ ID NO: 3, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is substituted by an R;

(ff) the amino acid sequence set forth as SEQ ID NO: 4, wherein Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 778 is an R;

(gg) the amino acid sequence set forth as SEQ ID NO: 5, wherein Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 778 is an R;

(hh) the amino acid sequence set forth as SEQ ID NO: 6, wherein Q at amino acid position 822 is an R, G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(ii) the amino acid sequence set forth as SEQ ID NO: 7, wherein G at amino acid position 856 is an R, and E at amino acid position 778 is an R;

(jj) the amino acid sequence set forth as SEQ ID NO: 8, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(kk) the amino acid sequence set forth as SEQ ID NO: 9, wherein E at amino acid position 778 is an R, S at amino acid position 647 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(ll) the amino acid sequence set forth as SEQ ID NO: 10, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(mm) the amino acid sequence set forth as SEQ ID NO: 11, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(nn) the amino acid sequence set forth as SEQ ID NO: 12, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(oo) the amino acid sequence set forth as SEQ ID NO: 13, wherein E at amino acid position 778 is an R, Q at amino acid position 822 is an R, S at amino acid position 647 is an R, and E at amino acid position 969 is an R;

(pp) the amino acid sequence set forth as SEQ ID NO: 14, wherein E at amino acid position 778 is an R, S at amino acid position 55 is an R, and E at amino acid position 969 is an R;

(qq) the amino acid sequence set forth as SEQ ID NO: 15, wherein E at amino acid position 778 is an R, G at amino acid position 856 is an R, and E at amino acid position 969 is an R; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16, wherein E at amino acid position 778 is an R, and E at amino acid position 969 is an R.

111. The method of claim 110, wherein said RGN polypeptide comprises an amino acid sequence having 100% sequence identity to the amino acid sequence selected from:

(a) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 52 is an R;

(b) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 55 is an R;

(c) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 86 is an R;

(d) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 472 is an R;

(e) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 533 is an R;

(f) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 541 is an R;

(g) the amino acid sequence set forth as SEQ ID NO: 1, wherein Y at amino acid position 643 is substituted by an R;

(h) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 647 is an R;

(i) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 653 is an R;

(j) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 745 is an R;

(k) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 774 is an R;

(l) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 778 is an R;

(m) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 780 is an R;

(n) the amino acid sequence set forth as SEQ ID NO: 1, wherein A at amino acid position 795 is an R;

(o) the amino acid sequence set forth as SEQ ID NO: 1, wherein Q at amino acid position 822 is an R;

(p) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 843 is an R;

(q) the amino acid sequence set forth as SEQ ID NO: 1, wherein G at amino acid position 856 is an R;

(r) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 871 is an R;

(s) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 872 is an R;

(t) the amino acid sequence set forth as SEQ ID NO: 1, wherein D at amino acid position 900 is an R;

(u) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 911 is an R;

(v) the amino acid sequence set forth as SEQ ID NO: 1, wherein T at amino acid position 913 is an R;

(w) the amino acid sequence set forth as SEQ ID NO: 1, wherein N at amino acid position 954 is an R;

(x) the amino acid sequence set forth as SEQ ID NO: 1, wherein V at amino acid position 958 is an R;

(y) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 968 is an R;

(z) the amino acid sequence set forth as SEQ ID NO: 1, wherein E at amino acid position 969 is an R;

(aa) the amino acid sequence set forth as SEQ ID NO: 1, wherein K at amino acid position 973 is an R;

(bb) the amino acid sequence set forth as SEQ ID NO: 1, wherein S at amino acid position 974 is an R;

(cc) the amino acid sequence set forth as SEQ ID NO: 1, wherein L at amino acid position 975 is an R;

(dd) the amino acid sequence set forth as SEQ ID NO: 2;

(ee) the amino acid sequence set forth as SEQ ID NO: 3;

(ff) the amino acid sequence set forth as SEQ ID NO: 4;

(gg) the amino acid sequence set forth as SEQ ID NO: 5;

(hh) the amino acid sequence set forth as SEQ ID NO: 6;

(ii) the amino acid sequence set forth as SEQ ID NO: 7;

(jj) the amino acid sequence set forth as SEQ ID NO: 8;

(kk) the amino acid sequence set forth as SEQ ID NO: 9;

(ll) the amino acid sequence set forth as SEQ ID NO: 10;

(mm) the amino acid sequence set forth as SEQ ID NO: 11;

(nn) the amino acid sequence set forth as SEQ ID NO: 12;

(oo) the amino acid sequence set forth as SEQ ID NO: 13;

(pp) the amino acid sequence set forth as SEQ ID NO: 14;

(qq) the amino acid sequence set forth as SEQ ID NO: 15; and

(rr) the amino acid sequence set forth as SEQ ID NO: 16.

112. The method of any one of claims 99, and 105-111, wherein said modified target DNA sequence comprises insertion of heterologous DNA into the target DNA sequence or deletion of at least one nucleotide from the target DNA sequence.

113. The method of any one of claims 99, and 105-111, wherein said modified target DNA sequence comprises mutation of at least one nucleotide in the target DNA sequence.

114. The method of any one of claims 95-113, wherein said method is performed *in vitro*, *in vivo*, or *ex vivo*.

115. The method of any one of claims 98-114, wherein said target polynucleotide sequence is a eukaryotic target DNA sequence.

116. The method of any one of claims 98-115, wherein the target polynucleotide sequence is within a cell.

117. The method of claim 116, wherein the cell is a eukaryotic cell.

118. The method of claim 117, wherein the eukaryotic cell is a mammalian cell.

119. The method of any one of claims 116-118, further comprising culturing the cell under conditions in which the RGN polypeptide is expressed and cleaves and modifies the target polynucleotide sequence to produce a nucleic acid molecule comprising a modified target

polynucleotide sequence; and selecting a cell comprising said modified target polynucleotide sequence.

120. A cell comprising a modified target polynucleotide sequence produced according to the method of claim 119.

121. A pharmaceutical composition comprising the cell of any one of claims 30, 96, and 120, and a pharmaceutically acceptable carrier.

122. A method for treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

administering to the subject the nucleic acid molecule of any one of claims 1-22, the vector of any one of claims 23-29, the RGN polypeptide of any one of claims 32-61, the RNP complex of claim 62, the system of any one of claims 63-95, the cell of any one of claims 30, 96 and 120, or the pharmaceutical composition of claim 97 or 121.

123. The method of claim 122, wherein said disease, disorder, or condition is associated with a mutation and said treating comprises correcting said mutation.

124. Use of the nucleic acid molecule of any one of claims 1-22, the vector of any one of claims 23-29, the RGN polypeptide of any one of claims 32-61, the RNP complex of claim 62, the system of any one of claims 63-95, the cell of any one of claims 30, 96 and 120, or the pharmaceutical composition of claim 97 or 121 in the manufacture of a medicament useful for treating a disease, disorder, or condition.

125. The use of claim 124, wherein said disease, disorder, or condition is associated with a mutation and an effective amount of said medicament corrects said mutation.

126. The RNP complex of claim 62 or the system of any one of claims 63-95 for use in binding a target polynucleotide sequence of a nucleic acid molecule.

127. The RNP complex of claim 62 or the system of any one of claims 63-95 for use in cleaving and/or modifying a target polynucleotide sequence of a nucleic acid molecule.

128. The nucleic acid molecule of any one of claims 1-22, the vector of any one of claims 23-29, the RGN polypeptide of any one of claims 32-61, the RNP complex of claim 62, the system of any one of claims 63-95, the cell of any one of claims 30, 96 and 120, or the pharmaceutical composition of claim 97 or 121 for use in treating a disease, disorder, or condition.

129. A method of increasing efficiency of cleaving and/or modifying a nucleic acid molecule comprising a target sequence, the method comprising delivering the system of any one of claims 63-95 or the RNP complex of claim 62 to the target sequence or to a cell comprising the target sequence, wherein cleavage or modification of the nucleic acid molecule occurs at greater efficiency as compared to cleavage or modification of the nucleic acid molecule by a method comprising delivering to the target sequence or to a cell comprising the target sequence a reference RGN system

or RNP complex, wherein the reference RGN system or RNP complex does not comprise said RGN polypeptide.

130. The method of claim 129, wherein the efficiency of cleaving and/or modifying the target sequence is increased by at least 15%.

131. The method of claim 129 or 130, wherein the efficiency of cleaving and/or modifying the target sequence is measured by next generation sequencing, Tracking of Indels by DEcomposition (TIDE) analysis, flow cytometry, or a combination thereof.

132. A fusion polypeptide comprising:

(a) the RGN polypeptide of any one of claims 32-48; and

(b) a heterologous polypeptide.

133. The fusion polypeptide of claim 132, wherein said heterologous polypeptide is operably fused to the N-terminus, to the C-terminus, or to an internal location of said RGN polypeptide.

134. The fusion polypeptide of claim 132 or 133, wherein said heterologous polypeptide is a polymerase editing polypeptide.

135. The fusion polypeptide of claim 132 or 133, wherein the heterologous polypeptide is a base-editing polypeptide.

136. The fusion polypeptide of claim 132 or 133, wherein the heterologous polypeptide is an effector domain, a detectable label, or a purification tag.

137. The fusion polypeptide of claim 136, wherein the effector domain is a cleavage domain, a deaminase domain, or an expression modulator domain.

138. A polynucleotide comprising a nucleotide sequence encoding the fusion polypeptide of any one of claims 132-137.

139. One or more polynucleotides encoding a polymerase editor (PE) comprising a polymerase and an RNA-guided nuclease (RGN) polypeptide of any one of claims 32-48.

140. A polymerase editor (PE) comprising a polymerase and an RNA-guided nuclease (RGN) polypeptide of any one of claims 32-48.

141. The PE of claim 140, wherein said polymerase is a reverse transcriptase.

142. The PE of claim 141, wherein said reverse transcriptase has at least 90% sequence identity to the amino acid sequence of SEQ ID NO: 254 or 255.

143. The PE of any one of claims 140-142, wherein said polymerase is operably fused to the N-terminus of said RGN polypeptide.

144. The PE of any one of claims 140-142, wherein said polymerase is operably fused to the C-terminus of said RGN polypeptide.

145. The PE of any one of claims 140-142, wherein said polymerase is operably fused to an internal location of said RGN polypeptide.

146. The PE of claim 145, wherein said polymerase is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

147. The PE of any one of claims 140-146, wherein said PE comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 135, and 162-181.

148. A PE system for modifying one or more target sequences in a target polynucleotide, said system comprising:

a) one or more polymerase editing guide RNAs (PEgRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more PEgRNAs, wherein the one or more PEgRNAs comprise an extension arm, wherein the extension arm comprises a primer binding site (PBS) and a DNA synthesis template sequence; and

b) a PE of any one of claims 140-147;

wherein said one or more PEgRNAs are capable of binding to said RGN polypeptide of said PE.

149. A method for modifying a target polynucleotide comprising a target sequence, said method comprising delivering a PE system according to claim 148 to said target sequence or a cell comprising the target sequence, wherein said method generates a modified target polynucleotide, and wherein components of said PE system are delivered simultaneously or sequentially to said target sequence or said cell comprising the target sequence.

150. One or more polynucleotides encoding a base editor comprising a base editing polypeptide and an RNA-guided nuclease (RGN) polypeptide of any one of claims 32-48.

151. A base editor comprising a base editing polypeptide and an RNA-guided nuclease (RGN) polypeptide of any one of claims 32-48.

152. The base editor of claim 151, wherein said base editing polypeptide is a deaminase.

153. The base editor of claim 152, wherein the deaminase has at least 90% sequence identity to an amino acid sequence of any one of SEQ ID NOs: 42-113, and 257.

154. The base editor of claim 151, wherein said base editing polypeptide is operably fused to the N-terminus of said RGN polypeptide.

155. The base editor of claim 151, wherein said base editing polypeptide is operably fused to the C-terminus of said RGN polypeptide.

156. The base editor of claim 151, wherein said base editing polypeptide is operably fused to an internal location of said RGN polypeptide.

157. The base editor of claim 156, wherein said base editing polypeptide is operably fused within said RGN polypeptide immediately after an amino acid at a position selected from the group consisting of:

a) an amino acid position corresponding to position 666 of SEQ ID NO: 1;

b) an amino acid position corresponding to position 785 of SEQ ID NO: 1; and

c) an amino acid position corresponding to position 910 of SEQ ID NO: 1.

158. The base editor of any one of claims 151-157, wherein said base editor comprises an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 258-270.

159. A base editing system for modifying one or more target sequences in a target polynucleotide, said system comprising:

a) one or more guide RNAs (gRNAs), or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more gRNAs; and

b) a base editor of any one of claims 151-158;

wherein said one or more gRNAs is capable of binding to said RGN polypeptide of said base editor.

160. A method for modifying a target polynucleotide comprising a target sequence, said method comprising delivering a base editing system according to claim 159 to said target sequence or a cell comprising the target sequence, wherein said method generates a modified target polynucleotide, and wherein components of said base editing system are delivered simultaneously or sequentially to said target sequence or said cell comprising the target sequence.

FIG. 1

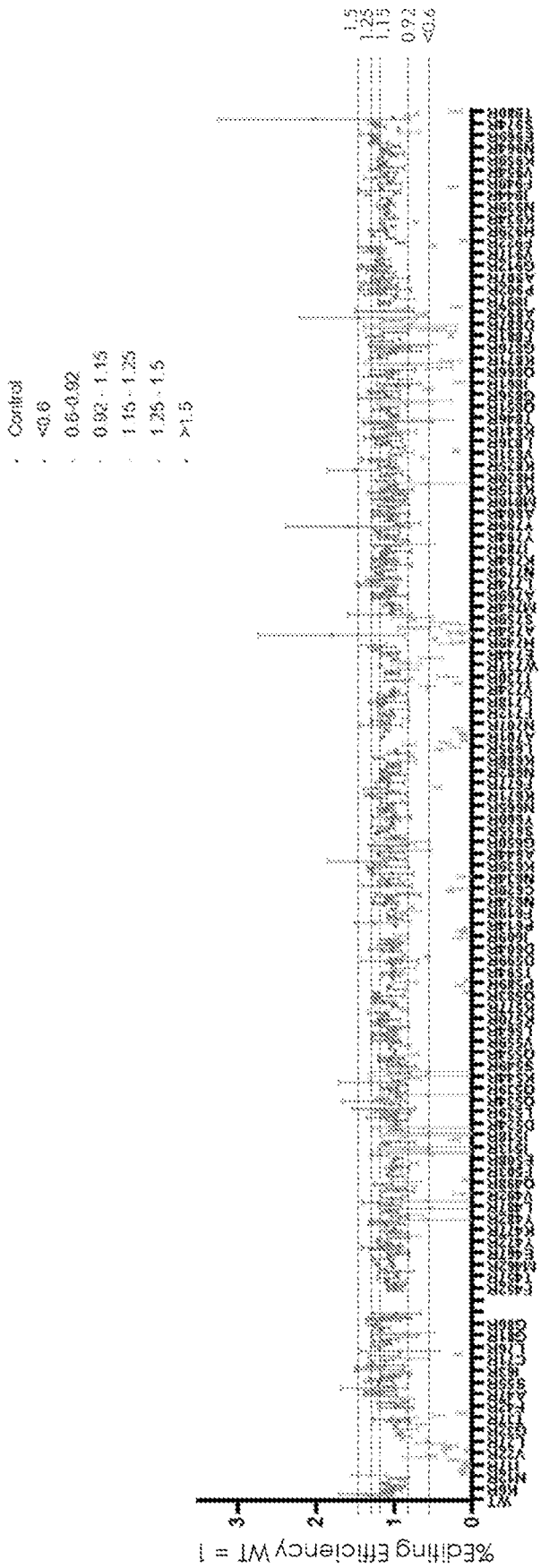

Control
<0.6
0.6-0.92
0.92 - 1.15
1.15 - 1.25
1.25 - 1.5
>1.5
1.5
1.25
1.15
0.92
<0.6
%Editing Efficiency WT = 1
0
1
2
3

FIG. 2

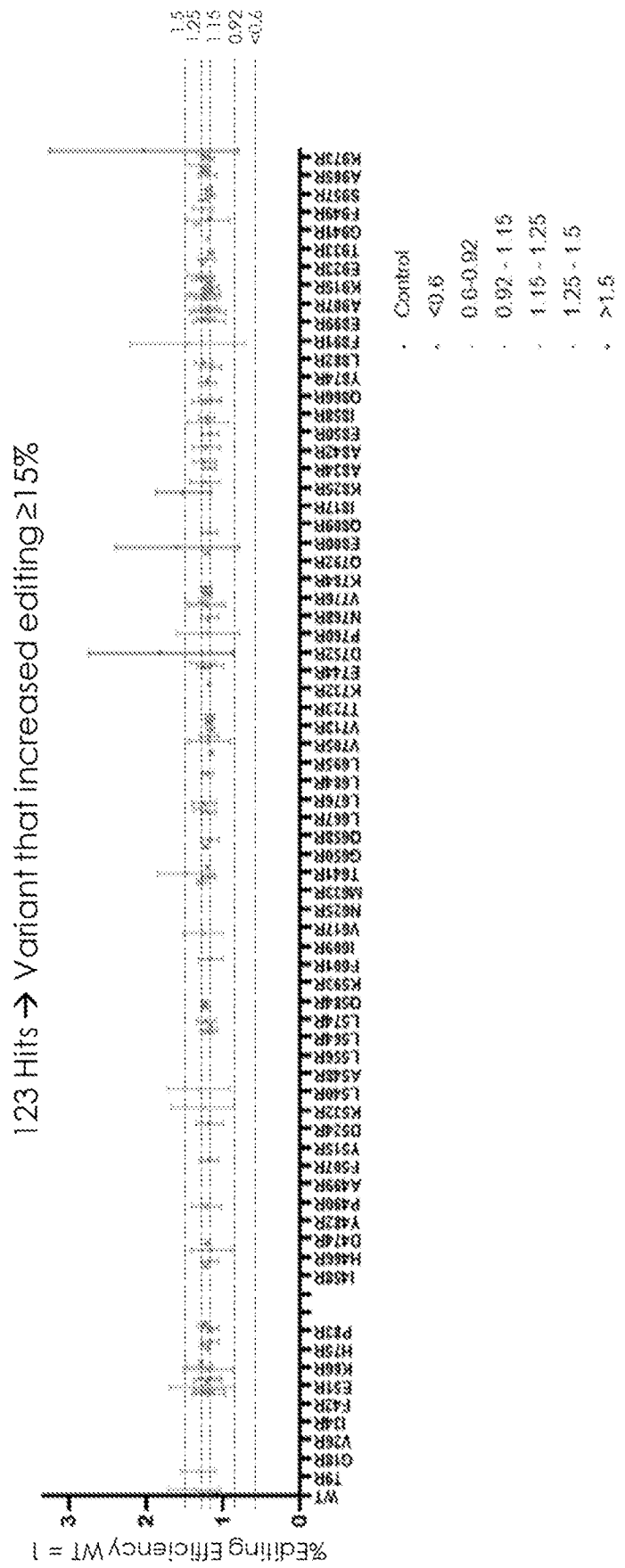

123 Hits → Variant that increased editing ≥15%
%Editing Efficiency WT = 1
0
1
2
3
1.5
1.25
1.15
0.92
<0.6
Control
<0.6
0.6-0.92
0.92 - 1.15
1.15 - 1.25
1.25 - 1.5
>1.5

FIG. 3

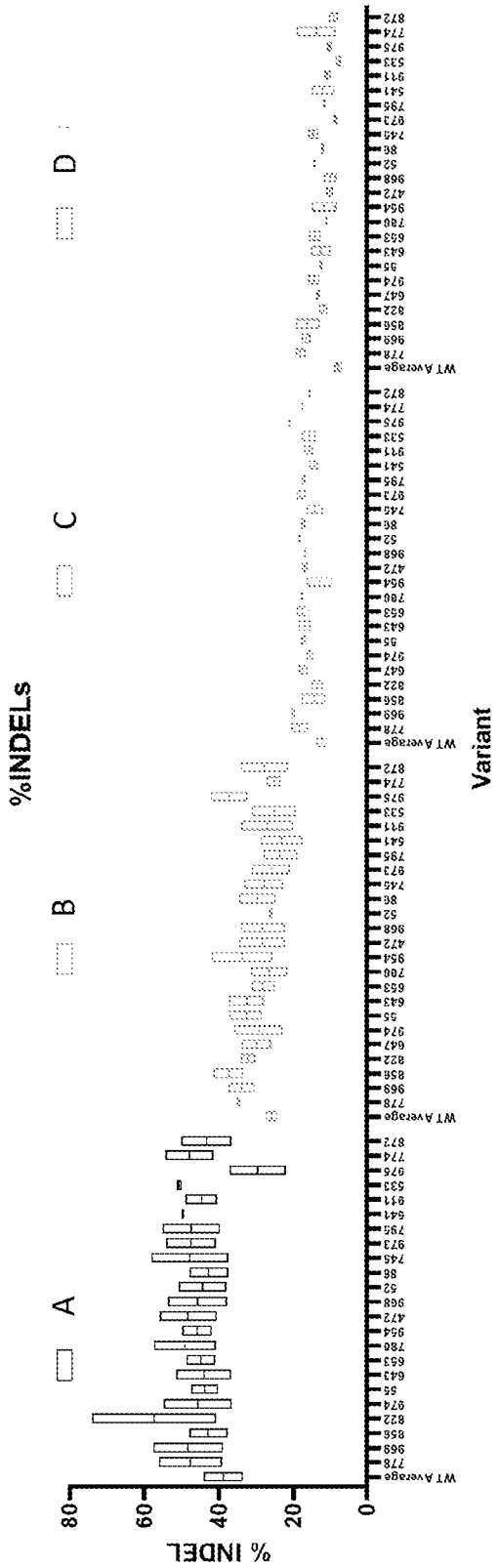


FIG. 4

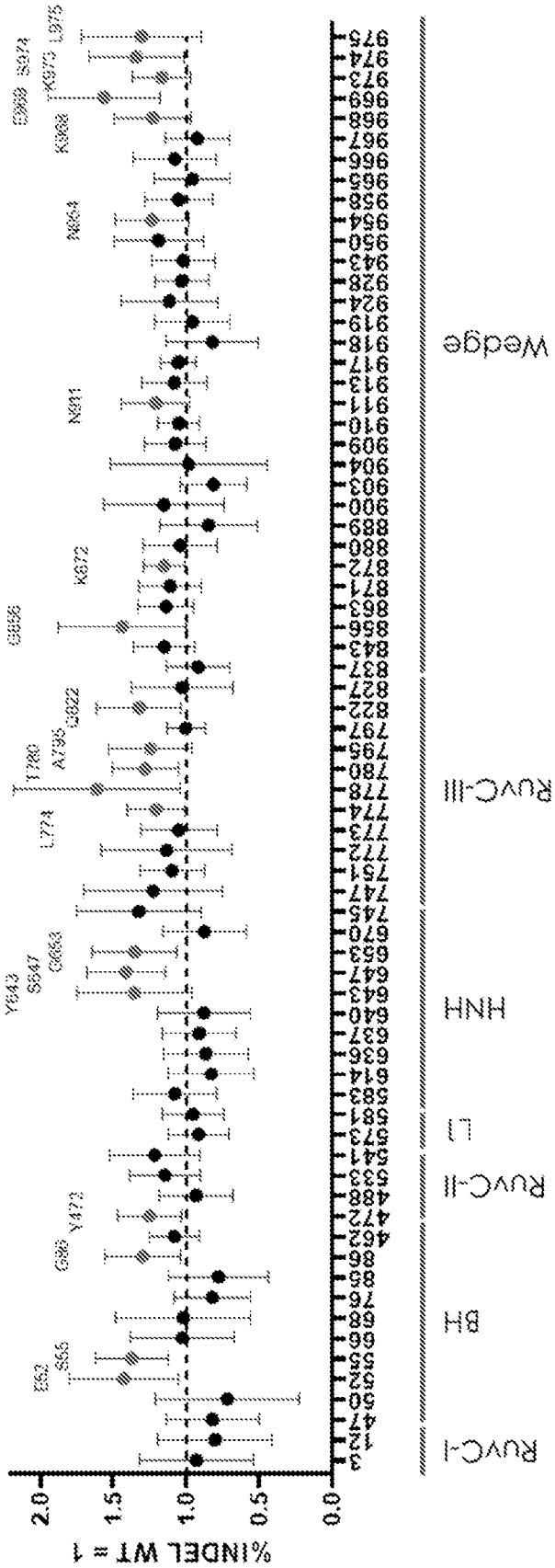

All Guides Normalized and Averaged (WT = 1)
%INDEL WT = 1
2.0
1.5
1.0
0.5
0.0
RuvC-I
BH
RuvC-II
L1
HNH
RuvC-III
Wedge

FIG. 5

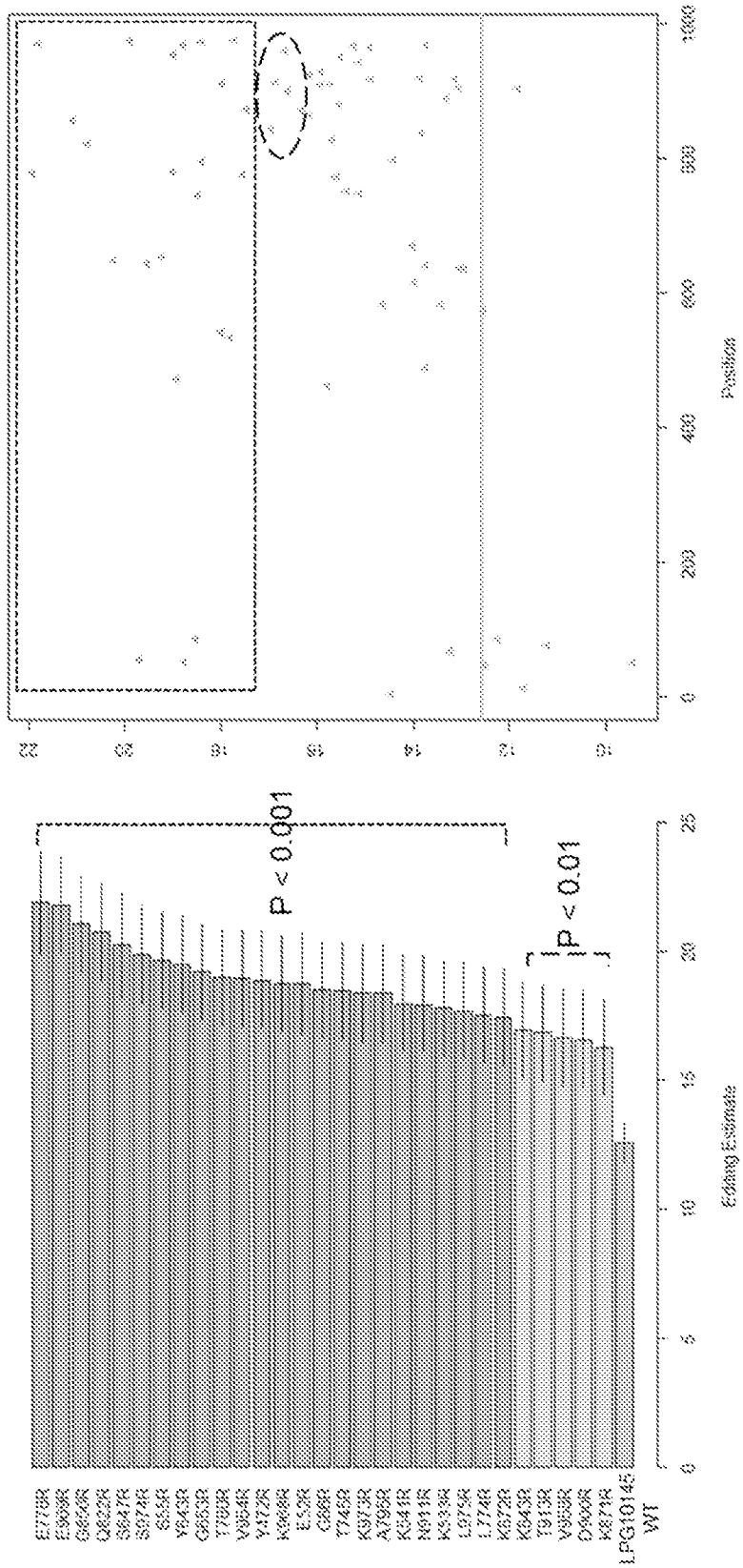

Position
Editing Estimate
P < 0.001
P < 0.01
E776R
E969R
G854R
Q822R
S647R
S974R
S55R
Y843R
G953R
T760R
V954R
Y472R
K966R
E52R
G66R
T745R
K973R
A795R
K941R
N911R
K333R
L975R
L774R
K672R
K943R
T913R
V953R
D900R
K874R
LPG10145
WT

FIG. 6

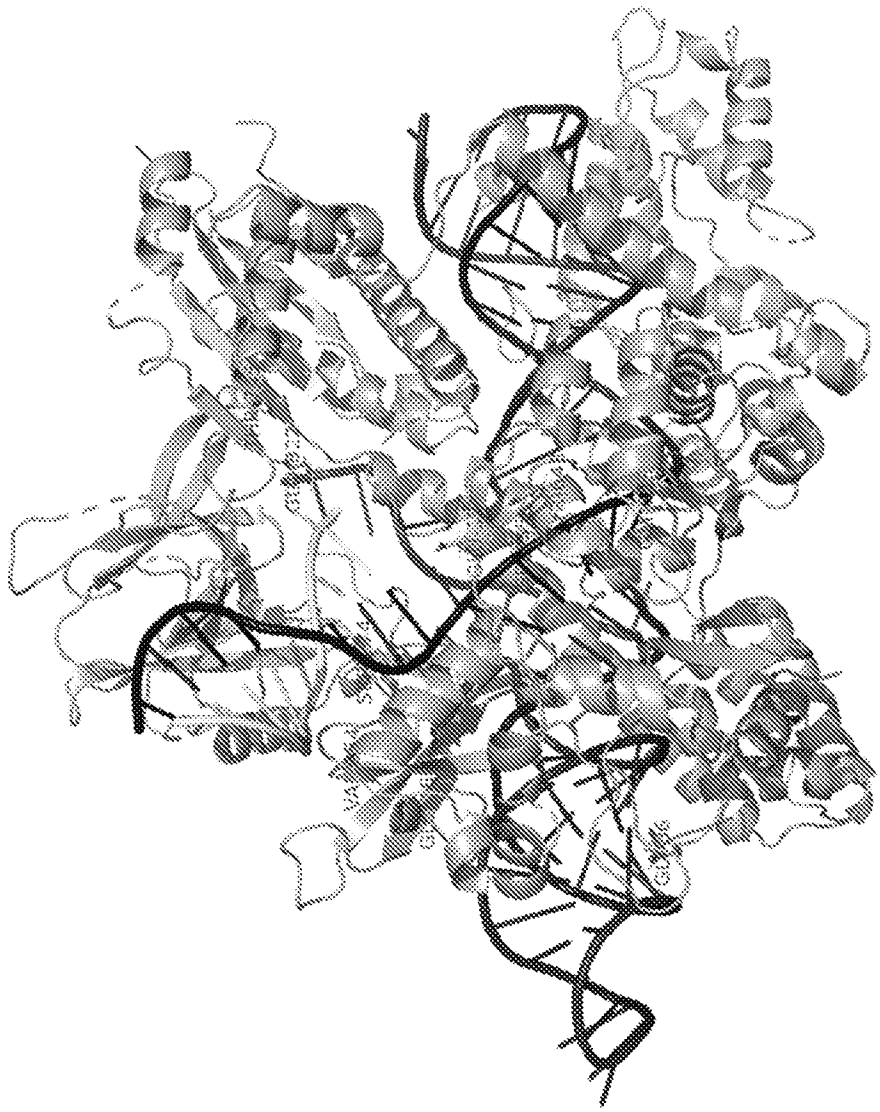


FIG. 7

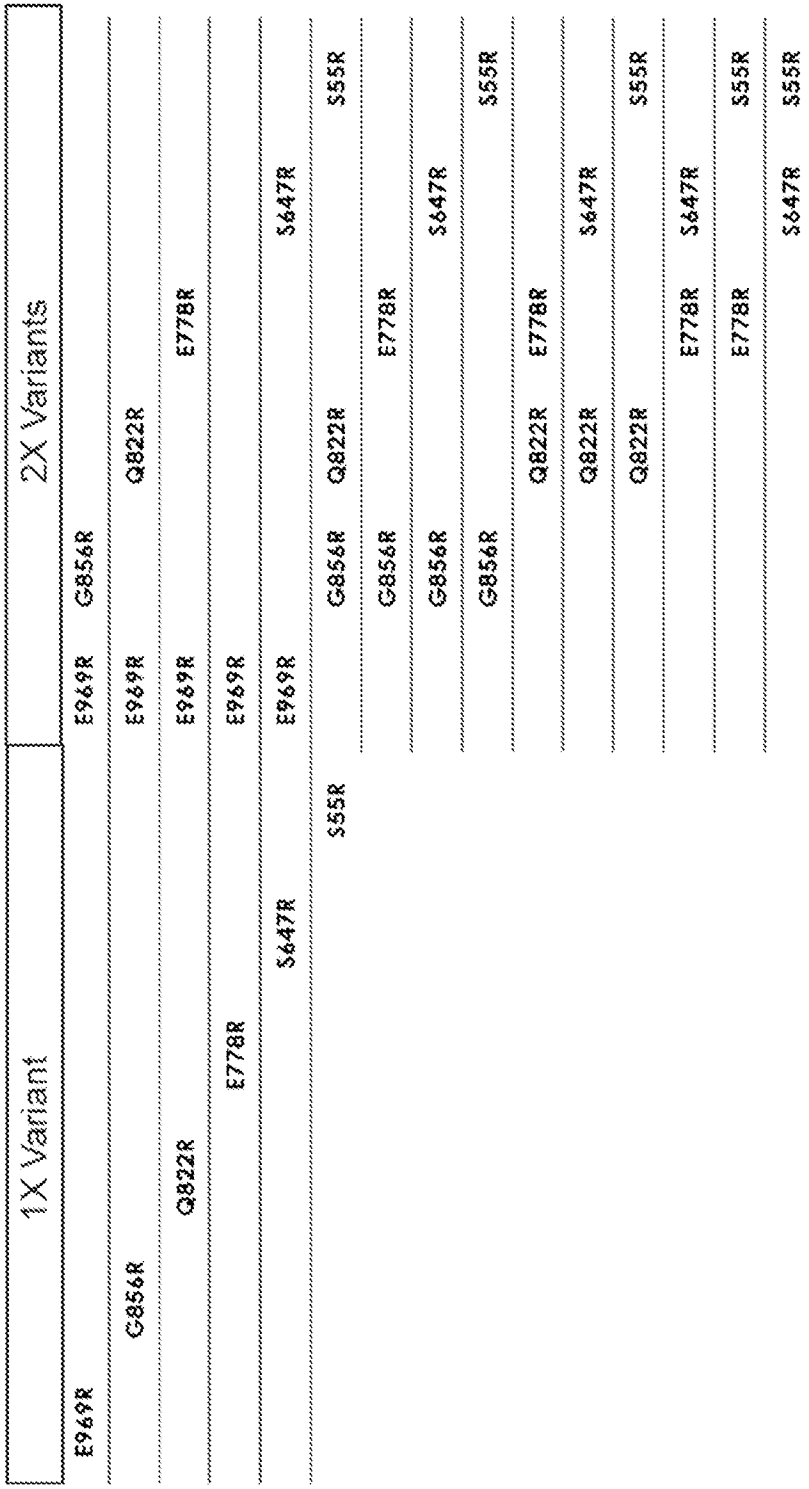

1X Variant
2X Variants
E969R
E969R G856R
G856R
E969R Q822R
Q822R
E969R E778R
E778R
E969R S647R
S647R
E969R S55R
S55R
G856R Q822R S55R
G856R E778R
G856R S647R
G856R S55R
Q822R E778R
Q822R S647R
Q822R S55R
E778R S647R
E778R S55R
S647R S55R

FIG. 7 cont'd.

3X Variants						4X Variants					
E969R	G856R	Q822R				E969R	G856R	Q822R	E778R		
E969R	G856R		E778R			E969R	G856R	Q822R		S647R	
E969R	G856R			S647R		E969R	G856R	Q822R			S55R
E969R	G856R				S55R	E969R	G856R		E778R	S647R	
E969R		Q822R	E778R			E969R	G856R		E778R		S55R
E969R		Q822R		S647R		E969R	G856R			S647R	S55R
E969R		Q822R			S55R	E969R		Q822R	E778R	S647R	
E969R			E778R	S647R		E969R		Q822R	E778R		S55R
E969R			E778R		S55R	E969R		Q822R		S647R	S55R
E969R				S647R	S55R	E969R			E778R	S647R	S55R
	G856R	Q822R	E778R				G856R	Q822R	E778R	S647R	
	G856R	Q822R		S647R			G856R	Q822R	E778R		S55R
	G856R	Q822R			S55R		G856R	Q822R		S647R	S55R
	G856R		E778R	S647R					E778R	S647R	S55R
	G856R		E778R		S55R			Q822R	E778R	S647R	S55R
	G856R			S647R	S55R						
		Q822R	E778R	S647R							
		Q822R	E778R		S55R						
		Q822R		S647R	S55R						
			E778R	S647R	S55R						

FIG. 7 cont'd.

5X Variants					
E969R	G856R	Q822R	E778R	S647R	
E969R	G856R	Q822R	E778R		S55R
E969R	G856R	Q822R		S647R	S55R
E969R	G856R		E778R	S647R	S55R
E969R		Q822R	E778R	S647R	S55R
	G856R	Q822R	E778R	S647R	S55R

6X Variant					
E969R	G856R	Q822R	E778R	S647R	S55R

FIG. 8

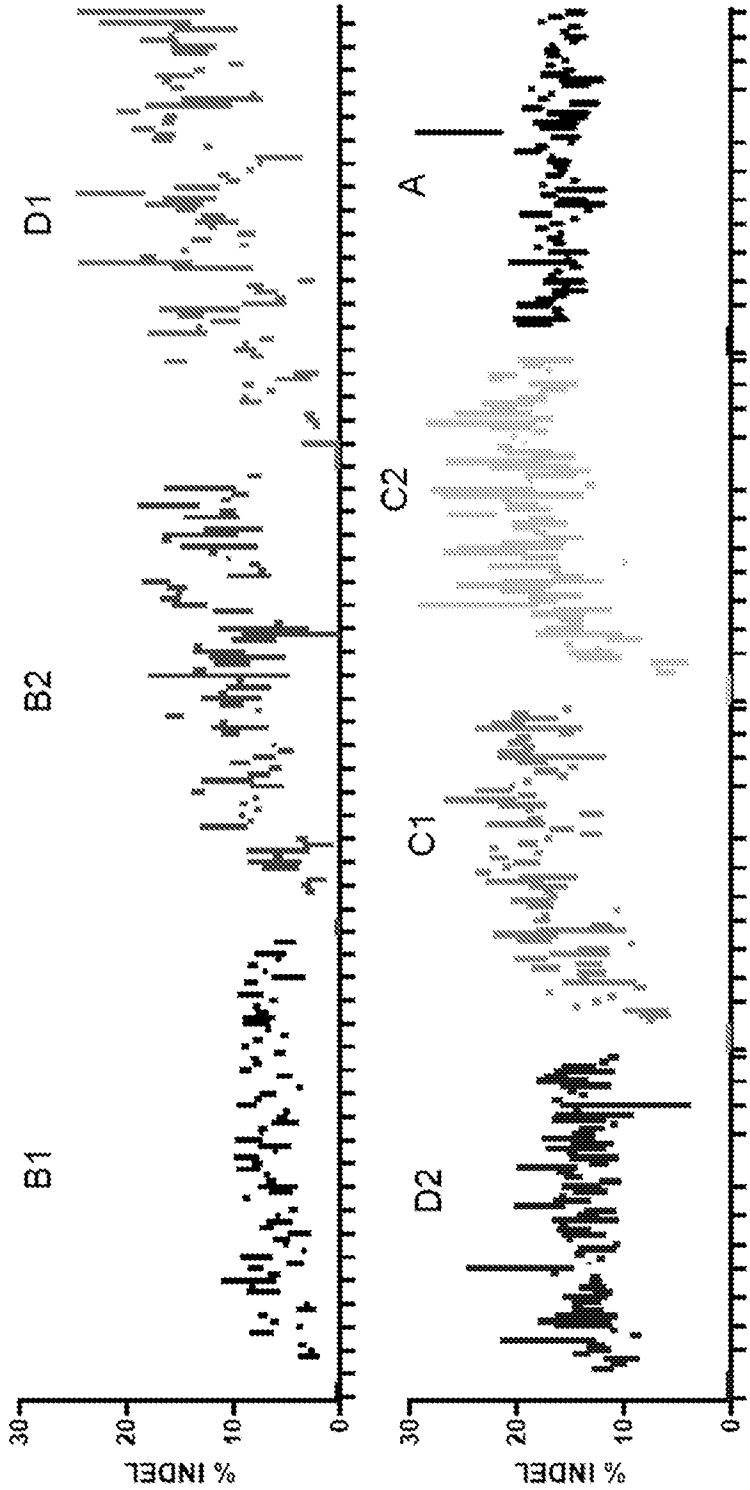


FIG. 9

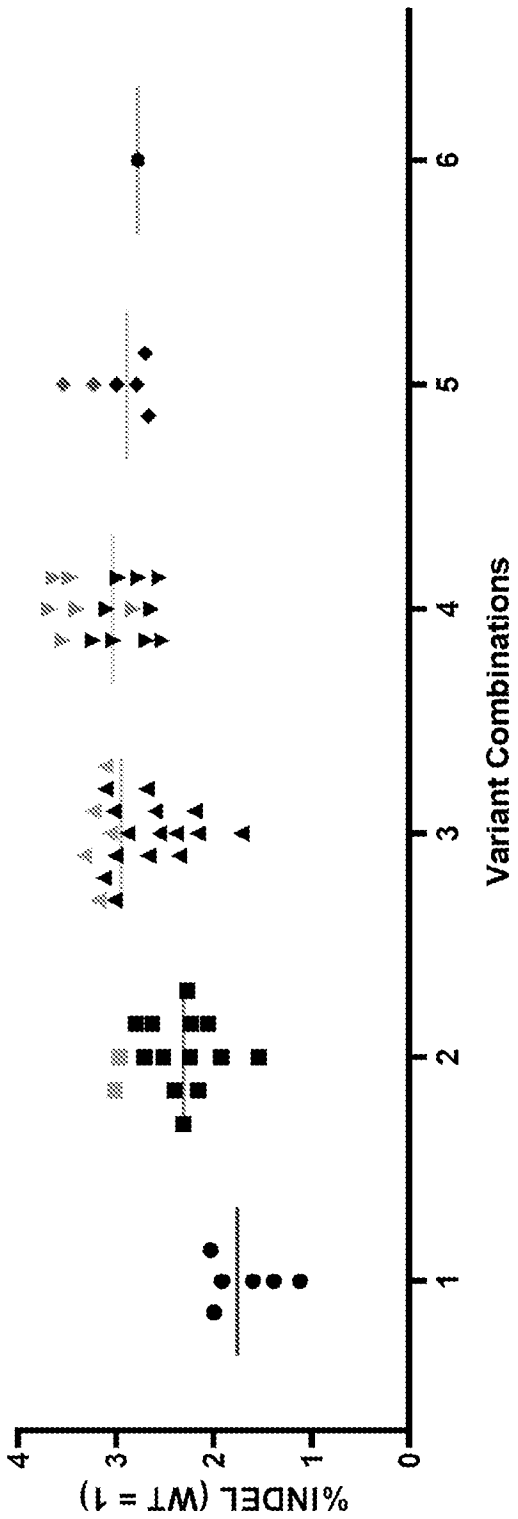
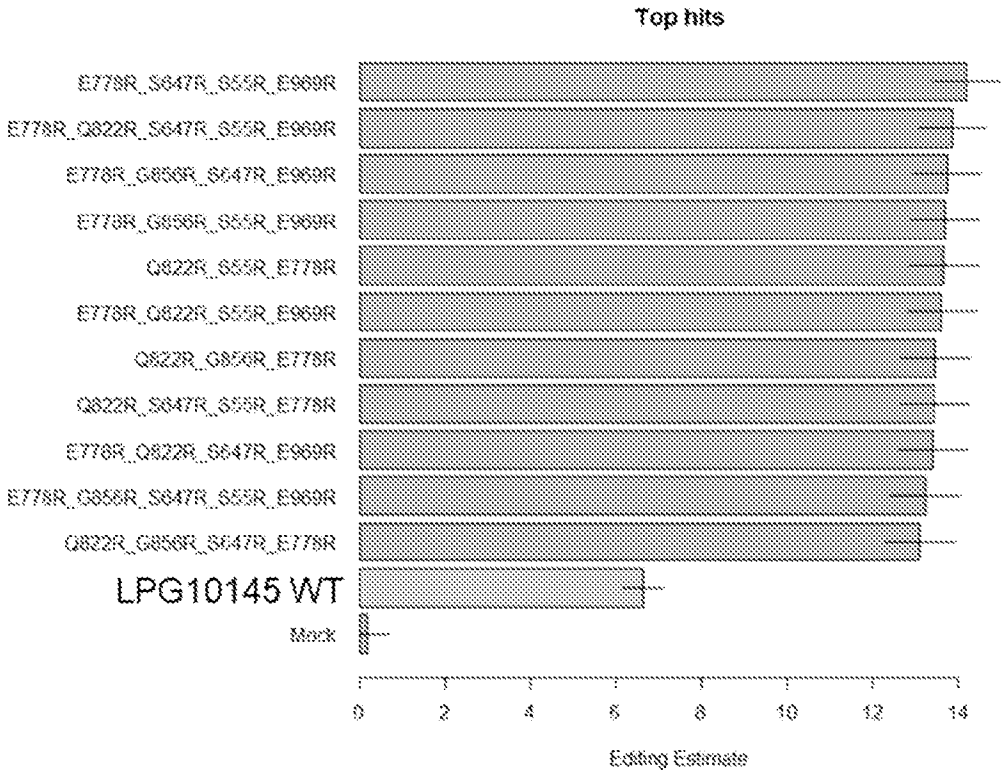
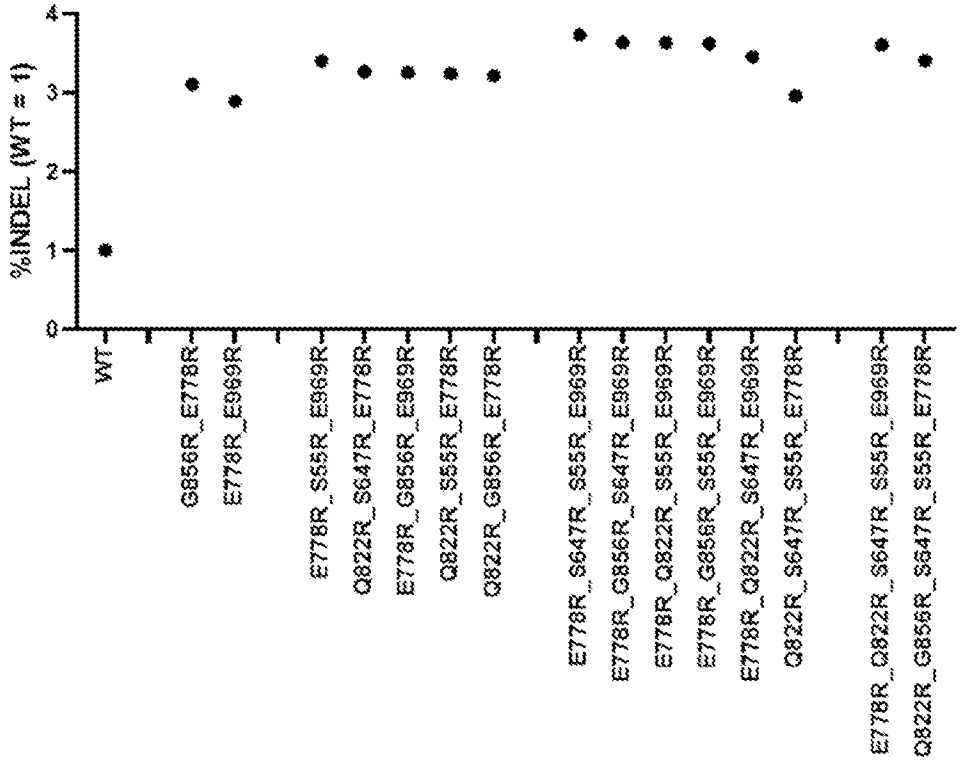


FIG. 10


Top hits
E778R_S647R_S55R_E969R
E778R_Q822R_S647R_S55R_E969R
E778R_G856R_S647R_E969R
E778R_G856R_S55R_E969R
Q822R_S55R_E778R
E778R_Q822R_S55R_E969R
Q822R_G856R_E778R
Q822R_S647R_S55R_E778R
E778R_Q822R_S647R_E969R
E778R_G856R_S647R_S55R_E969R
Q822R_G856R_S647R_E778R
LPG10145 WT
Mock
0
2
4
6
8
10
12
14
Editing Estimate


%INDEL (WT = 1)
0
1
2
3
4
WT
G856R_E778R
E778R_E969R
E778R_S55R_E969R
Q822R_S647R_E778R
E778R_G856R_E969R
Q822R_S55R_E778R
Q822R_G856R_E778R
E778R_S647R_S55R_E969R
E778R_G856R_S647R_E969R
E778R_Q822R_S55R_E969R
E778R_G856R_S55R_E969R
E778R_Q822R_S647R_E969R
Q822R_S647R_S55R_E778R
E778R_Q822R_S647R_S55R_E969R
Q822R_G856R_S647R_S55R_E778R