

ENGINEERED TYPE V RNA PROGRAMMABLE ENDONUCLEASES AND THEIR USES**1. CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims the priority benefit of U.S. provisional application no. 63/588,636, filed October 6, 2023, the contents of which are incorporated herein in their entireties by reference thereto.

2. SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML Sequence Listing, created on September 26, 2024, is named BRT-008WO_SL.xml and 451,059 bytes in size.

3. BACKGROUND

[0001] Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and the CRISPR -associated (Cas) genes, collectively known as the CRISPR-Cas or CRISPR/Cas systems, are currently understood to provide immunity to bacteria and archaea against phage infection. The CRISPR-Cas systems of prokaryotic adaptive immunity are an extremely diverse group of protein effectors, non-coding elements, as well as loci architectures, some examples of which have been engineered and adapted to produce important biotechnologies.

[0002] The components of the system involved in host defense include one or more effector proteins capable of modifying DNA or RNA and an RNA guide element that is responsible for targeting these protein activities to a specific sequence on the phage DNA or RNA. The RNA guide is composed of a CRISPR RNA (crRNA) and may require an additional trans-acting RNA (tracrRNA) to enable targeted nucleic acid manipulation by the effector protein(s). The crRNA consists of a segment termed "direct repeat", that is responsible for binding of the crRNA to the effector protein, and a segment termed "spacer sequence", that is complementary to the desired nucleic acid target sequence. CRISPR systems can be

[0003] CRISPR-Cas systems can be broadly classified into two classes: Class 1 systems are composed of multiple effector proteins that together form a complex around a crRNA, and Class 2 systems consist of a single effector protein that complexes with the crRNA guide to target DNA or RNA substrates. The single-subunit effector composition of the Class 2 systems provides a simpler component set for engineering and application and has thus far been an important source of programmable effectors. Thus, the discovery, engineering, and optimization of novel Class 2 systems may lead to widespread and powerful programmable technologies for genome engineering and beyond.

[0004] Editing genomes using the RNA-guided DNA targeting principle of CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR associated proteins), has been widely exploited in recent years. Five types of CRISPR-Cas systems (Type I, Type II and IIb, Type III, Type V, and Type VI) have been described. Most uses of CRISPR-Cas for genome editing have been with a Type II system. The main advantage provided by the bacterial Type II CRISPR-Cas system lies in the minimal requirement for programmable DNA interference: an endonuclease, Cas9, guided by a customizable dual-RNA structure. As initially demonstrated in the original Type II system of *Streptococcus pyogenes*, trans-activating CRISPR RNA (tracrRNA) binds to the invariable repeats of precursor CRISPR RNA (pre-crRNA) forming a dual-RNA that is essential for both crRNA co-maturation by RNase III in the presence of Cas9, and invading DNA cleavage by Cas9. As demonstrated in *Streptococcus pyogenes*, Cas9, guided by the duplex formed between mature activating tracrRNA and targeting crRNA, introduces site-specific double-stranded DNA (dsDNA) breaks in the invading cognate DNA. Cas9 is a multi-domain enzyme that uses an HNH nuclease domain to cleave the target strand (defined as complementary to the spacer sequence of crRNA) and a RuvC-like domain to cleave the non-target strand.

[0005] In addition to Type II CRISPR Cas 9 nucleases, a number of different Type V endonuclease nucleases have been described, such as Cas12a, Cas12b, Cas12e, Cas12f, Cas13a, Cas13b (Koonin *et al.*, Curr Opin Microbiol. 2017 Jun; 37: 67–78, and Makarova *et al.*, Nat Rev Microbiol. 2020 Feb;18(2):67-83.). Some of these systems require no tracr RNA (Cas 12a, Cas 13a, Cas 13b), whereas Cas 12b nucleases typically require a tracr RNA (Koonin *et al.*, Curr Opin Microbiol. 2017 Jun; 37: 67–78).

[0006] WO2022258753A1 discloses novel Type V endonuclease polypeptides, referred to as B-GEN.1, B-GEN.1.2 and B-GEN.2, that have particularly advantageous features for eukaryotic genome editing as compared to other endonucleases. The present disclosure provides engineered Type V endonuclease polypeptides with improved gene editing efficiency.

4. SUMMARY

[0007] The present disclosure relates to engineered Type V endonucleases with improved gene editing efficiency. Without being bound by theory, it is believed that the engineered Type V endonucleases have improved gene editing efficiency through better target interactions via their oligonucleotide binding domains (OBDs), e.g., OBD-II.

[0008] The disclosure provides engineered Type V endonucleases comprising an amino acid other than aspartic acid at the position corresponding to D504 of a Type V endonuclease of SEQ ID NO:1 (B-GEN.1) or D501 of a Type V endonuclease of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2). In some embodiments, the amino acid at the position corresponding to D504 of a Type V endonuclease of SEQ ID NO:1 (B-GEN.1) or D501 of a Type V endonuclease of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2) is arginine.

[0009] The disclosure also provides engineered Type V endonucleases comprising an OBD (e.g., an OBD-II) comprising the target-interacting sequence motif $GX_1X_2X_3X_4NX_5X_6X_7DX_8$ (SEQ ID NO:204), where each of X_1 through X_8 is any amino acid. In exemplary embodiments, the target-interacting sequence motif is any of SEQ ID NO:201, SEQ ID NO:202, and SEQ ID NO:3.

[0010] Exemplary engineered Type V endonucleases include B-GEN polypeptides. Engineered Type V endonucleases and B-GEN polypeptides include those described in Section 6.2 (and optionally comprising nuclear localization signals as described in Section 6.3 and/or linker sequences as described in Section 6.4) and in numbered embodiments 1 to 52.

[0011] The disclosure further provides engineered Type V endonuclease systems, e.g., engineered B-GEN Type V endonuclease systems comprising engineered B-GEN polypeptides and suitable guide RNAs and/or nucleic acids encoding them. Exemplary

engineered Type V endonuclease systems are disclosed in Section 6.5 and exemplary guide RNAs are disclosed in Section 6.6. In some embodiments, an engineered Type V endonuclease system is a ribonucleoprotein (RNP) complex, comprising an engineered Type V endonuclease or B-GEN polypeptide and a guide RNA. Ribonucleoprotein complexes are described in Section 6.7 and numbered embodiments 72 to 77.

[0012] The present disclosure further provides nucleic acids encoding the engineered Type V endonucleases and B-GEN polypeptides, for example expression vectors for engineered Type V endonucleases and B-GEN polypeptides, and recombinant cells engineered to express the engineered Type V endonucleases and B-GEN polypeptides. Exemplary nucleic acids are disclosed in Section 6.8 and numbered embodiments 53 to 62, exemplary recombinant cells and their use to produce engineered Type V endonucleases and B-GEN polypeptides are set forth in Section 6.10 and numbered embodiments 63 to 69, and exemplary vectors are disclosed in Section 6.9.

[0013] In certain aspects, provided herein is a method of targeting, editing, modifying, or manipulating a target DNA at one or more locations in a cell or in vitro. The methods generally entail introducing an engineered Type V endonuclease or B-GEN polypeptide system into the cell or into the in vitro environment under conditions that are suitable for the engineered Type V endonuclease or B-GEN polypeptide to make one or more nicks or cuts or base edits in the target DNA, wherein the engineered Type V endonucleases or B-GEN polypeptide is directed to the target DNA by the guide RNA in its processed or unprocessed form. As used herein, the term "Type V endonuclease or B-GEN polypeptide system" is intended to refer to any combination of nucleic acid and polypeptide components that can be delivered to a cell such that an RNP comprising the Type V endonuclease or B-GEN polypeptide is operably constituted in the cell such that editing can occur. Thus, a Type V endonuclease or B-GEN polypeptide system may include any combination of (a)(i) a Type V endonuclease or B-GEN polypeptide and/or (ii) one or more nucleic acids comprising a nucleotide sequence encoding a Type V endonuclease or B-GEN polypeptide and (b)(i) a guide RNA and/or (ii) a nucleic acid comprising a nucleotide sequence encoding a guide RNA.

[0014] In some embodiments, an RNP of the disclosure (comprising an engineered Type V endonuclease or B-GEN polypeptide and a guide RNA) is used to edit the genome of a cell.

In some embodiments, methods of genomic DNA editing using an RNP comprise nucleofection of a target cell comprising the genomic DNA with the RNP and exposing the target cell to conditions under which gene editing occurs, for example by culturing the target cell under conditions suitable for genomic editing by the engineered Type V endonuclease or B-GEN polypeptide.

[0015] In some embodiments, one or more viruses, e.g., one or more adeno-associated viruses (AAVs), are used to deliver an engineered Type V endonuclease or B-GEN polypeptide system, e.g., one or more viruses comprising one or more nucleic acids encoding an engineered Type V endonuclease or B-GEN polypeptide and one or more nucleic acids encoding a guide RNA, to a cell such that its genome can be edited. In some embodiments, methods of genomic DNA editing using one or more viruses comprise contacting a target cell comprising the genomic DNA with the one or more viruses and exposing the target cell to conditions under which gene editing occurs, for example by culturing the target cell under conditions suitable for expression of the engineered Type V endonuclease or B-GEN polypeptide and guide RNA and genomic editing by the engineered Type V endonuclease or B-GEN polypeptide as guided by the guide RNA.

[0016] In some embodiments, lipid nanoparticles (LNPs) are used to deliver an engineered Type V endonuclease or B-GEN polypeptide system, e.g., LNPs comprising one or more nucleic acids encoding an engineered Type V endonuclease or B-GEN polypeptide and a guide RNA or a nucleic acid encoding a guide RNA, to a cell such that its genome can be edited. In some embodiments, methods of genomic DNA editing using one or more viruses comprise contacting a target cell comprising the genomic DNA with the lipid nanoparticle and exposing the target cell to conditions under which gene editing occurs, for example by culturing the target cell under conditions suitable for expression of the engineered Type V endonuclease or B-GEN polypeptide and optionally the guide RNA and genomic editing by the engineered Type V endonuclease or B-GEN polypeptide as guided by the guide RNA.

[0017] Exemplary methods of editing cellular genomes using the engineered Type V endonucleases and B-GEN polypeptides are set forth in Section 6.13 and numbered embodiments 78 to 97.

[0018] Editing cellular genomes can be carried out in vitro (e.g., in cell culture), ex vivo, or in vivo (e.g., via administering an RNP, AAV or LNP to a subject for the purpose of gene therapy).

[0019] Exemplary cells comprising the engineered Type V endonucleases and B-GEN polypeptides and nucleic acids (e.g., cells contacted with RNPs, AAVs, or LNPs as disclosed herein) are set forth in Section 6.11 and numbered embodiments 98 to 110.

[0020] Additional features, advantages, and applications of the engineered Type V endonucleases and B-GEN polypeptides of the disclosure are more particularly described below.

5. BRIEF DESCRIPTION OF THE FIGURES

[0021] FIG. 1 shows the amino acid sequence alignment of three B-GEN enzymes: B-GEN.1 (SEQ ID NO:1), B-GEN.1.2 (SEQ ID NO:2), and B-GEN.2 (SEQ ID NO:3), together with an exemplary consensus sequence (SEQ ID NO:205). Residues of each B-GEN enzyme that differ from corresponding amino acids of the other two B-GEN enzymes are shown within dark boxes, whereas matching amino acids are unmarked.

[0022] FIGS. 2A-2C show phylogenetic assessments of endonucleases derived from various species in the order Bacillales. FIG. 2A is a phylogenetic tree displaying genera derived from genus *Bacillus*, adapted from Suzuki, 2018, Appl Microbiol Biotechnol. 102:10425-10437. FIG. 2B displays the phylogenetic origins of several endonucleases. FIG. 2C is a phylogenetic tree displaying endonucleases of genera derived from genus *Bacillus*, order Bacillales, using the tree creation algorithm of Geneious Prime software, based on NCBI blast output of the B-GEN.2 OBD II domain and the RuvC I domain sequences.

[0023] FIGS. 3A-3C display the AacC2c1 and B-GEN.2 enzyme structures. FIG. 3A is the AacC2c1 crystal structure (5U33) visualized using DNASTar. FIG. 3B shows the predicted structure of B-GEN.2 generated using AlphaFold2 and visualized using DNASTar. FIG. 3C shows the alignment of the crystal structure of AacC2c1 (5U33) with the predicted structure of B-GEN.2.

[0024] FIG. 4 shows the amino acid sequence alignment of B-GEN.1 and B-GEN.2 with 20 different Cas nucleases of the order Bacillales (created using the MUSCLE alignment algorithm of Geneious Prime software. Domains are marked on the Bth C2C1 sequence as

follows: Bolded amino acids represent oligonucleotide binding domains (OBD domains) (OBD-I and OBD-II), single continuous underlined amino acids represent recognition (REC) domains (REC1-I, REC1-II, and REC2), italicized amino acids represent the PAM-interacting (PI) domain, and double continuous underlined amino acids represent RuvC domains (RuvC-I, RuvC-II, RuvC-III) which together form the nuclease domain. The plus sign above the consensus sequence indicates the amino acid that corresponds to the D501 residue of B-GEN.2. Star signs above residues mark the active site catalytic residues within RuvC domains.

[0025] FIG. 5 is a Coomassie-stained image of and SDS-PAGE gel showing the results of a single-step heparin sulfate-based purification of B-GEN.2 D501R using the method described in Section 8.1.3.

[0026] FIG. 6 is a graph showing the B2M target site cutting efficiency of B-GEN.2 D501R relative to B-GEN.2 WT and Cpf1 in in vitro plasmid cutting assays for different nuclease:linearized plasmid ratios.

[0027] FIG. 7 is a bar graph showing the results of two assessments of B2M gene editing in iPSCs, using 50 pmol RNP comprising B-GEN.2 WT, B-GEN.2 D501R, or Cpf1. Percent indel formation on the Y-axis represents percent gene editing by the RNP formed with the indicated enzyme.

[0028] FIG. 8 shows superimposed structures of AacC2c1 (labeled on the image as 5U31) and B-GEN.2, generated using AlphaFold2. The aligned amino acid backbones displaying the target-interacting strands of the superimposed endonucleases are shown for both AacC2c1 (black residues) and B-GEN.2 (white residues), adjacent to the P-1 nucleotide of the target DNA, wherein the target-interacting strands of both endonucleases are located near the PAM recognition region of each enzyme and comprises the amino acid corresponding to the R507 of AacC2c1 (side chain shown hydrogen-bonding to one of the DNA the phosphate groups) and D501 of B-GEN.2 (in gray, farther from the target DNA strand). The amino acid sequence of the target-interacting strand of B-GEN.1 with the D504R substitution corresponds to SEQ ID NO:201, the amino acid sequence of the target-interacting strand of B-GEN.1.2 with the D501R substitution corresponds to SEQ ID NO:202, and amino acid sequence of the target-interacting strand of B-GEN.2 with the D501R substitution corresponds to SEQ ID NO:203. A consensus amino acid sequence for the

target-interacting strand among the endonucleases set forth in FIG. 4, and with the amino acid arginine at the position corresponding D501 of SEQ ID NO:3, is provided as SEQ ID NO:204.

6. DETAILED DESCRIPTION

6.1. Definitions

[0029] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present disclosure. In case of conflict, the present specification, including definitions, will control. Generally, nomenclature used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics, analytical chemistry, synthetic organic chemistry, medicinal and pharmaceutical chemistry, and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Throughout this specification and embodiments, the words "have" and "comprise," or variations such as "has," "having," "comprises," or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers. All publications and other references mentioned herein are incorporated by reference in their entirety. Although a number of documents are cited herein, this citation does not constitute an admission that any of these documents form part of the common general knowledge in the art.

[0030] **B-GEN polypeptide:** As used herein, the term "B-GEN polypeptide" refers to a polypeptide comprising a nuclease domain having an amino acid sequence related to or derived from a *Brevibacillus* Type V endonuclease, such as B-GEN.1 (SEQ ID NO:1), B-GEN.1.2 (SEQ ID NO:2), or B-GEN.2 (SEQ ID NO:3)). B-GEN.1 (SEQ ID NO:1) has a nuclease domain comprising RuvC I, RuvC II, and RuvC III sub-domains corresponding to SEQ ID NOS:8-10. B-GEN.1.2 (SEQ ID NO:2) has a nuclease domain comprising RuvC I,

RuvC II, and RuvC III sub-domains corresponding to SEQ ID NO:11-13. B-GEN.2 (SEQ ID NO:3) has a nuclease domain comprising RuvC I, RuvC II, and RuvC III sub-domains corresponding to SEQ ID NO:11-13. The term "B-GEN polypeptide" encompasses a polypeptide comprising amino acid sequences having at least 40% sequence identity to the RuvC I, RuvC II and RuvC III domains (individually or collectively) of any one of B-GEN.1, B-GEN.1.2 and B-GEN.2. "B-GEN polypeptide" also encompasses a variant of any of B-GEN.1 (SEQ ID NO:1), B-GEN.1.2 (SEQ ID NO:2), B-GEN.2 (SEQ ID NO:3), such as a variant comprising an amino acid sequence having at least 50% sequence identity SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 and/or (2) a variant comprising an amino acid sequence that differs from SEQ ID NO:4, SEQ ID NO:5 or SEQ ID NO:6 by up to 25 amino acids. In some embodiments, the B-GEN polypeptide has nuclease activity. The term "B-GEN polypeptide" encompasses engineered fusion polypeptides that includes the amino acid sequence of B-GEN.1 (SEQ ID NO:1), B-GEN.1.2 (SEQ ID NO:2), B-GEN.2 (SEQ ID NO:3), or any variant thereof as described in Section 6.2, for example an amino acid sequence (1) having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% sequence identity, at least 99% sequence identity, or 100% sequence identity to the nuclease domain or entire length of any of SEQ ID NOS:1 to 3 and/or (2) that differs from any one of SEQ ID NOS:1 to 3 by up to 25 amino acids, by up to 20 amino acids, by up to 15 amino acids, by up to 14 amino acids, by up to 13 amino acids, by up to 12 amino acids, by up to 11 amino acids, by up to 10 amino acids, by up to 9 amino acids, by up to 8 amino acids, by up to 7 amino acids, by up to 6 amino acids, or by up to 5 amino acids, together with additional sequences (e.g., one or more nuclear localization and/or linker sequences as described in Section 6.3 and/or Section 6.4). The term "B-GEN polypeptide" encompasses variants of in which the amino acid at the position corresponding to D504 of B-GEN.1 (SEQ ID NO:1) or D501 of B-GEN.1.2 (SEQ ID NO:2) or D501 of B-GEN.2 (SEQ ID NO:3) is not an aspartic acid, for example is an arginine.

[0031] Binding: As used herein, the term "binding" (e.g., with reference to an RNA-binding domain of a polypeptide) refers to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e.g., when a molecule X is said to interact with a molecule Y, it is meant the molecule X binds to molecule Y in a non-covalent manner). Not all components of a binding interaction need be sequence-

specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific. Binding interactions are generally characterized by a dissociation constant (K_d) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. "Affinity" refers to the strength of binding, increased binding affinity being correlated with a lower K_d .

[0032] Cell Therapy: As used herein, the term "cell therapy" refers to a therapy in which cellular material is administered to a patient. The cellular material may be intact, living cells. For example, T cells capable of fighting cancer cells via cell-mediated immunity may be administered in the course of immunotherapy. Cell therapy is also called cellular therapy or cytotherapy.

[0033] Coding Sequence: As used herein, the term "coding sequence" or "encoding nucleic acid" refers to a sequence within a nucleic acid (RNA or DNA) molecule that encodes a protein or RNA molecule. The coding sequence can further include initiation and termination signals operably linked to regulatory elements including a promoter and polyadenylation signal capable of directing expression in the cells of an individual or mammal to which the nucleic acid is introduced or administered. The coding sequence may be codon optimized for expression in a cell of interest.

[0034] Complement: As used herein, the terms "complement" and "complementary" in the context of a nucleic acid molecule refer to the ability to form Watson-Crick (e.g., A-T/U and C-G) or Hoogsteen base pairing between nucleotides or nucleotide analogs of nucleic acid molecules. "Complementarity" refers to a property shared between two nucleic acid sequences, such that when they are aligned antiparallel to each other, the nucleotide bases at each position will be complementary.

[0035] Corresponding (or Corresponds) to: The terms "corresponding to" or "corresponds to" as used in relation to a position in a reference sequence, e.g., the amino acid sequence of a B-GE_n polypeptide of any one of SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3, or the RuvC I, RuvC II, and RuvC III sub-domains of the nuclease domain thereof of any one of SEQ ID NO:8 or SEQ ID NO:11 (RuvC I), SEQ ID NO:9 or SEQ ID NO:12 (RuvC II), and SEQ ID NO:10 or SEQ ID NO:13 (RuvC II), is the sequence position in a query sequence that appears at the same position in an alignment of the reference and query sequence, for

example as shown in FIG. 1. Sequence alignment algorithms, such as, for example, Clustal Omega (ClustalW; found at www.ebi.ac.uk/Tools/msa/clustalo/), can be used to align a reference and query sequence, using the software's default parameters as of October 1, 2023.

[0036] Electroporation: As used herein, the term “electroporation” refers to a transfection technique, wherein an electrical pulse is used to create temporary pores in cell membranes to allow introduction of ribonucleoproteins or nucleic acid molecules, such as DNA or RNA (e.g., mRNA) molecules, into cells.

[0037] Encoding: The term “encoding” in relation to a nucleic acid (DNA or RNA) means that the nucleic acid comprises a nucleotide sequence coding for the amino acids of a polypeptide or the nucleotides of an RNA.

[0038] Engineered B-Gen polypeptide: As used herein, the term “engineered B-Gen polypeptide” refers to a variant polypeptide comprising at least one mutation (e.g., amino acid insertion, deletion, or substitution) as compared to the amino acid sequence of B-Gen.1 (SEQ ID NO:1), B-Gen.1.2 (SEQ ID NO:2), and B-Gen.2 (SEQ ID NO:3). An engineered B-Gen polypeptide of the disclosure encompasses a Type V endonuclease comprising an amino acid other than aspartic acid at the position corresponding to D504 of B-Gen.1 (SEQ ID NO:1), B-Gen.1.2 (SEQ ID NO:2), and B-Gen.2 (SEQ ID NO:3). In some embodiments, the amino acid at position 501 is an arginine.

[0039] Expression Cassette: As used herein, the term “expression cassette” refers to a DNA coding sequence operably linked to a promoter.

[0040] Guide RNA: As used herein, the term “guide RNA” refers to a ribonucleic acid having a DNA-targeting sequence (also referred to as “spacer” or “DNA-targeting segment”) and a protein-binding sequence (also referred to as “protein-binding segment”). The DNA-targeting sequence has sufficient complementarity with a target DNA (e.g., genomic DNA) sequence, to hybridize with the target DNA sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target DNA sequence. The DNA-targeting sequence generally includes the “protospacer-like” sequence described herein. The protein-binding sequence interacts with a site-specific modifying enzyme (e.g., a B-Gen polypeptide as described in Section 6.2 below). Site-specific cleavage of the target DNA occurs at locations determined by both (i) base pairing complementarity between the guide RNA and

the target DNA; and (ii) a short motif (referred to as the protospacer adjacent motif (PAM)) in the target DNA. The protein-binding segment of a guide RNA includes, in part, two complementary stretches of nucleotides that hybridize with one another to form a double stranded RNA duplex (dsRNA duplex). In some embodiments, a guide RNA is a single-stranded guide RNA (sgRNA).

[0041] A guide RNA and a site-specific modifying enzyme such as a B-GEN polypeptide may form a ribonucleoprotein complex (*e.g.*, bind via non-covalent interactions). The guide RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-specific modifying enzyme of the complex provides the endonuclease activity. In other words, the site-specific modifying enzyme is guided to a target DNA sequence (*e.g.*, a target sequence in a chromosomal nucleic acid; a target sequence in an extrachromosomal nucleic acid, *e.g.*, an episomal nucleic acid, a minicircle, *etc.*; a target sequence in a mitochondrial nucleic acid; a target sequence in a chloroplast nucleic acid; a target sequence in a plasmid; *etc.*) by virtue of its association with the protein-binding segment of the guide RNA.

[0042] Heterologous: As used herein, the term “heterologous” refers to a nucleotide or peptide that is not found in the native nucleic acid or polypeptide, respectively. A B-GEN.1, or B-GEN.1.2, or B-GEN.2 fusion protein described herein may in some embodiments comprise the RNA-binding domain of the B-GEN.1, or B-GEN.1.2, or B-GEN.2 polypeptide (or a variant thereof) fused to a heterologous polypeptide sequence (*e.g.*, a polypeptide sequence from a protein other than B-GEN.1 or B-GEN.2). The heterologous polypeptide may exhibit an activity (*e.g.*, enzymatic activity) that will also be exhibited by the B-GEN.1, or B-GEN.1.2, or B-GEN.2 fusion protein (*e.g.*, methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, *etc.*). A heterologous nucleic acid may be linked to a naturally occurring nucleic acid (or a variant thereof) (*e.g.*, by genetic engineering) to generate a fusion nucleic acid encoding a fusion polypeptide. As another example, in a fusion variant B-GEN.1, or B-GEN.1.2, or B-GEN.2 polypeptide, a variant B-GEN.1, or B-GEN.1.2, or B-GEN.2 polypeptide may be fused to a heterologous polypeptide (*e.g.*, a polypeptide other than B-GEN.1 or B-GEN.2), which exhibits an activity that will also be exhibited by the fusion variant B-GEN.1, or B-GEN.1.2, or B-GEN.2 polypeptide. A heterologous nucleic acid may be linked to a variant B-GEN.1, or B-GEN.1.2, or B-GEN.2

polypeptide (e.g., by genetic engineering) to generate a nucleic acid encoding a fusion variant B-GEN.1, or B-GEN.1.2, or B-GEN.2 polypeptide. "Heterologous," as used herein, additionally means a nucleotide or polypeptide in a cell that is not its native cell.

[0043] Host cell: As used herein, the terms "host cell" and "recombinant host cell" refer to a cell that has been genetically engineered, e.g., through introduction of a heterologous polypeptide or nucleic acid such as a vector or system of the disclosure. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. In some embodiments, a host cell carries a vector of the disclosure as an extrachromosomal heterologous expression vector. In some embodiments, a host cell comprises any one of the engineered B-GEN polypeptides disclosed herein, e.g., as introduced as an RNP complex. In other embodiments, a host cell has undergone gene editing by an engineered B-GEN polypeptide of the disclosure.

[0044] iPSC: As used herein, the terms "induced pluripotent stem cell" and "iPSC" refer to a type of pluripotent stem cell artificially prepared from a non-pluripotent cell, such as an adult somatic cell, partially differentiated cell or terminally differentiated cell, such as a fibroblast, a cell of hematopoietic lineage, a myocyte, a neuron, an epidermal cell, or the like, by introducing or contacting the cell with one or more reprogramming factors. iPSCs can be derived from multiple different cell types, including terminally differentiated cells. iPSCs have an embryonic stem (ES) cell-like morphology, growing as flat colonies with large nucleocytoplasmic ratios, defined borders and prominent nuclei. In addition, iPSCs express one or more key pluripotency markers known by one of ordinary skill in the art, including but not limited to Alkaline Phosphatase, SSEA3, SSEA4, Sox2, Oct3/4, Nanog, TRA160, TRA181, TDGF 1, Dnmt3b, FoxO3, GDF3, Cyp26a1, TERT, and zfp42.

[0045] Examples of methods of generating and characterizing iPSCs may be found in, for example, US Patent Publication Nos. US20090047263, US20090068742, US20090191159, US20090227032, US20090246875, and US20090304646 and PCT patent publications WO2013177133 and WO2022204567, the disclosures of each of which are incorporated herein by reference. Generally, to generate iPSCs, somatic cells are provided with reprogramming factors (e.g., Oct4, SOX2, KLF4, MYC, Nanog, Lin28, etc.) known in the art to reprogram the somatic cells to become pluripotent stem cells.

[0046] Nucleic Acid: As used herein, the terms “nucleic acid,” “oligonucleotide” and “nucleic acid” refer to at least two nucleotides covalently linked together. Nucleic acids may be single stranded or double stranded or may contain portions of both double stranded and single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA, or a hybrid, where the nucleic acid may contain combinations of deoxyribo- and ribo-nucleotides, and combinations of bases including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine and isoguanine. Nucleic acids may be obtained by chemical synthesis methods or by recombinant methods. The depiction of a single strand also defines the sequence of the complementary strand. Thus, reference to a single stranded nucleic acid herein also encompasses the complementary strand of a depicted single strand.

[0047] Nuclear Localization Signal: As used herein, the terms “nuclear localization signal” and “NLS” refer to an amino acid sequence that can facilitate the localization of a polypeptide to the nucleus of a eukaryotic cell.

[0048] Nuclease: As used herein, the terms “nuclease” and “endonuclease” are used interchangeably herein to mean an enzyme which possesses endonucleolytic catalytic activity for nucleic acid cleavage, as well as nuclease-inactivated variants thereof.

[0049] Nuclease Domain: As used herein, the terms “nuclease domain” and “cleavage domain” or “active domain” of a nuclease refer to the amino acid sequence or domain(s) within the nuclease which possesses the catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides. A single nuclease domain may consist of more than one isolated stretch of amino acids within a given polypeptide. In some embodiments, the boundaries of a nuclease domain of a B-GEN polypeptide are determined by aligning the B-GEN polypeptide with BthCas12b (Wu et al., 2017, Cell Research 27:705-708) and identifying the amino acids that align with the BthCas12b RuvC nuclease domain comprising RuvC I, RuvC II, and RuvC III sub-domains. A RuvC I domain of B-GEN.1 is set forth as SEQ ID NO:8, and a RuvC I domain of B-GEN.1.2 and B-GEN.2 is set forth as SEQ ID NO:11. A RuvC II domain of B-GEN.1 is set forth as SEQ ID NO:9, and a RuvC II domain of B-GEN.1.2 and B-GEN.2 is set forth as SEQ ID NO:12. A RuvC III domain of B-GEN.1 is

set forth as SEQ ID NO:10, and a RuvC III domain of B-GEN.1.2 and B-GEN.2 is set forth as SEQ ID NO:13.

[0050] Nucleofection: As used herein, the term “nucleofection” refers to an electroporation-based transfection method, which uses a combination of electrical parameters and cell-type specific reagents to transfer nucleic acids, such as DNA or RNA, and RNPs directly to the nuclei of target cells.

[0051] Operably linked: As used herein, the term “operably linked” refers to a functional relationship between two or more peptide or polypeptide domains or nucleic acid (e.g., DNA) segments. In the context of transcriptional regulation, the term refers to the functional relationship of a transcriptional regulatory sequence to a transcribed sequence. For example, a promoter or enhancer sequence is operably linked to a coding sequence if it stimulates or modulates the transcription of the coding sequence in an appropriate host cell or other expression system.

[0052] Percent (%) sequence identity: As used herein, the terms “percent sequence identity” “% sequence identity,” and the like in relationship to two amino acid sequences refer to the percentage of sequence identity as determined using the BLASTP algorithm (Tatusova and Madden, 1999, FEMS Microbiol. Lett. 174:247-250), which is available from the National Center for Biotechnology Information (NCBI) web site (www.ncbi.nlm.nih.gov), using the following settings: Matrix = Blosum62; Open gap = 11; Extension gap = 1; Penalties gap x_dropoff = 50; Expect = 10; Word size = 3; Filter on. The BLAST algorithm performs a two-step operation by first aligning a reference sequence (e.g., B-GEN polypeptide of any one of SEQ ID NO:1, SEQ ID NO:2, and SEQ ID NO:3, or the RuvC I, RuvC II, and RuvC III sub-domains of the nuclease domain thereof of any one of SEQ ID NO:8 or SEQ ID NO:11 (RuvC I), SEQ ID NO:9 or SEQ ID NO:12 (RuvC II), and SEQ ID NO:10 or SEQ ID NO:13 (RuvC II)) and a query sequence based and then determining the % sequence identity in a range of overlap between two aligned sequences. In addition to % sequence identity, BLASTP also determines the % sequence similarity based on the settings. In order to characterize the identity, subject sequences are aligned so that the highest order homology (match) is obtained.

[0053] Polypeptide, peptide and protein: As used herein, the terms “polypeptide,” “peptide” and “protein” refer to polymers of amino acids of any length. The polymer may in

various embodiments be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids.

[0054] Pluripotent: As used herein, the term “pluripotent” or “pluripotency” refers to the capacity of a cell to self-renew and to differentiate into cells of any of the three germ layers: endoderm, mesoderm, or ectoderm. “Pluripotent stem cells” or “PSCs” include, for example, embryonic stem cells derived from the inner cell mass of a blastocyst or derived by somatic cell nuclear transfer, and iPSCs derived from non-pluripotent cells.

[0055] Promoter: As used herein, the term “promoter” refers to a nucleotide sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a nucleic acid sequence. A promoter can be a constitutively active promoter (e.g., a promoter that is constitutively in an active “ON” state), it may be an inducible promoter (e.g., a promoter whose state, active/“ON” or inactive/“OFF”, is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein.), it may be a spatially restricted promoter (e.g., transcriptional control element, enhancer, *etc.*)(e.g., tissue specific promoter, cell type specific promoter, *etc.*), and it may be a temporally restricted promoter (e.g., the promoter is in the “ON” state or “OFF” state during specific stages of embryonic development or during specific stages of a biological process, e.g., hair follicle cycle in mice).

[0056] Protospacer Adjacent Motif: As used herein, the terms “protospacer adjacent motif” or “PAM” refer to a DNA sequence downstream (e.g., immediately downstream) of a target sequence on the non-target strand recognized by a Cas protein. A PAM sequence is located 3' of the target sequence on the non-target strand.

[0057] Recombinant: As used herein, the term “recombinant” in relation to a nucleic acid, polypeptide or cell refers to a nucleic acid (DNA or RNA), polypeptide or cell that is the product of genetic engineering, either directly or indirectly (e.g., is the progeny or replica of a nucleic acid, polypeptide or cell generated by genetic engineering methods). For example, a recombinant vector can be the product of various combinations of cloning, restriction, polymerase chain reaction (PCR) and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. DNA sequences encoding polypeptides can be assembled from cDNA fragments or from a series of synthetic oligonucleotides, to provide a synthetic nucleic

acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Genomic DNA comprising the relevant sequences can also be used in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not interfere with manipulation or expression of the coding regions and may indeed act to modulate production of a desired product by various mechanisms (see "DNA regulatory sequences," below). In addition or alternatively, DNA sequences encoding RNA (e.g., guide RNA) that is not translated may also be considered recombinant. Thus, e.g., the term "recombinant" nucleic acid refers to one which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is generally done to replace a codon with a codon encoding the same amino acid, a conservative amino acid, or a non-conservative amino acid. In addition or alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. When a recombinant nucleic acid encodes a polypeptide, the sequence of the encoded polypeptide can be naturally occurring ("wild type") or can be a variant (e.g., a mutant) of the naturally occurring sequence. Thus, the term "recombinant" polypeptide does not necessarily refer to a polypeptide whose sequence does not naturally occur. Instead, a "recombinant" polypeptide is encoded by a recombinant DNA sequence, but the sequence of the polypeptide can be naturally occurring ("wild type") or non-naturally occurring (e.g., a variant, a mutant, etc.). Thus, a "recombinant" polypeptide is the result of human intervention but may be a naturally occurring amino acid sequence. The term "non-naturally occurring" includes molecules that are markedly different from their naturally occurring counterparts, including chemically modified or mutated molecules.

[0058] Regulatory sequence: As used herein, the term "regulatory sequence" refers to a nucleic acid sequence which is required for expression of an operably linked sequence of interest, e.g., a guide RNA or an engineered B-GEN polypeptide sequence. In some instances, the regulatory sequence may be a promoter sequence and in other instances, the

regulatory sequence may include a promoter and an enhancer sequence and/or other regulatory elements which are required for expression of the pol. The regulatory sequence may, for example, be one which drives the expression of the operably linked sequence constitutively or in a tissue specific manner.

[0059] Ribonucleoprotein (RNP) complex, ribonucleoprotein (RNP) particle: As used herein, the terms “ribonucleoprotein complex” and “ribonucleoprotein particle” refer to a complex or particle including a nucleoprotein and a ribonucleic acid. A “nucleoprotein” as provided herein refers to a protein capable of binding a nucleic acid (e.g., RNA, DNA). Where the nucleoprotein binds a ribonucleic acid, it is referred to as “ribonucleoprotein.” The interaction between the ribonucleoprotein and the ribonucleic acid may be direct, e.g., by covalent bond, or indirect, e.g., by non-covalent bond (e.g., electrostatic interactions (e.g., ionic bond, hydrogen bond, halogen bond), van der Waals interactions (e.g., dipole-dipole, dipole-induced dipole, London dispersion), ring stacking (pi effects), hydrophobic interactions and the like). In embodiments, the ribonucleoprotein includes an RNA-binding motif non-covalently bound to the ribonucleic acid. For example, positively charged aromatic amino acid residues (e.g., lysine residues) in the RNA-binding motif may form electrostatic interactions with the negative nucleic acid phosphate backbones of the RNA, thereby forming a ribonucleoprotein complex. In some embodiments, any one of the engineered B-GEN polypeptides disclosed herein is in an RNP with a guide RNA.

[0060] Spacer: As used herein, the term “spacer” refers to a region of a gRNA molecule which is partially or fully complementary to a target sequence found in the + or - strand of genomic DNA. When complexed with a Cas protein, the gRNA directs the Cas protein to the target sequence in the genomic DNA. A spacer is typically 15 to 30 nucleotides in length (e.g., 20-25 nucleotides). The nucleotide sequence of a spacer can be, but is not necessarily, fully complementary to the target sequence. For example, in some embodiments, a spacer can contain one or more mismatches with a target sequence, e.g., the spacer can comprise one, two, or three mismatches with the target sequence.

[0061] Stem-Loop Structure: As used herein, the term “stem-loop structure” refers to a nucleic acid having a secondary structure that includes a region of nucleotides which are known or predicted to form a double strand (stem portion) that is linked on one side by a region of predominantly single-stranded nucleotides (loop portion). The terms “hairpin” and

"fold-back" structures are also used herein to refer to stem-loop structures. Such structures are well known in the art and these terms are used consistently with their known meanings in the art. As is known in the art, a stem-loop structure does not require exact base-pairing. Thus, the stem may include one or more base mismatches. Alternatively, the base-pairing may be exact, e.g., not include any mismatches.

[0062] Target cell: As used herein, the term "target cell" refers to a cell into which a nuclease, e.g., a B-GEN system of the disclosure, is introduced, for example a cell comprising target DNA in its genome. It should be understood that such term is intended to refer not only to the particular subject cell but to the progeny of such a cell. Because gene editing can take place in the cell as a result of the nuclease system, such progeny need not be identical to the parent cell into which the system was initially introduced but include gene edited counterparts of the cell. Such gene edited progeny are still included within the scope of the term "target cell" as used herein.

[0063] Target DNA: As used herein, the term "target DNA" refers to a polydeoxyribonucleotide that includes a "target site" or "target sequence." The terms "target site," "target sequence," "target protospacer DNA," or "protospacer-like sequence" are used interchangeably herein to refer to a nucleic acid sequence present in a target DNA to which a DNA-targeting segment (also referred to as a "spacer") of a guide RNA will bind, provided sufficient conditions for binding exist. For example, the target site (or target sequence) 5'-GAGCATATC-3' within a target DNA is targeted by (or is bound by, or hybridizes with, or is complementary to) the RNA sequence 5'-GAUAUGCUC-3'. Suitable DNA/RNA binding conditions include physiological conditions normally present in a cell. Other suitable DNA/RNA binding conditions (e.g., conditions in a cell-free system) are known in the art; see, e.g., Sambrook, J. and Russell, W., 2001. Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press. The strand of the target DNA that is complementary to and hybridizes with the guide RNA is referred to as the "complementary strand" and the strand of the target DNA that is complementary to the "complementary strand" (and is therefore not complementary to the guide RNA) is referred to as the "non-complementary strand" or "non-complementary strand." In some embodiments, the target DNA is genomic DNA.

[0064] Transfection: As used herein, the term “transfection” refers to the introduction of nucleic acid molecules, such as DNA or RNA (e.g., mRNA) molecules, into cells, e.g., into nuclei of target or production cells. In the context of the present disclosure, the term “transfection” encompasses any method known to the skilled person for introducing nucleic acid molecules into cells, e.g., into eukaryotic cells, such as into mammalian cells. Such methods encompass, for example, electroporation, lipofection, e.g., based on cationic lipids and/or liposomes, calcium phosphate precipitation, nanoparticle-based transfection, virus-based transfection, or transfection based on cationic polymers, such as DEAE-dextran or polyethylenimine. In some embodiments, the nucleic acid molecules are associated or complexed with a polypeptide, for example in the form of a ribonucleoprotein.

[0065] Vector: As used herein, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a “plasmid”, which refers to a circular double stranded DNA loop into which additional DNA segments can be incorporated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of nucleotide sequences to which they are operably linked. Such vectors are referred to herein as “expression vectors”. In some embodiments, a vector is a viral vector, e.g., an adenoviral vector, or an adeno-associated virus (AAV) vector.

6.2. Engineered Type V Endonucleases and B-GEN Polypeptides

[0066] The present disclosure provides engineered Type V endonucleases, e.g., engineered Bacillales Type V endonucleases. The engineered Type V endonucleases of the disclosure typically comprise an amino acid other than aspartic acid at the position corresponding to D504 of SEQ ID NO:1 (B-GEN.1) or D501 of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2).

[0067] In certain aspects, the disclosure also provides engineered Type V endonucleases comprising an OBD (e.g., an OBD-II) comprising the target-interacting sequence motif

$GX_1X_2X_3X_4NX_5X_6X_7DX_8$ (SEQ ID NO:204), where each of X_1 through X_8 is any amino acid. In some embodiments, (a) X_1 is selected from D, E, S, P, K, and R; (b) X_2 is independently selected from V, I, and A; (c) X_3 is independently selected from Y and F; (d) X_4 is independently selected from L and F; (e) X_5 is independently selected from I, L, F, V, and M; (f) X_6 is independently selected from S, V, T and A; (g) X_7 is independently selected from V, L and I; and (h) X_8 is independently selected from V, F, L and I. In exemplary embodiments, the target-interacting sequence motif is any of SEQ ID NO:201, SEQ ID NO:202, and SEQ ID NO:3.

[0068] In certain aspects, the disclosure also provides engineered Type V endonucleases, e.g., an engineered Type V endonuclease comprising an OBD comprising a target-interacting sequence motif as described above and/or comprising an amino acid other than aspartic acid at the position corresponding to D504 of a Type V endonuclease of SEQ ID NO:1 (B-GEN.1) and D501 of a Type V endonuclease of SEQ ID NO:2 (B-GEN.1.2) and SEQ ID NO:3 (B-GEN.2). In some embodiments, the amino acid at the position corresponding to D504 or D501 is arginine.

[0069] The engineered Type V endonucleases of the disclosure typically comprise an amino acid sequence that is at least 50% identical to the amino acid sequence of a Type V endonuclease of the order Bacillales, e.g., any amino acid sequence that is at least 50%, at least 60%, at least 70% or at least 80% identical to the amino acid sequence of any one of SEQ ID NOS:1, 2, 3, and 179-199, and comprises an OBD (e.g., an OBD-II) comprising the target-interacting sequence motif $GX_1X_2X_3X_4NX_5X_6X_7DX_8$ (SEQ ID NO:204), where each of X_1 through X_8 is any amino acid. In some embodiments, (a) X_1 is selected from D, E, S, P, K, and R; (b) X_2 is independently selected from V, I, and A; (c) X_3 is independently selected from Y and F; (d) X_4 is independently selected from L and F; (e) X_5 is independently selected from I, L, F, V, and M; (f) X_6 is independently selected from S, V, T and A; (g) X_7 is independently selected from V, L and I; and (h) X_8 is independently selected from V, F, L and I. In exemplary embodiments, the target-interacting sequence motif is any of SEQ ID NO:201, SEQ ID NO:202, and SEQ ID NO:203.

[0070] In certain aspects, the engineered Type V endonucleases are engineered B-GEN polypeptides. In some embodiments, the engineered B-GEN polypeptides comprise an amino acid sequence that is at least 50% identical to the entire length of B-GEN.1 (SEQ ID

NO:1), B-GEN1.2 (SEQ ID NO:2) or B-GEN.2 (SEQ ID NO:3) and/or differs from any one of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 by up to 25 amino acids. The amino acid sequence preferably comprises: (a) a target-interacting sequence motif of any one of SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203; and SEQ ID NO:204; (b) a RuvC I domain comprising an amino acid sequence having at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11; (c) a RuvC II domain comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12; (d) a RuvC III domain comprising an amino acid sequence having at least at least 80%, at least 85%, or at least 90% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13; or (e) any combination of two, three, or all four of (a), (b), (c) and (d).

[0071] In some embodiments, an engineered B-GEN polypeptide of the disclosure comprises an amino acid sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, or at least 95% identical to the entire length of SEQ ID NO:1, or SEQ ID NO:2 or SEQ ID NO:3. In some embodiments, an engineered B-GEN polypeptide comprises an amino acid sequence that is at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identical or at least 99.5% identical to the entire length of SEQ ID NO:1 or SEQ ID NO:2 or SEQ ID NO:3. The sequence preferably comprises: (a) a target-interacting sequence motif of any one of SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203; and SEQ ID NO:204; (b) a RuvC I domain comprising an amino acid sequence having at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11; (c) a RuvC II domain comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12; (d) a RuvC III domain comprising an amino acid sequence having at least at least 80%, at least 85%, or at least 90% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13; or (e) any combination of two, three, or all four of (a), (b), (c) and (d).

[0072] In some aspects, an engineered B-GEN polypeptide of the disclosure comprises (a) a target-interacting sequence motif of any one of SEQ ID NO:201, SEQ ID NO:202, SEQ ID

NO:203; and SEQ ID NO:204 and (b) RuvC I, RuvC II and RuvC III amino acid sequences that differ from corresponding sequences of B-GEN.1, B-GEN.1.2, or B-GEN.2 altogether by up to 25 amino acids across the three nuclease domains (SEQ ID NOS:8-10 for B-GEN.1 and SEQ ID NOS:11-13 for B-GEN.1.2 and B-GEN.2), optionally where there are no more than 3 amino acid differences in the RuvC III domain. In some embodiments, the engineered B-GEN polypeptide comprises an amino acid sequence having overall at least 70%, at least 80% or at least 90% to the sequence identity to the amino acid sequence of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3.

[0073] In yet further aspects, an engineered B-GEN polypeptide of the disclosure may comprise an amino acid sequence that differs from the entire sequence of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 by up to 25 amino acids, including the amino acid substitution D504R in relation SEQ ID NO:1 or D501R in relation to SEQ ID NO:2 or SEQ NO:3. In some embodiments, an engineered B-GEN polypeptide comprises an amino acid sequence that differs from the entire length of any one of SEQ ID NO:1, SEQ ID NO:2 or SEQ ID NO:3 by up to 25 amino acids, by up to 20 amino acids, by up to 15 amino acids, by up to 14 amino acids, by up to 13 amino acids, by up to 11 amino acids, by up to 10 amino acids, by up to 9 amino acids, by up to 8 amino acids, by up to 7 amino acids, by up to 6 amino acids, or by up to 5 amino acids, in each case including the amino acid substitution D504R in relation SEQ ID NO:1 or D501R in relation to SEQ ID NO:2 or SEQ NO:3.

[0074] The disclosure provides polypeptides, e.g., polypeptides having one more features as set forth in numbered embodiments 1 to 26, with at least 80% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 80% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0075] Some embodiments according to the disclosure are polypeptides, e.g., polypeptides having one more features as set forth in numbered embodiments 1 to 26, with at least 85% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each

case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 85% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0076] Other embodiments according to the disclosure are polypeptides, e.g., polypeptides having one more features as set forth in numbered embodiments 1 to 26, with at least 90% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 90% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0077] Further embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 95% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 95% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0078] Further embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 96% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to

D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 96% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0079] Additional embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 97% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 97% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0080] Other additional embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 98% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 98% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0081] Yet other embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 99% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ

NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 99% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0082] Yet other embodiments according to the disclosure are polypeptides, e.g., polypeptides having one or more features as set forth in numbered embodiments 1 to 26, with at least 99.5% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3 and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3, or nucleic acids comprising nucleotide sequences encoding polypeptides comprising nuclease sequences with least 99.5% sequence identity to any one of SEQ ID NO:1, SEQ ID NO:2, or SEQ ID NO:3, and in each case, including the amino acid arginine at the position corresponding to D504 of SEQ ID NO:1 or at the position corresponding to D501 of SEQ ID NO:2 or SEQ NO:3.

[0083] Exemplary B-GE_n consensus sequences, B-GE_n consensus I and B-GE_n consensus II, are provided herein as SEQ ID NO:7 and SEQ ID NO: 205, respectively.

[0084] Exemplary engineered B-GE_n sequences referred to herein as B-GE_n.1 D504R (SEQ ID NO:4), B-GE_n.1.2 D501R (SEQ ID NO:5), and B-GE_n.2 D501R (SEQ ID NO:6) are set forth in Table 1 below.

TABLE 1		
Name	Sequence	SEQ ID NO
B-GE _n .1 D504R aa sequence	MTIRSMKCLKKIYSGRSAPQLRQGLWRLHRLNNEGTAYYMDWLVMRQEA LPGKSKEEIRAELERRVRQQQEKNGVQNDQVPMDEVLSALRQLYELLVPS AVNNSGDAQTLRKFLSPLVDPNSEGGKGTSNAGAKPGWRKKQEAGDPSW EKDYERWLKRKQADPTAEILGKLETAGLKPLFPLYTNEVKDIRWMPPLTSK QYVRNWD RD MFQQAIEHLLSWETWNRKVNEERAKLKETVRRFEEQHLANG KDWLSPLQAYEANREQALRDMAISPSDRFRITRRQIKGWSELYERWNKLA PTASVEAYMQEVRHVQKKLG ^{GT} FGDADLYRFLAKPEN ^{VHI} WRDHQERLHY YAAYN ^{DL} HKRLMSAKEQA ^{AFT} L ^{PD} VAHPLWVRFDARDGNLFTYI ^{LQ} ADS SKQ ^{RSRR} YVNF ^{SR} FLWPVEDGYFEETENVKVELALS ^{KQ} FYRQ ^{VIV} HDNPT GKQKIT ^{FQ} DYSSKEILEGHLGGAKLQ ^{LD} RN ^{FL} RKSGRDFETGDFGPAFLN VVLRLKPKQE ^{VK} NGRLQ ^{SP} LGQALLV ^{KSR} PNDIPKVYGYKPDALA ^{AW} LEQ ASGEESLGSESLRQ ^{GFR} VMSIDL ^{GV} RSAA ^{IS} V ^{FS} VKGEK ^{TREG} DKVCYP VGETGLFAVH ^{DRS} FLLRLPGESSEKRVN ^{VER} DKRK ^{TERM} QIRYHIRT ^{LA} R VLR ^L LANKATPMDRIKAVQ ^{DV} LNDIESTRFMNDH ^{DH} VYNHALET ^{LR} TYAP	4

TABLE 1		
	DHQGIWEEQVIAAHRQLEHHVGVIVGEWRKNWGKDRRGTVGLSMDNIEEL DEMRRLLI SWSRRARY PREAKPFQVNESNPVHLLRHLQNLKEDRLKQLAN LIVMTALGYVYDSKEKKWKAAYPACQLILFEDLQRYRFHLDARSARENSQL MKWAHRSI PKYVWMQGEPEYGLQIGDVWAGFTSRYHAKTGAPGIRCKALTE KDFQQRLLLESIVAEGMFTLQEVGTLKPGDIVPAEGGELFVTLADDSGDR IVITHADINAQNQKREWLANSERFRVACRSVQIASQECFIPSSSVAK KMGKGVFVRDFS FHKDMEVYHWNQVKLTAKNVPTDHSDDLQDLQDYQAI LEEARESSSSYKTLFRDPSGFFFPDDVWVPQNIYWREVKKTTITALLRKRI MST	
B-Gen.1.2 D501R aa sequence	MPIRSFKLKLVTNHGDSYMDKLRRLGLWKTHVINRGIAYYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQRRNNWSEQTGTDELLSLSRVYELLV PSAIGEKGDQMLSRKFLSPLVDPNSEGGRGTA SGKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPFLPLFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERMLSWESWNNRRVAEEYQKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELA EVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQLRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFSSILPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVYYDYSSRFELTGVLGGAQIQFDRKHLKKAHRLAEGETGPIFLNVVV RVEPFLEVKNRRLRTP LGQVLQVNTRDWPVKVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGORTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARISLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELVS IQRKLELDQSEWISELEKLFDYIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISIWSIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHR SIPRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNSYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVTFADRSGASLA VIHADINAQNLOKRFWQHNSEVFRVPCKVKGGLVPVYEKMRKLF GKGL FVNIDDPESKEVYRWEHSTKMKS KTPVDLESEDI EHEELSDEWEDMQEG YKTLRLDPSGFFWSSDSWIPQKDFWIRVKSRIKLSLREQIR	5
B-Gen.2 D501R aa sequence	MPIRSFKLKLVTNHGDSYMDKLRRLGLWKTHVINRGIAYYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQRRNNWSEQTGTDELLSLSRVYELLV PSAIGEKGDQMLSRKFLSPLVDPNSEGGRGTA SGKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPFLPLFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERMLSWESWNNRRVAEEYKLLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELA EVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQLRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFSSILPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVYYDYSSRFELTGVLGGAQIQFDRKHLKKAHRLAEGETGPIFLNVVV RVEPFLEVKNRRLRTP LGQVLQVNTRDWPVKVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGORTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARISLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELVS IQKKLELDQSEWISELEKLFDYIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISIWSIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHR SIPRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNSYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVTFADRSGASLA VIHADINAQNLOKRFWQHNSEVFRVPCKVKGGLVPVYEKMRKLF GKGL	6

TABLE 1		
	FVNIDDPESKEVYRWEHSTKMKSKTTPVDLESEDIIDHEELSDEWEDMQEG YKTLLRDPSPGFFWSSDSWIPQKDFWIRVKSRIKSLREQIR	

[0085] In some embodiments, an engineered B-GEN polypeptide comprises an amino acid sequence that is identical to exemplary engineered B-GEN polypeptide of SEQ ID NO:4 or SEQ ID NO:5 or SEQ ID NO:6.

[0086] In some embodiments, the polypeptides further comprise nuclear localization signals, e.g., as set forth in Section 6.3, and/or linker sequences, e.g., as set forth in Section 6.4.

6.3. Nuclear Localization Signals

[0087] The engineered Type V endonucleases and B-GEN polypeptides of the disclosure may further comprise one or more nuclear localization signals (NLS's). In some embodiments, the engineered Type V endonucleases and B-GEN polypeptides of the present disclosure comprise one or more NLS's at their N-termini. In some embodiments, the engineered Type V endonucleases and B-GEN polypeptides of the disclosure comprise one or more NLS's at their C-termini. In some embodiments, the engineered Type V endonucleases and B-GEN polypeptides of the disclosure comprise one or more NLS's at both their N- and C-termini. The NLS's can be separated from the endonuclease sequences and from one another by linker sequences.

[0088] Accordingly, the disclosure provides engineered Type V endonucleases and B-GEN polypeptides comprising:

- (i) an engineered Type V endonucleases or B-GEN polypeptide sequence, as described in Section 6.2;
- (ii) one or more NLS sequences as described herein (e.g., in this Section 6.3); and
- (iii) one or optional linker sequences, e.g., as described in Section 6.4.

[0089] In certain aspects, an engineered Type V endonucleases or B-GEN polypeptide comprises a nuclease sequence and a first NLS sequence C-terminal to the nuclease sequence. An engineered Type V endonucleases or B-GEN polypeptide comprising a nuclease sequence and a first NLS sequence can further comprise a first linker sequence between the nuclease sequence and the first NLS sequence.

[0090] In certain aspects, an engineered Type V endonucleases or B-GEN polypeptide comprises more than one NLS sequence (e.g., more than one NLS sequences C-terminal to the nuclease sequence).

[0091] In some embodiments, an engineered Type V endonuclease or B-GEN polypeptide comprises a second NLS sequence C-terminal to the first NLS sequence. An engineered Type V endonucleases or B-GEN polypeptide comprising a second NLS sequence can further comprise a linker sequence between the first NLS sequence and the second NLS sequence.

[0092] In further embodiments, an engineered Type V endonuclease or B-GEN polypeptide comprises a third NLS sequence C-terminal to the second NLS sequence. An engineered Type V endonuclease or B-GEN polypeptide comprising a third NLS sequence can further comprise a linker sequence between the second NLS sequence and the third NLS sequence.

[0093] In additional embodiments, an engineered Type V endonuclease or B-GEN polypeptide comprises a fourth NLS sequence C-terminal to the third NLS sequence. An engineered Type V endonuclease or B-GEN polypeptide comprising a fourth NLS sequence can further comprise a linker sequence between the third NLS sequence and the fourth NLS sequence.

[0094] In some embodiments, an engineered Type V endonuclease or B-GEN polypeptide comprises an N-terminal NLS sequence in addition to the one or more NLS sequences C-terminal to the nuclease sequence. Thus, in certain aspects, an engineered Type V endonuclease or B-GEN polypeptide comprises an NLS sequence N-terminal to the nuclease sequence. An engineered Type V endonuclease or B-GEN polypeptide comprising an NLS sequence N-terminal to the nuclease sequence can further comprise a linker sequence between the NLS sequence and the nuclease sequence. In certain embodiments, an engineered Type V endonuclease or B-GEN polypeptide comprises more than one N-terminal NLS sequence (e.g., more than one NLS sequence N-terminal to the nuclease sequence that may be connected via one or more linkers).

[0095] Non-limiting examples of nuclear localization signals are listed in Table 2.

TABLE 2

NLS Sequence	NLS	SEQ ID NO
PKKKRKV	Wild-type SV40 large T protein	14
KRPAATKKAGQAKKKK	Nucleoplasmin	15
PAAKRVKLD	c-myc	16
MSRRRKANPTKLSENAKKLAKEVEN	EGL-13	17
KLKIKRPVK	TUS	18
(KR)-X(0,2)-(KR)-(KR)-x(3,10)-(RHK)-X(1,5)-PY	Basic PY-NLS	19
PKKKRMV	SV40 large T protein, variant 1	20
PKKKRKWEDP	SV40 large T protein, variant 2	21
CGYGPKKKRKVGG	SV40 large T protein, variant 3	22
CGYGPKKKRKV	SV40 large T protein, variant 4	23
CYDDEATADSQHSTPPKKKRKWEDPK DFESELLS	SV40 large T protein long NLS	24
CGGPKKKRKWVG	SV40 large T protein, variant 5	25
PKKKIKW	SV40 large T protein, variant 6	26
KRTADGSEFESPKKKRKV	SV40 large T protein, variant 7	27
TKKAGQAKKK	Nucleoplasmin min-NLS	28
TKKAGQAKKKKLD	Nucleoplasmin NLS variant 1	29
CGQAKKKKLD	Nucleoplasmin NLS variant 2	30
RQRRNELKRSP	c-myc NLS2	31
PKKARED	Polyoma large T protein	32
CGYGWSRKRPRPG	Polyoma virus large T protein	33
APTKRKGS	SV40 VP1 capsid polypeptide	34
APKRKSGVSKC	Polyoma virus major capsid protein VP1 (11 N-terminal aa)	35
PNKKKRK	SV40 VP2 capsid protein (39 kD)	36
EEDGPQKKKRRL	Polyoma virus capsid protein VP2	37
GKKRSKA	Yeast histone H2B	38
KRPRP	Adenovirus E1a	39
CGGLSSKRPRP	Adenovirus type 2/5 E1a	40
LVRKKRKTE ₃ SP	Xenopus N1, NLS1	41
LKD KDAKKSQKE	Xenopus N1, NLS2	42
GNKAKRQRST	V-Rel	43
PFLDRLRRDQK	NS1 protein of influenza A virus	44
SVTKKRKLE	Human lamin A	45
SASKRRRLE	Xenopus lamin A	46
ACIDKRVKLD	Human c-myc	47
SALIKKKKKMAP	Murine c-abl	48
PPKKRMRRRIE	Adenovirus 5 DBP	49
YRKCLQAGMNLARKTKKKIKGIQQATA	Rat glucocorticoid receptor	50
CGYGARKTKKKIK	Human glucocorticoid receptor	51

TABLE 2

NLS Sequence	NLS	SEQ ID NO
RKCLQAGMNLEARKTKK	Human glucocorticoid receptor NLS variant	52
RKFKKFNK	Rabbit progesterone receptor	53
CGYGIRKDRRGGR	Human estrogen receptor	54
CGYGARKLKKLGN	Human androgen receptor	55
GKRKNKPK	Chicken Ets1 core NLS	56
PLLKKIKQ	c-myb	57
PPQKKIKS	N-myc	58
PQPKKKP	p53	59
PQPKKKPL	p53 NLS variant	60
SKRWAKRKL	c-erb-A	61
MTGSKTRKHRGSGA	Yeast ribosomal protein L29 NLS	62
MTGSKHRKHGSGA	Yeast ribosomal protein L29 NLS, variant 1	63
RHRKHP	Yeast ribosomal protein L29 NLS, variant 2	64
KRRKHP	Yeast ribosomal protein L29 NLS, variant 3	65
KYRKHP	Yeast ribosomal protein L29 NLS, variant 4	66
KHRRHP	Yeast ribosomal protein L29 NLS, variant 5	67
KHKKHP	Yeast ribosomal protein L29 NLS, variant 6	68
RHLKHP	Yeast ribosomal protein L29 NLS, variant 7	69
KHRKYP	Yeast ribosomal protein L29 NLS, variant 8	70
KHRQHP	Yeast ribosomal protein L29 NLS, variant 9	71
LVRKKRKTE ₃ SP	Xenopus N1 NLS1	72
LKDKDAKSKQE	Xenopus N1 NLS2	73
ASKSRKRKL	Viral Jun	74
GGLCSARLHRHALLAT	Human T-cell leukemia virus Tax trans-activator protein	75
DTREKKKFLKRLLRLDE	Mouse nuclear MX1 protein (72 kD)	76
REKKKFLKRR	Mouse nuclear MX1 protein NLS variant	77
CGYGDRNKKKKE	Human retinoic acid receptor	78
RKRQRALMLRQAR	Human XPAC	79

TABLE 2

NLS Sequence	NLS	SEQ ID NO
EYLSRK GKLEL	T-DNA -linked VirD2 endonuclease	80
KKS KKKRC	Putative core NLS of yeast TRM1	206
QPQRYGGGRGRRW	Gag protein of human foamy retrovirus	81
NKKKRKLSRGSSQKTKGTSASAKARH KRRNRSSRS	SV40 Vp3 structural protein	82
RVTIRTWRWRPPK GKHRK	Simian sarcoma virus v-sis gene product	83
KRKIEEPEPEPKKAK	Putative bipartite NLS of Xenopus protein factor Xnf7	84
KKYENVVIKRSPRKGRPRKD	Yeast SWI5 gene product	85
GRKRAFHGDDPFGEGPPDKKGD	Herpes simplex virus ICP8 protein	86
KRPREDDDGEPSEKRARDDR	Bipartite NLS of VirD2 endonuclease	87
RMRIZFKNKGKDTAELRRRRVEVSVEL RKAKKDEQILKRRNV	IBB domain from importin-alpha	88
NQSSNFGPMKGGNFGGRSSGPYGGG GQYFAKPRNQGGY	hRNPAI M9 NLS	89
KRKGDEV DGVDEVAKKKSKK	Human poly(ADP-ribose) polymerase	90
RKLKKKIKKL	Hepatitis virus delta antigen	91
PKQKKRK	influenza virus NLS	92
SALIKKKKKMAP	mouse c-abl IV	93
VSRKRPRP	myoma T protein NLS1	94
PPKKARED	myoma T protein NLS2	95
GPAAKRVKLD	myc proto-oncogene protein [Homo sapiens]	96
KKRRIKQD	heat shock factor protein HSF8 [Solanum lycopersicum]	97
PKKKRKVEDPKKKRKVD	2x SV40, LrgT	98
PKKKRKVDPKKKRKVDPKKKRKV	3x SV40, LrgT	99
KKGKKKGK	monopartite NLS consensus cluster 1 (Dissertation Tatyana Goldberg 2016)	100
PKRRRGVVL	monopartite NLS consensus cluster 2 (Dissertation Tatyana Goldberg 2016)	101

TABLE 2

NLS Sequence	NLS	SEQ ID NO
EQLFKRRNV	monopartite NLS consensus cluster 3 (Dissertation Tatyana Goldberg 2016)	102
KRRRR	monopartite NLS consensus cluster 4 (Dissertation Tatyana Goldberg 2016)	103
KKRRR	monopartite NLS consensus cluster 5 (Dissertation Tatyana Goldberg 2016)	104
EGAPPAKRPR	monopartite NLS consensus cluster 6 (Dissertation Tatyana Goldberg 2016)	105
MLRRRRRKRAR	monopartite NLS consensus cluster 7 (Dissertation Tatyana Goldberg 2016)	106
RRKRR	monopartite NLS consensus cluster 8 (Dissertation Tatyana Goldberg 2016)	107
RKRK	monopartite NLS consensus cluster 9 (Dissertation Tatyana Goldberg 2016)	108
FKAVLEDILGEL	monopartite NLS minor cluster (Dissertation Tatyana Goldberg 2016)	109
KNRRL	NLSdb, protein origin P10152	110
RKRHW	NLSdb, protein origin Q09353	111
RRKKRR	NLSdb, protein origin Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1	112
RRKRSR	NLSdb, protein origin Q99PU7, D3ZHS6, Q92560, A2VDM8	113
KRGRKP	NLSdb, protein origin Q14781, P30658	114
KKRKLE	NLSdb, protein origin P02545, P48678, P48679, Q3ZD69	115
PKKKSrk	NLSdb, protein origin Q35914, Q01954	116
PKRGRGR	NLSdb, protein origin Q9FYS5, Q43386	117
KEKRKKR	NLSdb, protein origin E5RQA1	118
KKKKRKR	NLSdb, protein origin Q9Z1J1, Q9HCS4, Q924A0	119
RRGDGRRR	NLSdb, protein origin Q80WE1, Q5R9B4, Q06787, P35922	120

TABLE 2

NLS Sequence	NLS	SEQ ID NO
LSPSLSPL	NLSdb, protein origin Q9Y261, P32182, P35583	121
VNFSEFSK	NLSdb, protein origin P07156	122
IVINILSE	NLSdb, protein origin Q96EB6	123
PPAKRKCIF	NLSdb, protein origin Q6AZ28, O75928, Q8C5D8	124
QRPGPYDRP	NLSdb, SeqNLS prediction	125
KRKRGRPRK	NLSdb, protein origin Q8L7L5, A1L4X7, O80834, Q8LPN5	126
KIKELYRRR	NLSdb, protein origin O88907, O75925	127
MVQLRPRASR	NLSdb, SeqNLS prediction	128
KKRREKQRRR	NLSdb, protein origin Q5VK71	129
EGAPPAKRAR	NLSdb, protein origin P0C6L6, P25880, Q81835, P29833, P25882, P06934, P0C6L3, P29997, P0C6M5, P0C6M9, P29996, P0C6M1, P0C6L8, P0C6M2, P0C6L7, P25881	130
PKKGDKYDKTD	NLSdb, protein origin Q45FA5	131
KKKKSKDKKRK	NLSdb, protein origin P97376, Q14331	132
AHRAKKMSKTHA	NLSdb, protein origin P21827	133
KKGPSVQKRKKT	NLSdb, protein origin Q6ZN17	134
KGVKRKADTTTP	NLSdb, protein origin Q4R8Y1	135
KGVKRRADTTTP	NLSdb, protein origin Q91Y44, D4A7T3	136
KKPKWDDFKKKKK	NLSdb, protein origin Q15397, Q8BKS9, Q562C7	137
KRRRRRRRREKRKR	NLSdb, protein origin Q96GM8	138
DVRKRVQDLEQKM	NLSdb, protein origin P61635, Q6DV79, Q19S50, P52631	139
KKGKDEWFSRGKK	NLSdb, protein origin O60716	140
KKGKDEWFSRGKKP	NLSdb, protein origin P30999	141
ASPEYVNLPINGNG	NLSdb, SeqNLS prediction	142
YLRPVKKPKIRRK	NLSdb, protein origin Q7Z7C8, Q5ZMS1, Q9EQH4, A7MAZ4	143
KRKGKLNKGSKRKK	NLSdb, protein origin O15381	144
RRRGKNKVAAQNCRK	NLSdb, SeqNLS prediction	145
DKAKRVSRNKSEKKRR	NLSdb, protein origin O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4	146

TABLE 2

NLS Sequence	NLS	SEQ ID NO
EEQLRRRKNSRLNNTG	NLSdb, protein origin G5EFF5	147
HKKKHPDASVNFSEFSK	NLSdb, protein origin P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0	148
KKTGKNRKLKSKRVKTR	NLSdb, protein origin Q9Z301, O54943, Q8K3T2	149
KRSCRRRLAGHNERRRK	NLSdb, protein origin Q38740, Q38741, Q700W2, Q9S7A9, Q6Z461, P93015, Q94JW8, Q9S758	150
KRQRRKQSNRESARRSR	NLSdb, protein origin Q501B2	151
RGKGGKGLGKGGAKRHRK	NLSdb, SeqNLS prediction	152
RRRGFERFGPDNMGRKRK	NLSdb, protein origin Q63014, Q9DBR0	153
RRHQGGQGDDSSHKKERK	NLSdb, protein origin Q0IJ08, Q2TAE3, Q63470, Q13627, Q61214	154
KKKTGVIAPKRFVQRLKK	NLSdb, protein origin Q8LAM0, O24454	155
KRAMKDDSHGNSTSPKRRK	NLSdb, protein origin Q0E671	156
KVNFLDMSLDDIIYKELE	NLSdb, protein origin Q9P127	157
KKYENVVIKRSPRKRGRPRK	NLSdb, SeqNLS prediction	158
KRGNSSIGPNDLSKRKQRKK	NLSdb, SeqNLS prediction	159
KRASEDTTSGSPPKKSSAGPKR	NLSdb, protein origin Q9BZZ5, Q5R644	160
KRIHSVLSQSQIDPSKKVKRAK	NLSdb, SeqNLS prediction	161
EVLKVIRTGKRKKKAWKRMVTKVC	NLSdb, SeqNLS prediction	162
IINGRKLKLKKSRRRSSQTSNNSFTSRRS	NLSdb, SeqNLS prediction	163
AHFKISGEKRPSTDPGKKAKNPKKKKKKDP	NLSdb, protein origin Q76IQ7	164

[0096] In some embodiments, an engineered B-GE_n polypeptide comprises one or more NLS sequences only at the N-terminus of a B-GE_n.2 protein sequence. In other embodiments, a fusion protein of the disclosure comprises one or more NLS sequences only at the C-terminus of a B-GE_n.2 protein sequence. An exemplary engineered B-GE_n polypeptide with a C-terminal NLS is represented by SEQ ID NO:200.

[0097] In some embodiments, an engineered B-GEN polypeptide comprises multiple NLS sequences that are identical. In other embodiments, a fusion protein of the disclosure comprises multiple NLS sequences that are distinct.

[0098] In some embodiments, an engineered B-GEN polypeptide comprises a nucleoplasmin NLS (SEQ ID NO:15) on the N-terminus and an SV40 large T protein NLS (SEQ ID NO:14) on the C-terminus of a B-GEN polypeptide sequence as described in Section 6.2.

[0099] In some embodiments, an engineered B-GEN polypeptide comprises an SV40 large T protein NLS (SEQ ID NO:14) on the C-terminus of a B-GEN polypeptide sequence as described in Section 6.2.

[0100] In some embodiments, an engineered B-GEN polypeptide comprises a Nucleoplasmin NLS (SEQ ID NO:15) on the N-terminus of a B-GEN polypeptide sequence as described in Section 6.2).

6.4. The Linker Sequence

[0101] The disclosure provides engineered B-GEN polypeptides, which are in the form of fusion proteins comprising a B-GEN polypeptide sequence, e.g., having an amino acid sequence as described in Section 6.2, fused with one or more NLS sequences, optionally via a peptide linker.

[0102] In some embodiments, a B-GEN polypeptide sequence is connected to an NLS sequence on its N- and/or C-terminus via a peptide linker. In other embodiments, a B-GEN polypeptide sequence is connected to one or more NLS sequences with a peptide linker connecting a pair of individual polypeptide sequences, such as between two NLS sequences or between an NLS sequence and a B-GEN polypeptide sequence.

[0103] In some embodiments, a B-GEN polypeptide of the disclosure comprises multiple NLS sequences that are connected by identical linkers. In other embodiments, a B-GEN polypeptide of the disclosure comprises multiple NLS sequences that are connected by distinct linkers.

[0104] Peptide linkers suitable for use in the B-GEN polypeptides of the disclosure include those disclosed in Chen *et al.*, 2013, Adv Drug Deliv Rev. 65(10):1357-1369. Nonlimiting examples of such linkers are reproduced below in Table 3.

TABLE 3		
Linker Sequence	Linker Type	SEQ ID NO
(GGGGS) ₃	Flexible	165
GGG(EAAAK) ₃	Mixed	166
A(EAAAK) ₄ ALEA(EAAAK) ₄ A	Rigid	167
GGGGSLVPRGSGGGGS	Flexible	168
GAAPAAAPAKQEAAAPAPAAKAEAPAAAPAAKA	Proline rich, rigid, interdomain	169
TRHRQPRGWE	Cleavable	170

6.5. B-Gen Type V CRISPR-Cas Systems

[0105] The disclosure provides engineered Type V endonuclease and B-Gen polypeptide systems incorporating the engineered Type V endonuclease and B-Gen polypeptides of the disclosure or nucleic acids encoding them.

[0106] In some embodiments, an engineered Type V endonuclease or B-Gen polypeptide system comprises the following components:

- (a) an engineered Type V endonuclease or B-Gen polypeptide, e.g., as described in Section 6.2, or a nucleic acid encoding such engineered B-Gen polypeptide, e.g., as described in Section 6.8,
- (b) a heterologous guide RNA (gRNA), e.g., as described in Section 6.6, or a nucleic acid (e.g., a vector as described in Section 6.9) that allows the generation of such gRNA in situ, wherein the gRNA comprise(s):
 - i. an engineered DNA targeting segment composed of RNA and capable of hybridizing to a target sequence in a nucleic acid locus,
 - ii. a tracr mate sequence composed of RNA, and
 - iii. a tracr RNA sequence composed of RNA,

wherein the tracr mate sequence hybridizes to the tracr sequence, and wherein (i), (ii), and (iii) are arranged in a 5' to 3' orientation. The gRNA can be a single guide RNA (sgRNA). Within a sgRNA, a tracr mate sequence and a tracr sequence are generally connected by a suitable loop sequence and form a stem-loop structure.

[0107] In particular embodiments, an engineered Type V endonuclease or B-Gen polypeptide system comprises the following components:

- (a) an engineered Type V endonuclease or B-GEN polypeptide, e.g., as described in Section 6.2,
- (b) a heterologous guide RNA (gRNA), e.g., as described in Section 6.6, which comprise(s):
 - i. an engineered DNA targeting segment composed of RNA and capable of hybridizing to a target sequence in a nucleic acid locus,
 - ii. a tracr mate sequence composed of RNA, and
 - iii. a tracr RNA sequence composed of RNA,

wherein the tracr mate sequence hybridizes to the tracr sequence, and wherein (i), (ii), and (iii) are arranged in a 5' to 3' orientation. The gRNA can be a single guide RNA (sgRNA). Within a sgRNA, a tracr mate sequence and a tracr sequence are generally connected by a suitable loop sequence and form a stem-loop structure.

[0108] In some embodiments, such engineered Type V endonucleases or B-GEN polypeptides are delivered to target cells in a composition known as a ribonucleoprotein (RNP) complex described in Section 6.7.

[0109] In particular embodiments, engineered Type V endonucleases or B-GEN polypeptide systems comprise the following components:

- (a) a nucleic acid, e.g., as described in Section 6.8, encoding an engineered Type V endonucleases or B-GEN polypeptide, e.g., as described in Section 6.2, or
- (b) a nucleic acid (e.g., a vector as described in Section 6.9) that allows the generation of a heterologous guide RNA (gRNA), e.g., as described in Section 6.6, wherein the gRNA comprise(s):
 - i. an engineered DNA targeting segment composed of RNA and capable of hybridizing to a target sequence in a nucleic acid locus,
 - ii. a tracr mate sequence composed of RNA, and
 - iii. a tracr RNA sequence composed of RNA,

wherein the tracr mate sequence hybridizes to the tracr sequence, and wherein (i), (ii), and (iii) are arranged in a 5' to 3' orientation. The gRNA can be a single guide RNA (sgRNA). Within a sgRNA, a tracr mate sequence and a tracr sequence are generally connected by a suitable loop sequence and form a stem-loop structure.

[0110] Any reference to “RNA” or “guide RNA” encompasses RNA molecules comprising non-natural as well as natural nucleobases, for example one or more of the nucleic acid modifications described in Section 6.8.3.

[0111] In some embodiments, the nucleic acid(s) encoding the engineered Type V endonuclease or B-Gen polypeptide and / or the sgRNAs contain a suitable promoter for the expression in a cellular or in vitro environment.

[0112] In some embodiments, the nucleic acid(s) encoding the engineered Type V endonuclease or B-Gen polypeptide and / or the sgRNAs is in the form of a viral vector, for example as described in Section 6.9.3.

6.6. Guide RNAs (gRNAs) and single guide RNAs (sgRNAs)

[0113] The systems, compositions, and methods described herein in some embodiments employ a genome-targeting nucleic acid that can direct the activities of an engineered B-Gen polypeptide to a specific target sequence within a target nucleic acid. In some embodiments, the genome-targeting nucleic acid is an RNA. A genome-targeting RNA is referred to as a “guide RNA” or “gRNA” herein. A guide RNA has at least a spacer sequence that can hybridize to a target nucleic acid sequence of interest and a CRISPR repeat sequence (such a CRISPR repeat sequence is also referred to as a “tracr mate sequence”). In Type II systems, the gRNA also has a second RNA called the tracrRNA sequence. In the Type II guide RNA (gRNA), the CRISPR repeat sequence and tracrRNA sequence hybridize to each other to form a duplex. In the Type V guide RNA (gRNA), the crRNA forms a duplex. In both systems, the duplex binds a site-specific polypeptide such that the guide RNA and site-direct polypeptide form a complex. The genome-targeting nucleic acid provides target specificity to the complex by virtue of its association with the site-specific polypeptide. The genome-targeting nucleic acid thus directs the activity of the site-specific polypeptide.

[0114] In some embodiments, the genome-targeting nucleic acid is a double-molecule guide RNA. In some embodiments, the genome-targeting nucleic acid is a single-molecule guide

RNA or single guide RNA (sgRNA). A double-molecule guide RNA has two strands of RNA. The first strand has in the 5' to 3' direction, an optional spacer extension sequence, a spacer sequence and a minimum CRISPR repeat sequence. The second strand has a minimum tracrRNA sequence (complementary to the minimum CRISPR repeat sequence), a 3' tracrRNA sequence and an optional tracrRNA extension sequence. A single-molecule guide RNA (sgRNA) in a Type II system has, in the 5' to 3' direction, an optional spacer extension sequence, a spacer sequence, a minimum CRISPR repeat sequence, a single-molecule guide linker, a minimum tracrRNA sequence, a 3' tracrRNA sequence and an optional tracrRNA extension sequence. The optional tracrRNA extension may have elements that contribute additional functionality (e.g., stability) to the guide RNA. The single-molecule guide linker links the minimum CRISPR repeat and the minimum tracrRNA sequence to form a hairpin structure. The optional tracrRNA extension has one or more hairpins.

[0115] A single-molecule guide RNA (sgRNA) in a Type V system has, in the 5' to 3' direction, a minimum CRISPR repeat sequence and a spacer sequence. Or alternatively, a single-molecule guide RNA (sgRNA) in a Type V system has, in the 5' to 3' direction, optional tracr extension sequence, a tracr RNA sequence, a single molecule guide linker, a minimum CRISPR repeat sequence, a spacer sequence, and an optional spacer extension sequence.

[0116] Alternatively, a single-molecule guide RNA (sgRNA) in a Type V system has, in the 5' to 3' direction, an optional extension sequence, a minimum CRISPR repeat sequence, a spacer sequence, and an optional spacer extension sequence.

[0117] In some embodiments, an sgRNA in a Type V system comprises, in the 5' to 3' direction, an optional extension sequence, an artificial nuclease binding RNA sequence and a spacer sequence, and an optional spacer extension sequence.

[0118] Particularly useful sgRNAs for the B-GEN.2 CRISPR Cas nucleases according to the disclosure, and potentially for other Type V endonuclease nucleases, are disclosed in Table 4.

TABLE 4		
sgRNA	Sequence	SEQ ID NO
B-GEN.2-sgRNA_v4	guuuagcuau aggcuaauaa gauaguugug ucaagugcuu cggagaccua acacgucauc	171

TABLE 4		
	cagucacaac ggcuaucaaa aaauagaugg uuagcac	
B-GEN.2-sgRNA_v4.2	uuagcuauag gcuaauaaga uaguuguguc aagugcuucg gagaccuaac acgucaucca gucacaacgg cuaaaaaugg uuagcac	172
B-GEN.2-sgRNA_v4.3	uuagcuauag gcuaauaaga uaguuguguc aagugcuucg gagaccuaac acgucaucca gucacaacgg cuaaaaauag ccagcac	173
B-GEN.2-sgRNA_v4.4	uuagcuauag gcuaauaaga uaguuguguc aagugcuucg gagaccuaac acgucuccag ucacaacggc uaaaaauagc cagcac	174
B-GEN.2-sgRNA_v4.5	uuagcuauag gcuaauaaga uaguugugac aagugcuucg gagaccuaac acgucuccag ucacaacggc uaaaaauagc cagcac	175

[0119] Exemplary genome-targeting nucleic acids are described, for example, in WO2018002719. In general, a CRISPR repeat sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a DNA targeting segment flanked by CRISPR repeat sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex includes the CRISPR repeat sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the CRISPR repeat sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm and may further account for secondary structures, such as self-complementarity within either the tracr sequence or CRISPR repeat sequence. In some embodiments, the degree of complementarity between the tracr sequence and CRISPR repeat sequence along the 30 nucleotides length of the shorter of the two when optimally aligned is about or more than 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and CRISPR repeat sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In some embodiments, the transcript or transcribed nucleic acid sequence has at least two or more hairpins.

[0120] Suitable tracr sequences for use with B-Gen.2 or B-Gen.1 in CRISPR Cas systems are listed in Table 5. Alternatively, variants of these sequences could be employed. Variants could include either parts or truncated versions of such sequences and/or sequences with base modifications in one or more places of these sequences. The respective RNA sequences are disclosed in SEQ ID NOs:176 and 177, respectively.

TABLE 5		
Name	SEQ ID NO:	RNA sequence
tracr_B-Gen.2	176	UAAUAAAUAUAUGAGUUACAGGCUCAUGUCAUUCGUGC GUGUGCUGCGAGGCUCUACAACGUCUCAACACGGGAUC ACGUUAUUUCCACUAAGUGAUUGGAC
tracr_B-Gen.1	177	CUGGAUAUAUAAGAUGUUUAGCUAUAGGCUAUAAGAUA GUUGUGUCAAGUGCUUCGGAGACCUAACACGUCAUCCAGU CACAACGGCUAUCUAUAUUUCCACUAAC

[0121] The spacer of a guide RNA includes a nucleotide sequence that is complementary to a sequence in a target DNA. In other words, the spacer of a guide RNA interacts with a target DNA in a sequence-specific manner via hybridization (e.g., base pairing). As such, the nucleotide sequence of the spacer may vary and determines the location within the target DNA that the guide RNA and the target DNA will interact. The DNA- targeting segment of a guide RNA can be modified (e.g., by genetic engineering) to hybridize to any desired sequence within a target DNA.

[0122] In some embodiments, the spacer has a length of from 10 nucleotides to 30 nucleotides. In some embodiments, the spacer has a length of from 13 nucleotides to 25 nucleotides. In some embodiments, the spacer has a length of from 15 nucleotides to 23 nucleotides. In some embodiments, the spacer has a length of from 18 nucleotides to 22 nucleotides, e.g., from 20 to 22 nucleotides.

[0123] In some embodiments, the percent complementarity between the DNA-targeting sequence of the spacer and the protospacer of the target DNA is at least 60% (e.g., at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 97%, at least 98%, at least 99%, or 100%) over the 20-22 nucleotides.

[0124] In some embodiments, the protospacer is directly adjacent to a suitable PAM sequence on its 3' end or such PAM sequence is part of the DNA targeting sequence in its 3' portion.

[0125] Suitable PAM sequences are listed in Table 6, wherein the engineered DNA targeting segment is, on its 3' end, directly adjacent to the PAM sequence on the targeted DNA segment, or such PAM sequence is part of the targeted DNA sequence in its 5' portion.

Table 6	
B-GEN	PAM Sequence
B-GEN.1	“DTTN” with “D” representing “A” or “T” or “G”, preferably “ATT”
B-GEN.1.2	“DTTN” with “D” representing “A” or “T” or “G”, preferably “ATT”
B-GEN.2	“DTTN” with “D” representing “A” or “T” or “G”, preferably “ATT”

[0126] Modifications of guide RNAs can be used to enhance the formation or stability of the CRISPR-Cas genome editing complex comprising guide RNAs and a Cas endonuclease such as B-GEN.1 or B-GEN.2. Modifications of guide RNAs can also or alternatively be used to enhance the initiation, stability or kinetics of interactions between the genome editing complex with the target sequence in the genome, which can be used for example to enhance on-target activity. Modifications of guide RNAs can also or alternatively be used to enhance specificity, e.g., the relative rates of genome editing at the on-target site as compared to effects at other (off-target) sites.

[0127] Modifications can also or alternatively be used to increase the stability of a guide RNA, e.g., by increasing its resistance to degradation by ribonucleases (RNases) present in a cell, thereby causing its half-life in the cell to be increased. Modifications enhancing guide RNA half-life can be particularly useful in embodiments in which a Cas endonuclease such as a B-GEN.1, or B-GEN.1.2, or B-GEN.2 is introduced into the cell to be edited via an RNA that needs to be translated in order to generate B-GEN.1, or B-GEN.1.2, or B-GEN.2 endonuclease, since increasing the half-life of guide RNAs introduced at the same time as the RNA encoding the endonuclease can be used to increase the time that the guide RNAs and the encoded Cas endonuclease co-exist in the cell.

6.6.1. Additional Sequences

[0128] In some embodiments, a guide RNA comprises at least one additional segment at either the 5' or 3' end. For example, a suitable additional segment can comprise a 5' cap

(e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (e.g., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and protein complexes); a sequence that forms a dsRNA duplex (e.g., a hairpin); a sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, etc.); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like) a modification or sequence that provides for increased, decreased, and/or controllable stability; and combinations thereof.

6.6.1.1. Stability Control Sequences

[0129] A stability control sequence influences the stability of an RNA (e.g., a guide RNA). A non-limiting example of a suitable stability control sequence is a transcriptional terminator segment (e.g., a transcription termination sequence). A transcriptional terminator segment of a guide RNA can have a total length of from 10 nucleotides to 100 nucleotides, e.g., from 10 nucleotides (nt) to 20 nt, from 20 nt to 30 nt, from 30 nt to 40 nt, from 40 nt to 50 nt, from 50 nt to 60 nt, from 60 nt to 70 nt, from 70 nt to 80 nt, from 80 nt to 90 nt, or from 90 nt to 100 nt. For example, the transcriptional terminator segment can have a length of from 15 nucleotides (nt) to 80 nt, from 15 nt to 50 nt, from 15 nt to 40 nt, from 15 nt to 30 nt or from 15 nt to 25 nt.

[0130] In some embodiments, the transcription termination sequence is one that is functional in a eukaryotic cell. In some embodiments, the transcription termination sequence is one that is functional in a prokaryotic cell.

[0131] Nucleotide sequences that can be included in a stability control sequence (e.g., transcriptional termination segment, or in any segment of the guide RNA to provide for increased stability) include, for example, a Rho-independent trp termination site.

6.7. Ribonucleoprotein (RNP) Complexes

[0132] In some embodiments, the engineered Type V endonucleases and B-GEN polypeptides are delivered in a composition known as a ribonucleoprotein or RNP complex.

An RNP complex is assembled by combining a Cas endonuclease, such as an engineered Type V or B-GEN endonuclease, with a ribonucleic acid, e.g., a guide RNA (gRNA).

[0133] In some embodiments, the ribonucleoprotein complex comprises an engineered B-GEN endonuclease, e.g., as described in Section 6.2, complexed with a suitable ribonucleic acid. In some embodiments, the ribonucleic acid is a gRNA or an sgRNA, which are described further in Section 6.6. In some embodiments, the RNP complex comprises an engineered B-GEN polypeptide and an sgRNA listed in Table 4 or another suitable sgRNA.

[0134] In some embodiments, an engineered B-GEN polypeptide and an sgRNA are present in a molar ratio from about 1:1 to about 1:4. In some embodiments, the molar ratio is from about 1:1 to about 1:3. In some embodiments, the molar ratio is from about 1:1 to about 1:2.5. In some embodiments, the molar ratio is from about 1:1 to about 1:2. In some embodiments, the molar ratio is from about 1:1 to about 1:1.5. In some embodiments, the molar ratio of polypeptide to an sgRNA is 1:1.

[0135] One of the most common techniques for delivery of RNPs is electroporation, which generates pores in the cell membrane, allowing for entry of the RNP into the cytoplasm. Further, electroporation can be combined with cell-type specific reagents in a technique known as nucleofection, which forms pores in the nuclear membrane, allowing for entry of a DNA template.

[0136] In some embodiments, an engineered Type V endonuclease or B-GEN polypeptide in an RNP complex is delivered into target cells via nucleofection.

6.8. Nucleic Acids

[0137] The disclosure provides nucleic acids (e.g., DNA or RNA) encoding B-GEN Type V CRISPR-Cas proteins (e.g., an engineered B-GEN polypeptide), nucleic acids encoding gRNAs or sgRNAs of the disclosure, nucleic acids encoding both engineered B-GEN polypeptides and gRNAs or sgRNAs, and pluralities of nucleic acids, for example comprising a nucleic acid encoding an engineered B-GEN polypeptide and a gRNA or an sgRNA.

[0138] Nucleic acids encoding an engineered B-GEN polypeptide can be codon optimized, e.g., where at least one non-common codon or less-common codon has been replaced by a codon that is common in a host cell or target cell. For example, a codon optimized nucleic

acid can direct the synthesis of an optimized messenger mRNA, e.g., optimized for expression in a mammalian expression system.

6.8.1. B-GEN Coding Sequences

[0139] In some embodiments, a nucleic acid described herein comprises one or more modifications which can be used, for example, to enhance activity, stability or specificity, alter delivery, reduce innate immune responses in host cells, further reduce the protein size, or for other enhancements, as further described herein and known in the art. In some embodiments, such modifications will result in an engineered B-GEN polypeptide whose nucleic acid sequence component has at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% amino acid sequence identity to the sequence of SEQ ID NOs: 4, 5, or 6.

6.8.2. Codon Optimization

[0140] In certain embodiments, modified nucleic acids are used in a CRISPR-B-GEN.1, or B-GEN.1.2, or B-GEN.2 system described herein, in which the guide RNAs and/or a DNA or an RNA comprising a nucleic acid sequence encoding an engineered B-GEN polypeptide can be modified, as described below. Such modified nucleic acids can be used in the CRISPR-B-GEN.1, or B-GEN.1.2, or B-GEN.2 system to edit any one or more genomic loci. In some embodiments, such modifications in the nucleic acids of the disclosure are achieved via codon-optimization, e.g., codon-optimized based on specific host cells in which the encoded polypeptide is expressed. It will be appreciated by the skilled artisan that any nucleotide sequence and/or recombinant nucleic acid of the present disclosure can be codon optimized for expression in any species of interest. Codon optimization is well known in the art and involves modification of a nucleotide sequence for codon usage bias using species specific codon usage tables. The codon usage tables are generated based on a sequence analysis of the most highly expressed genes for the species of interest. In a non-limiting example, when the nucleotide sequences are to be expressed in the nucleus, the codon usage tables are generated based on a sequence analysis of highly expressed nuclear genes for the species of interest. The modifications of the nucleotide sequences are determined by comparing the species-specific codon usage table with the codons present in the native nucleic acid sequences.

[0141] In some embodiments, an engineered B-GEN polypeptide described herein is expressed from a codon-optimized nucleic acid sequence. For example, if the intended host cell or target cell were a human cell, a human codon-optimized nucleic acid sequence encoding an engineered B-GEN polypeptide comprising the amino acid sequence of B-GEN.1, or B-GEN.1.2, or B-GEN.2 (or a B-GEN.1, or B-GEN.1.2, or B-GEN.2 variant, e.g., enzymatically inactive variant) would be suitable. As another non-limiting example, if the intended host cell or target cell were a mouse cell, then a mouse codon-optimized nucleic acid sequence an engineered B-GEN polypeptide comprising the amino acid sequence of B-GEN.1, or B-GEN.1.2, or B-GEN.2 (or a B-GEN.1, or B-GEN.1.2, or B-GEN.2 variant, e.g., enzymatically inactive variant) would be suitable.

[0142] Strategies and methodologies for codon optimization are known in the art and have been described for various systems including, but not limited to yeast (Outchkourov *et al.*, Protein Expr Purif, 24(1):18-24 (2002)) and E. coli (Feng *et al.*, Biochemistry, 39(50):15399-15409 (2000)). In some embodiments, the codon optimization is performed by using GeneGPS® Expression Optimization Technology (ATUM) and using the manufacturer's recommended expression optimization algorithms. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in a human cell. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in an E. coli cell. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in an insect cell. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in a Sf9 insect cell. In some embodiments, the expression optimization algorithms used in codon optimization procedure are defined to avoid putative poly-A signals (e.g., AATAAA and ATTAAA) as well as long (greater than 4) stretches of A's which can lead to polymerase slippage.

[0143] As is well understood in the art, codon optimization of a nucleotide sequence results in a nucleotide sequence having less than 100% identity (e.g., less than 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%, or 99%) to the native nucleotide sequence but which still encodes a polypeptide having the same function as that encoded by the original, native nucleotide sequence. Thus, in representative embodiments of the

disclosure, the nucleotide sequence and/or recombinant nucleic acid of the disclosure can be codon optimized for expression in the particular species of interest.

[0144] In some embodiments, a codon-optimized nucleic acid sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.2%, 99.5%, 99.8%, 99.9%, or 100% sequence identity to SEQ ID NO: 4, SEQ ID NO:5, or SEQ ID NO:6. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression of the encoded engineered B-GEN polypeptide in a target cell or a host cell. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in a human cell. Generally, the nucleic acids of the disclosure are codon-optimized for increased expression in any human cells. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in an E. coli cell. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in an insect cell. Generally, the nucleic acids of the disclosure are codon-optimized for increased expression in any insect cells. In some embodiments, the nucleic acids of the disclosure are codon-optimized for increased expression in a Sf9 insect cell expression system.

[0145] Polyadenylation signals can also be chosen to optimize expression in the intended host.

6.8.3. Nucleic Acid Modifications

[0146] In some embodiments, a nucleic acid (e.g., a guide RNA, a nucleic acid comprising a nucleotide sequence encoding a guide RNA; a nucleic acid encoding a site-specific modifying enzyme such as an engineered B-GEN polypeptide of the disclosure; *etc.*) includes a modification or sequence that provides for an additional desirable feature (e.g., modified or regulated stability; subcellular targeting; tracking, e.g., a fluorescent label; a binding site for a protein or protein complex; *etc.*). Non-limiting examples include: a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (e.g., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and/or protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (e.g., a hairpin); a modification or sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule,

conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, *etc.*); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[0147] In some embodiments, a guide RNA includes an additional segment at either the 5' or 3' end that provides for any of the features described above. For example, a suitable third segment can include a 5' cap (e.g., a 7-methylguanylate cap (m7G)); a 3' polyadenylated tail (e.g., a 3' poly(A) tail); a riboswitch sequence (e.g., to allow for regulated stability and/or regulated accessibility by proteins and protein complexes); a stability control sequence; a sequence that forms a dsRNA duplex (e.g., a hairpin); a sequence that targets the RNA to a subcellular location (e.g., nucleus, mitochondria, chloroplasts, and the like); a modification or sequence that provides for tracking (e.g., direct conjugation to a fluorescent molecule, conjugation to a moiety that facilitates fluorescent detection, a sequence that allows for fluorescent detection, *etc.*); a modification or sequence that provides a binding site for proteins (e.g., proteins that act on DNA, including transcriptional activators, transcriptional repressors, DNA methyltransferases, DNA demethylases, histone acetyltransferases, histone deacetylases, and the like); and combinations thereof.

[0148] Modifications can also or alternatively be used to decrease the likelihood or degree to which RNAs introduced into cells elicit innate immune responses. Such responses, which have been well characterized in the context of RNA interference (RNAi), including small-interfering RNAs (siRNAs), as described below and in the art, tend to be associated with reduced half-life of the RNA and/or the elicitation of cytokines or other factors associated with immune responses.

[0149] One or more types of modifications can also be made to RNAs encoding an engineered B-Gen polypeptide that are introduced into a cell, including, without limitation, modifications that enhance the stability of the RNA (such as by decreasing its degradation by RNases present in the cell), modifications that enhance translation of the resulting product (e.g., the endonuclease), and/or modifications that decrease the likelihood or degree to which the RNAs introduced into cells elicit innate immune responses. Combinations of modifications, such as the foregoing and others, can likewise be used. In

the case of an engineered B-Gen polypeptide, for example, one or more types of modifications can be made to guide RNAs (including those exemplified above), and/or one or more types of modifications can be made to RNAs encoding an engineered B-Gen polypeptide (including those exemplified above).

[0150] By way of illustration, guide RNAs used in the CRISPR-B-Gen system or other smaller RNAs can be readily synthesized by chemical means, enabling a number of modifications to be readily incorporated, as illustrated below and described in the art. While chemical synthetic procedures are continually expanding, purifications of such RNAs by procedures such as high-performance liquid chromatography (HPLC, which avoids the use of gels such as PAGE) tends to become more challenging as nucleic acid lengths increase significantly beyond a hundred or so nucleotides. One approach used for generating chemically modified RNAs of greater length is to produce two or more molecules that are ligated together. Much longer RNAs, such as those encoding a B-Gen.1, or B-Gen.1.2, or B-Gen.2 endonuclease, are more readily generated enzymatically. While fewer types of modifications are generally available for use in enzymatically produced RNAs, there are still modifications that can be used to, e.g., enhance stability, reduced the likelihood or degree of innate immune response, and/or enhance other attributes, as described further below and in the art; and new types of modifications are regularly being developed.

[0151] By way of illustration of various types of modifications, especially those used frequently with smaller chemically synthesized RNAs, modifications can include one or more nucleotides modified at the 2' position of the sugar, in some embodiments a 2'-O-alkyl, 2'-O-alkyl-O-alkyl or 2'-fluoro-modified nucleotide. In some embodiments, RNA modifications include 2'-fluoro, 2'-amino and 2' O-methyl modifications on the ribose of pyrimidines, basic residues or an inverted base at the 3' end of the RNA. Such modifications are routinely incorporated into oligonucleotides and these oligonucleotides have been shown to have a higher T_m (e.g., higher target binding affinity) than; 2'- deoxy oligonucleotides against a given target.

[0152] A number of nucleotide and nucleoside modifications have been shown to make the oligonucleotide into which they are incorporated more resistant to nuclease digestion than the native oligonucleotide; these modified oligonucleotides survive intact for a longer time than unmodified oligonucleotides. Specific examples of modified oligonucleotides include

those comprising modified backbones, for example, phosphorothioates, phosphotriesters, methyl phosphonates, short chain alkyl or cycloalkyl intersugar linkages or short chain heteroatomic or heterocyclic intersugar linkages. Some oligonucleotides are oligonucleotides with phosphorothioate backbones and those with heteroatom backbones, particularly CH₂-NH-O-CH₂, CH₂-N(CH₃)-O-CH₂ (known as a methylene(methylimino) or MMI backbone), CH₂-O-N(CH₃)-CH₂, CH₂-N(CH₃)-N(CH₃)-CH₂ and O-N(CH₃)-CH₂-CH₂ backbones; amide backbones (see De Mesmaeker *et al.*, 1995, *Acc. Chem. Res.*, 28:366-374); morpholino backbone structures (see Summerton and Weller, US Patent No. 5,034,506); peptide nucleic acid (PNA) backbone (wherein the phosphodiester backbone of the oligonucleotide is replaced with a polyamide backbone, the nucleotides being bound directly or indirectly to the aza nitrogen atoms of the polyamide backbone, see Nielsen *et al.*, 1991, *Science* 254:1497). Phosphorus-containing linkages include, but are not limited to, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'; see US patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050.

[0153] Morpholino-based oligomeric compounds are described in Braasch and Corey, *Biochemistry*, 41(14): 4503-4510 (2002); Genesis, Volume 30, Issue 3, (2001); Heasman, *Dev. Biol.*, 243:209-214 (2002); Nasevicius *et al.*, *Nat. Genet.*, 26:216-220 (2000); Lacenra *et al.*, *Proc. Natl. Acad. Sci.*, 97: 9591-9596 (2000); and US Patent No. 5,034,506, issued Jul. 23, 1991. Cyclohexenyl nucleic acid oligonucleotide mimetics are described in Wang *et al.*, *J. Am. Chem. Soc.*, 122: 8595-8602 (2000).

[0154] Modified oligonucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages,

mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts; see US patent nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; and 5,677,439, each of which is herein incorporated by reference.

[0155] One or more substituted sugar moieties can also be included, e.g., one of the following at the 2' position: OH, SH, SCH₃, F, OCN, OCH₃, OCH₃O(CH₂)_nCH₃, O(CH₂)_nNH₂ or O(CH₂)_nCH₃ where n is from 1 to 10; C₁ to C₁₀ lower alkyl, alkoxyalkoxy, substituted lower alkyl, alkaryl or aralkyl; Cl; Br; CN; CF₃; OCF₃; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; SOCH₃; SO₂CH₃; ONO₂; NO₂; N₃; NH₂; heterocycloalkyl; heterocycloalkaryl; aminoalkylamino; polyalkylamino; substituted silyl; an RNA cleaving group; a reporter group; an intercalator; a group for improving the pharmacokinetic properties of an oligonucleotide; or a group for improving the pharmacodynamic properties of an oligonucleotide and other substituents having similar properties. In some embodiments, a modification includes 2'-methoxyethoxy (2'-O-CH₂CH₂OCH₃, also known as 2'-O-(2-methoxyethyl)) (Martinet *a*, *Helv. Chim. Acta*, 1995, 78, 486). Other modifications include 2'-methoxy (2'-O-CH₃), 2'-propoxy (2'-OCH₂CH₂CH₃) and 2'-fluoro (2'-F). Similar modifications may also be made at other positions on the oligonucleotide, particularly the 3' position of the sugar on the 3' terminal nucleotide and the 5' position of 5' terminal nucleotide. Oligonucleotides may also have sugar mimetics such as cyclobutyls in place of the pentofuranosyl group. In some embodiments, both a sugar and an internucleoside linkage, e.g., the backbone, of the nucleotide units are replaced with novel groups. The base units are maintained for hybridization with an appropriate nucleic acid target compound. One such oligomeric compound, an oligonucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA compounds, the sugar-

backbone of an oligonucleotide is replaced with an amide containing backbone, for example, an aminoethylglycine backbone. The nucleobases are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative United States patents that teach the preparation of PNA compounds include, but are not limited to, US patent nos. 5,539,082; 5,714,331; and 5,719,262. Further teaching of PNA compounds can be found in Nielsen *et al.*, Science, 254: 1497-1500 (1991).

[0156] Guide RNAs can also include, additionally or alternatively, nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include adenine (A), guanine (G), thymine (T), cytosine (C) and uracil (U). Modified nucleobases include nucleobases found only infrequently or transiently in natural nucleic acids, e.g., hypoxanthine, 6-methyladenine, 5-Me pyrimidines, particularly 5-methylcytosine (also referred to as 5-methyl-2' deoxycytosine and often referred to in the art as 5-Me-C), 5-hydroxymethylcytosine (HMC), glycosyl HMC and gentobiosyl HMC, as well as synthetic nucleobases, e.g., 2-aminoadenine, 2-(methylamino)adenine, 2-(imidazolylalkyl)adenine, 2-(aminoalkylamino)adenine or other heterosubstituted alkyladenines, 2-thiouracil, 2-thiothymine, 5-bromouracil, 5-hydroxymethyluracil, 8-azaguanine, 7-deazaguanine, N6 (6-aminohexyl)adenine and 2,6-diaminopurine. Kornberg, A, DNA Replication, W. H. Freeman & Co., San Francisco, pp75-77 (1980); Gebeyehu *et al.*, Nucl. Acids Res. 15:4513 (1997). A "universal" base known in the art, e.g., inosine, can also be included. 5-Me-C substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 degrees C. (Sanghvi, Y. S., in Crooke, S. T. and Lebleu, B., eds., Antisense Research and Applications, CRC Press, Boca Raton, 1993, pp. 276-278) and are embodiments of base substitutions.

[0157] Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl uracil and cytosine, 6-azo uracil, cytosine and thymine, 5-uracil (pseudo-uracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other a-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylquanine and 7-

methyladenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine.

[0158] Other useful nucleobases include those disclosed in United States Patent No. 3,687,808, those disclosed in "The Concise Encyclopedia of Polymer Science And Engineering", pages 858-859, Kroschwitz, J.I., ed. John Wiley & Sons, 1990, those disclosed by Englisch *et al.*, Angewandte Chemie, International Edition, 1991, 30, page 613, and those disclosed in Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S.T. and Lebleu, B. ea., CRC Press, 1993. Certain of these nucleobases are particularly useful for increasing the binding affinity of the oligomeric compounds of the disclosure. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and -O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2 oc (Sanghvi, Y.S., Crooke, S.T. and Lebleu, B., eds, "Antisense Research and Applications", CRC Press, Boca Raton, 1993, pp. 276-278) and are embodiments of base substitutions, even more particularly when combined with 2'-O-methoxyethyl sugar modifications. Modified nucleobases are described in US patent nos. 3,687,808, as well as 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,596,091; 5,614,617; 5,681,941; 5,750,692; 5,763,588; 5,830,653; 6,005,096; and US Patent Application Publication 20030158403.

[0159] It is not necessary for all positions in a given oligonucleotide to be uniformly modified, and in fact more than one of the aforementioned modifications may be incorporated in a single oligonucleotide or even at within a single nucleoside within an oligonucleotide.

[0160] In some embodiments, the guide RNAs and/or mRNA encoding an endonuclease such as B-GEN.1, or B-GEN.1.2, or B-GEN.2 of the disclosure are capped using any one of current capping methods such as mCAP, ARCA or enzymatic capping methods to create viable mRNA constructs that remain biologically active and avoid self/non-self intracellular responses. In some embodiments, the guide RNAs and/or mRNA encoding an endonuclease such as B-GEN.1, or B-GEN.1.2, or B-GEN.2 of the disclosure are capped by using a CleanCap™ (TriLink) co-transcriptional capping method.

[0161] In some embodiments, the guide RNAs and/or mRNA encoding an endonuclease of the disclosure includes one or more modifications selected from the group consisting of pseudouridine, N1-methylpseudouridine, and 5-methoxyuridine. In some embodiments, one or more N1-methylpseudouridines are incorporated into the guide RNAs and/or mRNA encoding an endonuclease of the disclosure in order to provide enhanced RNA stability and/or protein expression and reduced immunogenicity in animal cells, such as mammalian cell (e.g., human and mice). In some embodiments, the N1-methylpseudouridine modifications are incorporated in combination with one or more 5-methylcytidines.

[0162] In some embodiments, the guide RNAs and/or mRNA (or DNA) encoding an endonuclease such as B-GEN.1, or B-GEN.1.2, or B-GEN.2 are chemically linked to one or more moieties or conjugates that enhance the activity, cellular distribution, or cellular uptake of the oligonucleotide. Such moieties include but are not limited to, lipid moieties such as a cholesterol moiety (Letsinger *et al.*, 1989, Proc. Natl. Acad. Sci. USA 86: 6553-6556); cholic acid (Manoharan *et al.*, 1994, Bioorg. Med. Chem. Lett. 4: 1053-1060); a thioether, e.g., hexyl-S-tritylthiol (Manoharan *et al.*, 1992, Ann. N. Y Acad. Sci. 660:306-309 and Manoharan *et al.*, 1993, Bioorg. Med. Chem. Lett. 3:2765-2770); a thiocholesterol (Oberhauser *et al.*, 1992, Nucl. Acids Res. 20: 533-538); an aliphatic chain, e.g., dodecandiol or undecyl residues (Kabanov *et al.*, 1990, FEBS Lett., 259: 327-330 and Svinarchuk *et al.*, 1993, Biochimie, 75: 49-54); a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan *et al.*, 1995, Tetrahedron Lett. 36:3651-3654 and Shea *et al.*, 1990, Nucl. Acids Res. 18: 3777-3783); a polyamine or a polyethylene glycol chain (Mancharan *et al.*, 1995, Nucleosides & Nucleotides 14:969-973); adamantane acetic acid (Manoharan *et al.*, 1995, Tetrahedron Lett. 36:3651-3654); a palmityl moiety (Mishra *et al.*, 1995, Biochim. Biophys. Acta 1264:229-237); or an octadecylamine or hexylamino-carbonyl-t oxysterol moiety (Crooke *et al.*, 1996, J. Pharmacol. Exp. Ther., 277: 923-937). See also US Patent Nos. 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717; 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335; 4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241; 5,391,723; 5,416,203; 5,451,463; 5,510,475; 5,512,667;

5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 and 5,688,941.

[0163] Sugars and other moieties can be used to target proteins and complexes including nucleotides, such as cationic polysomes and liposomes, to particular sites. For example, hepatic cell directed transfer can be mediated via asialoglycoprotein receptors (ASGPRs); see, e.g., Hu, *et al.*, 2014, Protein Pept Lett. 21(10):1025-30. Other systems known in the art and regularly developed can be used to target biomolecules of use in the present case and/or complexes thereof to particular target cells of interest.

[0164] These targeting moieties or conjugates can include conjugate groups covalently bound to functional groups such as primary or secondary hydroxyl groups. Suitable conjugate groups include intercalators, reporter molecules, polyamines, polyamides, polyethylene glycols, polyethers, groups that enhance the pharmacodynamic properties of oligomers, and groups that enhance the pharmacokinetic properties of oligomers. Typical conjugate groups include cholesterol, lipids, phospholipids, biotin, phenazine, folate, phenanthridine, anthraquinone, acridine, fluoresceins, rhodamines, coumarins, and dyes. Groups that are capable of enhancing the pharmacodynamic properties include groups that improve uptake, enhance resistance to degradation, and/or strengthen sequence-specific hybridization with the target nucleic acid. Groups that are capable of enhancing the pharmacokinetic properties include groups that improve uptake, distribution, metabolism or excretion of the compounds of the present disclosure. Representative conjugate groups are disclosed in International Patent Application No. PCT/US92/09196, filed Oct. 23, 1992, and US Patent No. 6,287,860, which are incorporated herein by reference. Conjugate moieties include, but are not limited to, lipid moieties such as a cholesterol moiety, cholic acid, a thioether, e.g., hexyl-5-tritylthiol, a thiocholesterol, an aliphatic chain, e.g., dodecandiol or undecyl residues, a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H- phosphonate, a polyamine or a polyethylene glycol chain, or adamantane acetic acid, a palmityl moiety, or an octadecylamine or hexylamino-carbonyloxy cholesterol moiety. See, e.g., US Patent Nos. 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717; 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335;

4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241, 5,391,723; 5,416,203, 5,451,463; 5,510,475; 5,512,667; 5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 and 5,688,941.

[0165] Longer nucleic acids that are less amenable to chemical synthesis and are generally produced by enzymatic synthesis can also be modified by various means. Such modifications can include, for example, the introduction of certain nucleotide analogs, the incorporation of particular sequences or other moieties at the 5' or 3' ends of molecules, and other modifications. By way of illustration, the mRNA encoding B-GEN.1, or B-GEN.1.2, or B-GEN.2 is approximately 4kb in length and can be synthesized by in vitro transcription. Modifications to the mRNA can be applied to, e.g., increase its translation or stability (such as by increasing its resistance to degradation with a cell), or to reduce the tendency of the RNA to elicit an innate immune response that is often observed in cells following introduction of exogenous RNAs, particularly longer RNAs such as that encoding B-GEN.1 or B-GEN.2.

[0166] Numerous such modifications have been described in the art, such as polyA tails, 5' cap analogs (e.g., Anti Reverse Cap Analog (ARCA) or m7G(5')ppp(5')G (mCAP)), modified 5' or 3' untranslated regions (UTRs), use of modified bases (such as Pseudo-UTP, 2-Thio-UTP, 5-Methylcytidine-5'-Triphosphate (5-Methyl-CTP) or N6-Methyl-ATP), or treatment with phosphatase to remove 5' terminal phosphates. These and other modifications are known in the art, and new modifications of RNAs are regularly being developed.

[0167] There are numerous commercial suppliers of modified RNAs, including for example, TriLink Biotech, Axolabs, Bio-Synthesis Inc., Dharmacon and many others. As described by TriLink, for example, 5-Methyl-CTP can be used to impart desirable characteristics such as increased nuclease stability, increased translation or reduced interaction of innate immune receptors with in vitro transcribed RNA. 5'-Methylcytidine-5'-Triphosphate (5-Methyl-CTP), N6-Methyl-ATP, as well as Pseudo-UTP and 2-Thio-UTP, have also been shown to reduce innate immune stimulation in culture and in vivo while enhancing translation as illustrated in publications by Konmann *et al.* and Warren *et al.* referred to below.

[0168] It has been shown that chemically modified mRNA delivered in vivo can be used to achieve improved therapeutic effects; see, e.g., Kormann *et al.*, Nature Biotechnology 29, 154-157 (2011). Such modifications can be used, for example, to increase the stability of the RNA molecule and/or reduce its immunogenicity. Using chemical modifications such as Pseudo-U, N6-Methyl-A, 2-Thio-U and 5-Methyl-C, it was found substituting just one quarter of the uridine and cytidine residues with 2-Thio-U and 5-Methyl-C respectively, resulted in a significant decrease in toll-like receptor (TLR) mediated recognition of the mRNA in mice. By reducing the activation of the innate immune system, these modifications can therefore be used to effectively increase the stability and longevity of the mRNA in vivo; see, e.g., Konmann *et al.*, supra.

[0169] It has also been shown that repeated administration of synthetic messenger RNAs incorporating modifications designed to bypass innate anti-viral responses can reprogram differentiated human cells to pluripotency. See, e.g., Warren, *et al.*, Cell Stem Cell, 7(5):618-30 (2010). Such modified mRNAs that act as primary reprogramming proteins can be an efficient means of reprogramming multiple human cell types. Such cells are referred to as induced pluripotency stem cells (iPSCs), and it was found that enzymatically synthesized RNA incorporating 5-Methyl-CTP, Pseudo- UTP and an Anti Reverse Cap Analog (ARCA) could be used to effectively evade the cell's antiviral response; see, e.g., Warren *et al.*, supra. Other modifications of nucleic acids described in the art include, for example, the use of polyA tails, the addition of 5' cap analogs (such as m7G(5')ppp(5')G (mCAP)), modifications of 5' or 3' untranslated regions (UTRs), or treatment with phosphatase to remove 5' terminal phosphates-and new approaches are regularly being developed.

[0170] A number of compositions and techniques applicable to the generation of modified RNAs for use herein have been developed in connection with the modification of RNA interference (RNAi), including small-interfering RNAs (siRNAs). siRNAs present particular challenges in vivo because their effects on gene silencing via mRNA interference are generally transient, which can require repeat administration. In addition, siRNAs are double-stranded RNAs (dsRNA) and mammalian cells have immune responses that have evolved to detect and neutralize dsRNA, which is often a by-product of viral infection. Thus, there are mammalian enzymes such as PKR (dsRNA-responsive kinase), and potentially retinoic acid-inducible gene I (RIG-I), that can mediate cellular responses to dsRNA, as well as Toll-like

receptors (such as TLR3, TLR7 and TLR8) that can trigger the induction of cytokines in response to such molecules; see, e.g., the reviews by Angart *et al.*, *Pharmaceuticals* (Basel) 6(4): 440-468 (2013); Kanasty *et al.*, *Molecular Therapy* 20(3): 513-524 (2012); Burnett *et al.*, *Biotechnol J.* 6(9):1130-46 (2011); Judge and MacLachlan, *Hum Gene Ther* 19(2):111-24 (2008); and references cited therein.

[0171] A large variety of modifications have been developed and applied to enhance RNA stability, reduce innate immune responses, and/or achieve other benefits that can be useful in connection with the introduction of nucleic acids into human cells as described herein; see, e.g., the reviews by Whitehead KA *et al.*, *Annual Review of Chemical and Biomolecular Engineering*, 2:77-96 (2011); Gaglione and Messere, *Mini Rev Med Chem*, 10(7):578-95 (2010); Chernolovskaya *et al.*, *Curr Opin Mol Ther.*, 12(2):158-67 (2010); Deleavey *et al.*, *Curr Protoc Nucleic Acid Chem Chapter 16:Unit 16.3* (2009); Behlke, *Oligonucleotides* 18(4):305-19 (2008); Fucini *et al.*, *Nucleic Acid Ther* 22(3): 205-210 (2012); Bremsen *et al.*, *Front Genet* 3:154 (2012).

[0172] As noted above, there are a number of commercial suppliers of modified RNAs, many of which have specialized in modifications designed to improve the effectiveness of siRNAs. A variety of approaches are offered based on various findings reported in the literature. For example, Dharmacon notes that replacement of a non-bridging oxygen with sulfur (phosphorothioate, PS) has been extensively used to improve nuclease resistance of siRNAs, as reported by Kale, *Nature Reviews Drug Discovery* 11:125-140 (2012). Modifications of the 2'-position of the ribose have been reported to improve nuclease resistance of the internucleotide phosphate bond while increasing duplex stability (T_m), which has also been shown to provide protection from immune activation. A combination of moderate PS backbone modifications with small, well-tolerated 2'-substitutions (2'-O-, 2'-Fluoro, 2'-Hydro) has been associated with highly stable siRNAs for applications in vivo, as reported by Soutschek *et al.* *Nature* 432:173-178 (2004); and 2'-O-Methyl modifications have been reported to be effective in improving stability as reported by Volkov, *Oligonucleotides* 19:191-202 (2009). With respect to decreasing the induction of innate immune responses, modifying specific sequences with 2'-O-Methyl, 2'-Fluoro, 2'-Hydro have been reported to reduce TLR7/TLR8 interaction while generally preserving silencing activity; see, e.g., Judge *et al.*, *Mol. Ther.* 13:494-505 (2006); and Cekaite *et al.*, *J. Mol. Biol.* 365:90-

108 (2007). Additional modifications, such as 2-thiouracil, pseudouracil, 5-methylcytosine, 5-methyluracil, and N6-methyladenosine have also been shown to minimize the immune effects mediated by TLR3, TLR7, and TLR8; see, e.g., Kariko *et al.*, Immunity 23:165-175 (2005).

[0173] As is also known in the art, and commercially available, a number of conjugates can be applied to nucleic acids such as RNAs for use herein that can enhance their delivery and/or uptake by cells, including for example, cholesterol, tocopherol and folic acid, lipids, peptides, polymers, linkers and aptamers; see, e.g., the review by Winkler, Ther. Deliv. 4:791-809 (2013), and references cited therein.

6.9. Vectors

[0174] The present disclosure provides vectors comprising the nucleic acids of the disclosure, for example as described in Section 6.8. In some embodiments, the nucleic acid comprises a nucleic acid encoding an engineered B-GEN polypeptide as described in Section 6.2. In some embodiments, the engineered B-GEN polypeptide coding sequence is codon-optimized, at least for the portion encoding the nuclease component of the engineered B-GEN polypeptide.

[0175] The vector (or nucleotide sequence) may further encode a gRNA.

[0176] In some embodiments, the vector comprising the nucleotide sequence may be an expression vector.

[0177] In some embodiments the expression vector is production vector for an engineered B-GEN polypeptide, for example useful for the expression / production of the engineered B-GEN polypeptide in host cell. Following expression / production of the engineered B-GEN polypeptide in a host cell, the engineered B-GEN polypeptide can be incorporated into an RNP for nucleofection of a target cell.

[0178] Alternatively, the expression vector comprising the nucleotide sequence may be a delivery vector for an engineered B-GEN polypeptide, for example useful for introduction of the engineered B-GEN polypeptide coding sequence into a target cell intended for gene editing. Following expression / production of the engineered B-GEN polypeptide in the target cell, the engineered B-GEN polypeptide, together with guide RNA molecules, is capable of editing the target cell. In some embodiments, a delivery vector further includes

coding sequences for the gRNAs. In other embodiments, a separate nucleic acid encoding the gRNA is introduced into the target cell.

[0179] Expression vectors contemplated include, but are not limited to, viral vectors based on vaccinia virus, poliovirus, adenovirus, adeno-associated virus, SV40, herpes simplex virus, human immunodeficiency virus, retrovirus (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus) and other recombinant vectors. Other vectors contemplated for eukaryotic target cells include, but are not limited to, the vectors pXT1, pSG5, pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). Additional vectors contemplated for eukaryotic cells include, but are not limited to, the vectors pCTx-1, pCTx-2, and pCTx-3. Other vectors can be used so long as they are compatible with the intended host or target cell.

[0180] In some embodiments, an expression vector has one or more transcription and/or translation control elements. Depending on the expression cell/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. can be used in the vector. The vector can also contain a ribosome binding site for translation initiation and a transcription terminator.

[0181] Non-limiting examples of suitable eukaryotic promoters (i.e., promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, human elongation factor-1 promoter (EF1), a hybrid construct having the cytomegalovirus (CMV) enhancer fused to the chicken beta-actin promoter (CAG), murine stem cell virus promoter (MSCV), phosphoglycerate kinase-1 locus promoter (PGK), and mouse metallothionein-I.

[0182] In some embodiments, a promoter is an inducible promoter (e.g., a heat shock promoter, tetracycline-regulated promoter, steroid-regulated promoter, metal-regulated promoter, estrogen receptor-regulated promoter, etc.). In some embodiments, a promoter is a constitutive promoter (e.g., CMV promoter, UBC promoter). In some embodiments, the promoter is a spatially restricted and/or temporally restricted promoter (e.g., a tissue specific

promoter, a cell type specific promoter, etc.). In some embodiments, a vector does not have a promoter for at least one gene to be expressed in a host cell if the gene is going to be expressed, after it is inserted into a genome, under an endogenous promoter present in the genome.

[0183] For expressing small RNAs, including guide RNAs, various promoters such as RNA polymerase III promoters, including for example U6 and H1, can be advantageous. Thus, such promoters may advantageously be incorporated into delivery vectors. Descriptions of and parameters for enhancing the use of such promoters are known in art, and additional information and approaches are regularly being described; see, e.g., Ma, H. *et al.*, Molecular Therapy - Nucleic Acids 3, e161 (2014) doi:10.1038/mtna.2014.12.

[0184] In some embodiments, the vector is a self-inactivating vector that either inactivates the viral sequences or the components of the CRISPR machinery or other elements. Self-inactivating vectors are particularly useful for delivery vectors, to select against cells that retain engineered B-GEN polypeptide coding sequences after gene editing is complete.

[0185] In some embodiments, the expression vectors are RNA vectors. In other embodiments, the expression vectors are DNA vectors.

6.9.1. RNA Vectors

[0186] The expression vectors of the disclosure may be RNA vectors.

[0187] Particularly suitable vectors are RNA virus-based virus replicons, such as alphaviruses and Paramyxoviruses. Alphavirus and Paramyxovirus replicons do not involve a DNA intermediate for replication and thus provide a safer alternative to several other commonly used viral vectors including lentiviral and retroviral vectors (Yoshioka *et al.*, 2013, Cell Stem Cell. 13(2):246-54; Yoshioka and Dowdy, 2017, PLOS ONE 12:e0182018). Alphaviruses are lipid-enveloped, positive-sense RNA viruses, which constitute a genus of more than 30 viruses in the Togaviridae family, including Eastern, Western, and Venezuelan equine encephalitis viruses (EEEV, WEEV, and VEEV, respectively), chikungunya (CHIK), Sindbis, Ross River, and O'nyong-nyong viruses, among others. Sendai viruses (SeV) are enveloped, single-stranded negative-sense Paramyxoviruses that replicate episomally in a host cell cytoplasm.

[0188] Accordingly, in some embodiments, an RNA vector is derived from an RNA virus, such as an alphavirus, a paramyxovirus, a flavivirus, a rhabdovirus, a measles virus, or a picornavirus.

[0189] In some embodiments, the RNA vector is a single stranded RNA replicon. In some embodiments, the single stranded RNA replicon is a positive strand. In some other embodiments, the single stranded RNA replicon is a negative strand. In some embodiments the RNA vector comprises one or more coding sequences for one or more engineered B-Gen polypeptides and a self-replication element.

[0190] RNA replicons of the disclosure typically include regulatory elements, a subgenomic (SG) promoter, operably linked to the engineered B-Gen polypeptide coding sequence(s). The sequences comprising engineered B-Gen polypeptide coding sequence are typically flanked by 5' and 3' UTR sequence, and the 3' UTR sequence is typically followed by a polyadenylation signal.

[0191] The RNA vector construct may be produced from a DNA template (, a DNA plasmid construct). By way of example, the RNA construct may be transcribed from a DNA template by using a SP6 or T7 in vitro transcription kit.

[0192] RNA vectors are particularly useful as delivery vectors.

6.9.2. DNA Vectors

[0193] In some embodiments, an expression vector of the disclosure is a DNA vector. The present disclosure provides two types of DNA vectors: (1) a DNA vector that is a production vector or delivery vector and (2) a DNA vector from which an RNA vector of the disclosure (as described in Section 6.9.1) can be transcribed, as described in Section 6.9.2.2. The DNA vector from which an RNA replicon of the disclosure can be transcribed is sometimes referred to herein as a "template vector".

[0194] In some embodiments, the DNA vector of the disclosure is a nonintegrating DNA vector. For example, the vector can be an episomal vector. For instance, a number of DNA viruses, such as adenoviruses, Simian vacuolating virus 40 (SV40), bovine papilloma virus (BPV), or budding yeast ARS (Autonomously Replicating Sequences)-containing plasmids may be used without genomic integration.

[0195] In some embodiments, the DNA vectors of the disclosure include an origin of replication. Examples of origins of replications that may be incorporated into a DNA vector of the disclosure include the replication origin of a lymphotropic herpes virus, a gammaherpesvirus, an adenovirus, a bovine papilloma virus, or a yeast. In some embodiments, the replication origin is from a lymphotropic herpes virus or a gammaherpesvirus corresponding to oriP of EBV, as a self-replication element. In some embodiments, the lymphotropic herpes virus is Epstein Barr virus (EBV), Kaposi's sarcoma herpes virus (KSHV), Herpes virus saimiri (HS), or Marek's disease virus (MDV). Epstein Barr virus (EBV) and Kaposi's sarcoma herpes virus (KSHV) are also examples of a gammaherpesvirus.

[0196] In certain embodiments, a vector of the disclosure comprises a replication origin of EBV, *OriP*. *OriP* is the site at or near which DNA replication initiates and is composed of two *cis*-acting sequences approximately 1 kilobase pair apart known as the family of repeats (FR) and the dyad symmetry (DS). FR is composed of 21 imperfect copies of a 30 bp repeat and contains 20 high affinity EBNA-1-binding sites. When FR is bound by EBNA-1, it both serves as a transcriptional enhancer of promoters in *cis* up to 10 kb away. DS is sufficient for initiation of DNA synthesis in the presence of EBNA-1 and initiation occurs either at or near DS.

[0197] One or more of the expression cassettes in a replicating DNA vector may further comprise a nucleotide sequence encoding a trans-acting factor that binds to the replication origin to replicate an extra-chromosomal template. Alternatively or additionally, the somatic cell may express such a trans-acting factor.

[0198] In other embodiments, the DNA vectors of the disclosure lack an origin of replication.

[0199] The DNA vectors of the disclosure typically comprise one or more promoters, SP6 or T7, to drive expression of the engineered B-GE_n polypeptide in the case of DNA vectors that are intended to be production vectors or expression of an RNA replicons in the case of DNA vectors that are intended to be template vectors.

6.9.2.1. Expression Vectors

[0200] In some embodiments, the expression vector is a DNA vector comprising expression cassettes for expression of one or more proteins of interest, operably linked to a regulatory element comprising a promoter suitable for driving expression of the engineered B-GE_n

polypeptide in the cell type of interest. Examples of promoters suitable for driving expression of proteins in mammalian cells include cytomegalovirus (CMV) promoters, EF1a promoters, SV40 promoter, Ubc promoter, human beta actin promoters, PGK1 promoters and CAG promoters.

[0201] DNA vectors that are used for direct expression of an engineered B-GEN polypeptide (rather than as a template for expression of an RNA vector as described in Section 6.9.2.2) need not include RNA replicon self-replication sequences, for example the nsP1-nsP4 proteins of VEEV or NP, P and L proteins of Sendai virus.

[0202] In some embodiments, a DNA expression vector is a non-replicating DNA vector. In some embodiments, a DNA expression vector is a replicating DNA vector.

6.9.2.2. Template Vectors

[0203] The DNA vectors of the disclosure may also serve as templates for transcription of an RNA replicon as described herein. Thus, the “expression cassettes” included in template vectors are intended for transcription from the RNA replicon produced by transcription of the RNA replicon.

[0204] Accordingly, the template vectors of the disclosure comprise a nucleotide sequence encoding an RNA replicon as described herein under the control of a regulatory element, e.g., the SP6 or T7 promoter.

[0205] In some embodiments, a template DNA vector is a non-replicating DNA vector. In some embodiments, a template DNA vector is a replicating DNA vector.

[0206] In some embodiments, the template vectors are used for in vitro transcription of an RNA replicon that is subsequently introduced into a cell to drive expression of the engineered B-GEN polypeptide.

6.9.3. Viral Vectors

[0207] A recombinant adeno-associated virus (AAV) vector may be used for delivery. Known techniques to produce rAAV particles in the art is to provide a cell with a polynucleotide to be delivered between two AAV invert terminal repeats (ITRs), AAV rep and cap genes and helper virus functions. Production of rAAV requires that the following components are present within a single cell (denoted herein as a packaging cell): a polynucleotide of interest between two ITRs, AAV rep and cap genes separate from (i.e., not

in) the AAV genome, and helper virus functions. The AAV rep and cap genes may be from any AAV serotype for which recombinant virus can be derived and may be from a different serotype of AAV than that of ITRs on a packaged polynucleotide, including, but not limited to, AAV serotypes AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-7, AAV-8, AAV-9, AAV-10, AAV-11, AAV-12, AAV-13 and AAV rh.74. Production of pseudotyped rAAV is disclosed in, for example, WO 01/83692.

AAV Serotype	Genbank Accession No.
AAV-1	NC_002077.1
AAV-2	NC_001401.2
AAV-3	NC_001729.1
AAV-38	AF028705.1
AAV-4	NC_001829.1
AAV-5	NC_006152.1
AAV-6	AF028704.1
AAV-7	NC_006260.1
AAV-8	NC_006261.1
AAV-9	AX753250.1
AAV-10	AY631965.1
AAV-11	AY631966.1
AAV-12	00813647.1
AAV-13	EU285562.1

[0208] A method of generating a packaging cell is to create a cell line that stably expresses all the necessary components for AAV particle production. For example, a plasmid (or multiple plasmids) comprising a polynucleotide of interest between AAV ITRs, AAV rep and cap genes separate from the AAV genome, and a selectable marker, such as a neomycin

resistance gene, are integrated into the genome of a cell. AAV genomes have been introduced into bacterial plasmids by procedures such as GC tailing (Samulski et al., 1982, Proc. Natl. Acad. Sci. USA, 79:2077-2081), addition of synthetic linkers containing restriction endonuclease cleavage sites (Laughlin et al., 1983, Gene, 23:65-73) or by direct, blunt-end ligation (Senapathy & Carter, 1984, J. Biol. Chem., 259:4661-4666). The packaging cell line is then infected with a helper virus such as adenovirus. The advantages of this method are that the cells are selectable and are suitable for large-scale production of rAAV. Other examples of suitable methods employ adenovirus or baculovirus rather than plasmids to introduce rAAV genomes and/or rep and cap genes into packaging cells.

[0209] General principles of rAAV production are reviewed in, for example, Carter, 1992, Current Opinions in Biotechnology, 1533-539; and Muzyczka, 1992, Curr. Topics in Microbial. and Immunol., 158:97-129). Various approaches are described in Ratschin et al., Mol. Cell. Biol. 4:2072 (1984); Hermonat et al., Proc. Natl. Acad. Sci. USA, 81:6466 (1984); Tratschin et al., Mol. Cell. Biol. 5:3251 (1985); McLaughlin et al., J. Virol., 62:1963 (1988); and Lebkowski et al., 1988 Mol. Cell. Biol., 7:349 (1988). Samulski et al. (1989, J. Virol., 63:3822-3828); US Patent No. 5,173,414; WO 95/13365 and corresponding US Patent No. 5,658,776; WO 95/13392; WO 96/17947; PCT/US98/18600; WO97/09441 (PCT/US96/14423); WO 97/08298 (PCT/US96/13872); WO 97/21825 (PCT/US96/20777); WO 97/06243 (PCT/FR96/01064); WO 99/11764; Perrin et al. (1995) Vaccine 13:1244-1250; Paul et al. (1993) Human Gene Therapy 4:609-615; Clark et al. (1996) Gene Therapy 3:1124-1132; US Patent. No. 5,786,211; US Patent No. 5,871,982; and US Patent. No. 6,258,595.

[0210] AAV vector serotypes used for transduction are dependent on target cell types. For example, the following exemplary cell types are known to be transduced by the indicated AAV serotypes among others.

Tissue/Cell Type	Serotype
Liver	AAV8, AAV9
Skeletal muscle	AAV1, AAV7, AAV6, AAV8, AAV9
Central nervous system	AAV5, AAV1, AAV4

Tissue/Cell Type	Serotype
RPE	AAV5, AAV4
Photoreceptor cells	AAV5
Lung	AAV9
Heart	AAV8
Pancreas	AAV8
Kidney	AAV2

[0211] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

6.10. Host Cells and Recombinant Expression

[0212] In some embodiments, host cells may be employed to express gRNAs, sgRNAs, or the engineered B-GE_n polypeptides of the disclosure. Suitable host cells include naturally occurring cells; genetically modified cells (e.g., cells genetically modified in a laboratory), and cells manipulated in vitro in any way. In some embodiments, a host cell is isolated.

[0213] The host cell can be a eukaryote or prokaryote and includes, for example, yeast (such as *Pichia pastoris* or *Saccharomyces cerevisiae*), bacteria (such as *E. coli* or *Bacillus subtilis*), insect Sf9 cells (such as baculovirus-infected SF9 cells) or mammalian cells (such as Human Embryonic Kidney (HEK) cells, Chinese hamster ovary cells, HeLa cells, human 293 cells and monkey COS-7 cells).

[0214] Host cells may be from established cell lines, or they may be primary cells, where “primary cells”, “primary cell lines”, and “primary cultures” are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, e.g., splittings, of the culture. For example, primary cultures include cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5

times, 10 times, or 15 times, but not enough times go through the crisis stage. Primary cell lines can be maintained for fewer than 10 passages in vitro. In some embodiments, host cells are PSCs (e.g., iPSCs, or ESCs), or PSC-derived cells (e.g., PSC-derived neurons, PSC-derived microglial cells, PSC-derived cardiomyocytes, PSC-derived cells of the eye).

[0215] If the cells are primary cells, such cells may be harvested from an individual by any suitable method. An appropriate solution may be used for dispersion or suspension of the harvested cells. The harvested cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will generally be frozen in 10% dimethyl sulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

6.11. Target Cells

[0216] In some embodiments, the engineered Type V endonuclease or B-GE_n CRISPR-Cas system is introduced into target cells or populations of target cells. Methods for introducing proteins and nucleic acids to target cells are described further in Section 6.12.

[0217] The target cells and target cell populations of the disclosure can be cells in which gene editing by the systems of the disclosure has taken place, or cells in which the components of a system of the disclosure have been introduced or expressed but gene editing has not taken place, or a combination thereof. In various embodiments, a cell population can comprise, for example, a population in which at least 1%, at least 5%, at least 10%, at least 15%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, or at least 70% of the cells have undergone gene editing by a system of the disclosure.

[0218] In some embodiments, the methods of the disclosure may be employed to induce transcriptional modulation in mitotic or post-mitotic cells in vivo and/or ex vivo and/or in vitro. In some embodiments, the methods of the disclosure may be employed to induce DNA cleavage, DNA modification, and/or transcriptional modulation in mitotic or post-mitotic cells in vivo and/or ex vivo and/or in vitro (e.g., to produce genetically modified cells that can be reintroduced into an individual).

[0219] Because the guide RNA provides specificity by hybridizing to target DNA, a mitotic and/or post-mitotic cell can be any of a variety of target cell, and modified in vivo or in vitro. Suitable target cells include, but are not limited to, a bacterial cell; an archaeal cell; a single-celled eukaryotic organism; a plant cell; an algal cell, e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like; a fungal cell; an animal cell; a cell from an invertebrate animal (e.g., an insect, a cnidarian, an echinoderm, a nematode, etc.); a eukaryotic parasite (e.g., a malarial parasite, e.g., *Plasmodium falciparum*; a helminth; etc.); a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal); a mammalian cell, e.g., a rodent cell, a human cell, a non-human primate cell, etc. In some embodiments, the target cell can be any human cell. Suitable target cells include naturally occurring cells; genetically modified cells (e.g., cells genetically modified in a laboratory, e.g., by the “hand of man”); and cells manipulated in vitro in any way. In some embodiments, a target cell is isolated.

[0220] Any type of cell may be of interest as a host cell or target cell for in vivo or in vitro modification. In various embodiments, the host cell or target cell is a stem cell (e.g., PSC such as an embryonic stem (ES) cell or an induced pluripotent stem cell (iPSC)), a germ cell; a somatic cell, e.g., a fibroblast, a hematopoietic cell, an immune cell (e.g., a T-lymphocyte, a B-lymphocyte, a dendritic cell, or a macrophage), a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.). Cells may be from established cell lines, or they may be primary cells, where “primary cells”, “primary cell lines”, and “primary cultures” are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, e.g., splittings, of the culture. For example, primary cultures include cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Primary cell lines can be maintained for fewer than 10 passages in vitro. Target cells are, in some embodiments, unicellular organisms, or are grown in culture. In some embodiments, a host cell is the same as a target cell. In some embodiments, a target cell is modified to become another cell type such that the resultant host cell is different from the target cell. As an example, a target cell may be a PSC (e.g., iPSC) which is then differentiated into a PSC-derived cell (such as a

PSC-derived neuron) such that the host cell is a neuron. Alternatively, a target cell is not a PSC, but is subsequently de-differentiated or reprogrammed into a PSC.

[0221] If the cells are primary cells, such cells may be harvested from an individual by any suitable method. For example, leukocytes may be suitably harvested by apheresis, leukocytophoresis, density gradient separation, *etc.*, while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, *etc.* are most suitably harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution will generally be a balanced salt solution, e.g., normal saline, phosphate-buffered saline (PBS), Hank's balanced salt solution, *etc.*, suitably supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, e.g., from 5-25 mM. Suitable buffers include HEPES, phosphate buffers, lactate buffers, *etc.* The cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will generally be frozen in 10% dimethyl sulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

[0222] In some embodiments, target cells are located in a subject and the methods of the disclosure comprise administering B-GEN CRISPR-Cas system (e.g., a ribonucleoprotein comprising an engineered B-GEN polypeptide of the disclosure) to a subject, e.g., a mammalian subject such as a human or livestock subject. Exemplary *in vivo* cells to which a B-GEN CRISPR-Cas system can be delivered include but are not limited to fibroblasts, hematopoietic cells, immune cells (e.g., T-lymphocytes, B-lymphocytes, dendritic cells, or macrophages), neurons, neuroglia, muscle cells (e.g., cardiomyocytes), bone cells, hepatocytes, and pancreatic cells.

6.11.1. Pluripotent Stem Cells (PSCs)

[0223] In some embodiments, the target cells are stem cells, for example particularly pluripotent stem cells (PSCs) such as induced pluripotent stem cells (iPSCs) or human embryonic stem cells (hESCs), which can be differentiated and used to generate large numbers of a specific cell type that can be delivered for regenerative medicine in patients

with many different diseases. Differentiation, in the context of PSCs, is the process of lineage specification and can be achieved using cell specific protocols.

[0224] Following modification of a PSC, e.g., an hESC or an iPSC, by introduction of an engineered Type V endonuclease or B-GEN polypeptide of the disclosure, the PSC can be differentiated into a cell type of interest for cell therapy. In some embodiments, the PSC's genome is edited by the engineered B-GEN polypeptide of the disclosure prior to differentiation.

[0225] The PSCs, e.g., PSCs comprising an engineered Type V endonuclease or B-GEN polypeptide of the disclosure or whose genomes have been edited by an engineered Type V endonuclease or B-GEN polypeptide of the disclosure, can be differentiated into cells suitable for therapy, including the cells in the endoderm (e.g., lung, thyroid, or pancreatic cells, or progenitors thereof), ectoderm (e.g., skin, neuronal, or pigment cells, or progenitors thereof) and mesoderm (e.g., cardiac cells, skeletal muscle cells, red blood cells, smooth muscle cells, or progenitors or precursors thereof) lineages.

[0226] In some embodiments, a PSC of the disclosure is differentiated into a cardiac cell. In various embodiments, the cardiac cell is a cardiac progenitor cell or a mature or immature (atrial or ventricular) cardiomyocyte.

[0227] In other embodiments, a PSC of the disclosure is differentiated into an oligodendrocyte progenitor cells or an oligodendrocyte.

[0228] In other embodiments, a PSC of the disclosure is differentiated into a neural lineage cell, for example a neural crest cells, an astrocyte, a dopaminergic neuron progenitor cell, a dopaminergic neuron cells, a midbrain dopaminergic neuron progenitor cell, a midbrain dopaminergic neuron, an authentic midbrain dopamine (DA) neuron, a dopaminergic neuron precursor cell, a floor plate midbrain progenitor cell, a floor plate midbrain DA neuron.

[0229] In other embodiments, a PSC of the disclosure is differentiated into a photoreceptor cell, a photoreceptor precursor cell, a retinal pigmented epithelium cell, a neural retinal cell, or a neural retinal progenitor cell.

[0230] In some embodiments, a PSC of the disclosure is differentiated into a microglial cell or a microglial progenitor cell.

[0231] In some embodiments, a PSC of the disclosure is differentiated into a macrophage.

[0232] In some embodiments, a PSC of the disclosure is differentiated into an enteric progenitor cell or an enteric cell.

[0233] In some embodiments, a PSC of the disclosure is differentiated into an immune cell, e.g., a T-lymphocyte, a B-lymphocyte, a dendritic cell, or a macrophage.

[0234] In some embodiments, the PSCs may be genetically engineered (e.g., to produce a functional protein that is defective in a patient, to produce a therapeutic protein, to include a shutoff switch, or to evade immune detection, thereby supporting allogeneic applications) prior to differentiation into a cell type of interest.

6.12. Methods of Introducing Nucleic Acids into Host and Target Cells

[0235] In some embodiments, the methods of the disclosure include introducing into a host or target cell (or a population of host or target cells) one or more nucleic acids comprising a nucleotide sequence encoding a guide RNA and/or a nucleotide sequence (e.g., a codon-optimized nucleotide sequence) encoding an engineered B-GEN polypeptide. In some embodiments, the methods of the disclosure comprise introducing into a host or target cell (or a population of host or target cells) a guide RNA and/or a nucleotide sequence (e.g., a codon-optimized nucleotide sequence) encoding an engineered B-GEN polypeptide.

[0236] In some embodiments, a target cell (e.g., a cell comprising DNA that is targeted by a guide RNA for editing by an engineered B-GEN polypeptide) is in a cell in vitro, e.g., in cell culture. In some embodiments, a target cell is in vivo, e.g., in a subject (e.g., a mammal such as a human) for whom gene therapy is intended.

[0237] In some embodiments, the nucleotide sequence encoding a guide RNA and/or an engineered B-GEN polypeptide is operably linked to an inducible promoter. In some embodiments, a nucleotide sequence encoding a guide RNA and/or an engineered B-GEN polypeptide is operably linked to a constitutive promoter.

[0238] A guide RNA, or a nucleic acid comprising a nucleotide sequence encoding same, can be introduced into a host or target cell by any of a variety of well-known methods. Similarly, where a method involves introducing into a host or target cell a nucleic acid comprising a nucleotide sequence (e.g., a codon-optimized nucleotide sequence) encoding an engineered B-GEN polypeptide, such a nucleic acid can be introduced into a host or target cell by any of a variety of well-known methods. Guide nucleic acids (RNA or DNA;

e.g., a guide RNA or one or more DNA molecules encoding a guide RNA) and/or engineered B-GEN polypeptide-encoding nucleic acids (RNA or DNA) can be delivered by viral or non-viral delivery vehicles known in the art.

[0239] Methods of introducing a nucleic acid into a host or target cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a stem cell or progenitor cell. Suitable methods include, e.g., viral or bacteriophage infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, nucleofection, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam *et al.*, *Adv Drug Deliv Rev.* 2012 Sep 13. pii: 50169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like, including but not limiting to exosome delivery.

[0240] Polynucleotide sequence (e.g., a) may be delivered by non-viral delivery vehicles including, but not limited to, nanoparticles, liposomes, ribonucleoproteins, positively charged peptides, small molecule RNA- conjugates, aptamer-RNA chimeras, and RNA-fusion protein complexes. Some exemplary non-viral delivery vehicles are described in Peer and Lieberman, *Gene Therapy*, 18: 1127-1133 (2011) (which focuses on non-viral delivery vehicles for siRNA that are also useful for delivery of other nucleic acids).

[0241] Suitable systems and techniques for delivering a nucleic acid of the disclosure (e.g., mRNA and sgRNA) for gene editing includes lipid nanoparticles (LNPs). As used herein, the term "lipid nanoparticles" includes liposomes irrespective of their lamellarity, shape or structure and lipoplexes as described for the introduction of nucleic acids and/or polypeptides into cells. These lipid nanoparticles can be complexed with biologically active compounds (e.g., nucleic acids and/or polypeptides) and are useful as in vivo delivery vehicles. In general, any method known in the art can be applied to prepare the lipid nanoparticles comprising one or more nucleic acids of the present disclosure and to prepare complexes of biologically active compounds and said lipid nanoparticles. Examples of such methods are widely disclosed, e.g., in *Biochim Biophys Acta* 1979, 557:9; *Biochim et Biophys Acta* 1980, 601:559; *Liposomes: A practical approach* (Oxford University Press, 1990); *Pharmaceutica Acta Helvetiae* 1995, 70:95; *Current Science* 1995, 68:715; *Pakistan*

Journal of Pharmaceutical Sciences 1996, 19:65; Methods in Enzymology 2009, 464:343). Particularly suitable systems and techniques for preparing LNP formulations comprising one or more nucleic acids and/or polypeptides of the present disclosure include, but are not limited to, those developed by Intellia (see e.g., WO2017173054A1), Alnylam (see, e.g., WO2014008334A1), Modernatx (see, e.g., WO2017070622A1 and WO2017099823A1), TranslateBio, Acuitas (see, e.g., WO2018081480A1), Genevant Sciences, Arbutus Biopharma, Tekmira, Arcturus, Merck (see, e.g., WO2015130584A2), Novartis (see, e.g., WO2015095340A1), and Dicerna; all of which are herein incorporated by reference in their entireties.

[0242] Suitable nucleic acids comprising nucleotide sequences encoding an engineered B-GE_n polypeptide and/or a guide RNA include expression vectors. In some embodiments, the expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., US Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, a recombinant retroviral construct, etc. Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g. Li *et al.*, Invest Ophthalmol Vis Sci 35:2543 2549, 1994; Borrás *et al.*, Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto *et al.*, Hum Gene Ther 5:10881097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g. Ali *et al.*, Hum Gene Ther 9:81 86, 1998, Flannery *et al.*, PNAS 94:6916 6921, 1997; Bennett *et al.*, Invest Ophthalmol Vis Sci 38:2857 2863, 1997; Jomary *et al.*, Gene Ther 4:683-690, 1997, Rolling *et al.*, Hum Gene Ther 10:641 648, 1999; Ali *et al.*, Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski *et al.*, J. Vir. (1989) 63:3822-3828; Mendelson *et al.*, Viral. (1988) 166:154-165; and Flotte *et al.*, PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g. Miyoshi *et al.*, PNAS 94:10319 23, 1997; Takahashi *et al.*, J Virol 73:7812 7816, 1999); a retroviral vector (e.g. Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[0243] In some embodiments, a B-Gen ribonucleoprotein comprising a B-Gen endonuclease and an sgRNA is delivered to target cells through nucleofection, which is a method of delivering nucleic acid to the cells by the use of cell-specific reagents and electrical parameters to create transient small pores in the cell membrane.

[0244] Engineered B-Gen Type V CRISPR-Cas systems may be delivered into target cells by delivery vectors, such as viral vectors. Engineered B-Gen Type V CRISPR-Cas systems may also be delivered into target cells by non-viral delivery vehicles including, but not limited to, nanoparticles, liposomes, ribonucleoproteins, positively charged peptides, small molecule RNA- conjugates, aptamer-RNA chimeras, and RNA-fusion protein complexes. Some exemplary non-viral delivery vehicles are described in Peer and Lieberman, Gene Therapy, 18: 1127-1133 (2011).

[0245] In some embodiments, an engineered B-Gen Type V CRISPR-Cas system is delivered into target cells via the use of delivery vectors as described in Section 6.9.

6.13. Methods of Gene Editing

[0246] The disclosure also provides methods of gene editing using an engineered Type V endonuclease or B-Gen polypeptide system. The methods of gene editing may comprise a method for targeting, editing, modifying, or manipulating a target DNA at one or more locations in the genome of a host or target cell (or a multitude of host or target cells), whether in vitro, ex vivo or in vivo, or a target DNA in a cell-free environment. Typically, the methods of gene editing comprise introducing an engineered Type V endonuclease or B-Gen polypeptide system into a host or target cell (or a population of host or target cells) or a target DNA sequence comprising cell-free environment under conditions that are suitable for the engineered Type V endonuclease or B-Gen polypeptide to make one or more modifications (e.g., nicks or cuts or base edits) in the target DNA, wherein the engineered Type V endonuclease or B-Gen polypeptide is directed to the target DNA by the guide RNA in its processed or unprocessed form.

[0247] Methods of gene editing may comprise introducing an engineered Type V endonuclease or B-Gen polypeptide system of the disclosure into a host or target cell (or a population of host or target cells) in an RNP complex and/or introducing (e.g., via a virus such as AAV or via a LNP) one or more nucleic acids comprising a guide RNA or a nucleotide sequence encoding a guide RNA and a nucleotide sequence (e.g., a codon-

optimized nucleotide sequence) encoding an engineered Type V endonuclease or B-GEN polypeptide into a host or target cell (or a population of host or target cells). An engineered Type V endonuclease or B-GEN polypeptide system (e.g., in an RNP complex; one or more nucleic acid molecules encoding a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide; or a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide) of the disclosure can be introduced into a host or target cell (or a population of host or target cells) by any of a variety of well-known viral or non-viral delivery methods. The methods gene editing may be employed in a host or target cell (or a population of host or target cells) in vivo, ex vivo or in vitro.

[0248] In some embodiments, the methods of gene editing comprise introducing an engineered Type V endonuclease or B-GEN polypeptide system (e.g., in an RNP complex; one or more nucleic acid molecules encoding a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide; or a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide) of the disclosure to make one or more modifications (e.g., nicks or cuts or base edits) in the target DNA in vivo, in a subject (e.g., a mammalian subject such as a human subject) in which gene editing, e.g., for gene therapy purposes, is desirable.

[0249] In some embodiments, the methods of gene editing comprise introducing an engineered Type V endonuclease or B-GEN polypeptide system (e.g., in an RNP complex; one or more nucleic acid molecules encoding a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide; or a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide) of the disclosure to make one or more modifications (e.g., nicks or cuts or base edits) in the target DNA ex vivo.

[0250] In some embodiments, the methods of gene editing comprise introducing an engineered Type V endonuclease or B-GEN polypeptide system (e.g., in an RNP complex; one or more nucleic acid molecules encoding a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide; or a guide RNA and one or more nucleotide sequences (e.g., one or more codon-optimized nucleotide sequences) encoding an engineered Type V endonuclease or B-GEN polypeptide) of the disclosure to make one or more modifications (e.g., nicks or cuts or base edits) in the target DNA in vitro.

[0251] In some embodiments, methods of gene editing comprise contacting a cell with an engineered Type V endonuclease or B-GEN polypeptide system in the form of an RNP complex comprising the engineered Type V or B-GEN polypeptide and a guide RNA. Illustrative details regarding the preparation, composition, and delivery of RNP complexes are described in Section 6.7. In some embodiments, an RNP complex can also be delivered into cell using delivery methods that increase the porosity of the plasma membrane of the cell, e.g., via electroporation or nucleofection.

[0252] In some embodiments, methods of gene editing comprise contacting a cell with an engineered Type V endonuclease or B-GEN polypeptide system in the form of a LNP, e.g., a LNP comprising an RNP (e.g., an RNP as described in the preceding paragraph) or an LNP comprising a nucleic acid encoding the engineered Type V endonuclease or B-GEN polypeptide and a guide RNA or a nucleic acid encoding the guide RNA. Any method known in the art can be applied to prepare LNPs, for example as described in Section 6.12.

[0253] In some embodiments, methods of gene editing comprise contacting a cell with an engineered Type V endonuclease or B-GEN polypeptide system in the form of one or more viruses, e.g., one or more AAVs whose genome(s) comprise one or more nucleotide sequences encoding the engineered Type V endonuclease or B-GEN polypeptide and a nucleic acid encoding the guide RNA. The use of viruses to introduce transgenes, e.g., nucleic acids encoding the engineered Type V endonuclease or B-GEN polypeptide and guide RNA coding sequences, is known in the art and described in Section 6.9.3. Other methods of, and delivery vectors for, introducing nucleic acids encoding an engineered Type V endonuclease or B-GEN polypeptide or guide RNA are described in Sections 6.9 and 6.12.

6.14. Pharmaceutical Compositions

[0254] Also disclosed herein are pharmaceutical formulations and medicaments comprising a B-GEEn protein, gRNA, nucleic acid or plurality of nucleic acids, system, particle, or plurality of particles of the disclosure together with a pharmaceutically acceptable excipient.

[0255] Suitable excipients include, but are not limited to, salts, diluents, (*e.g.*, Tris-HCl, acetate, phosphate), preservatives (*e.g.*, Thimerosal, benzyl alcohol, parabens), binders, fillers, solubilizers, disintegrants, sorbents, solvents, pH modifying agents, antioxidants, antinfective agents, suspending agents, wetting agents, viscosity modifiers, tonicity agents, stabilizing agents, and other components and combinations thereof. Suitable pharmaceutically acceptable excipients can be selected from materials which are generally recognized as safe (GRAS) and may be administered to an individual without causing undesirable biological side effects or unwanted interactions. Suitable excipients and their formulations are described in Remington's Pharmaceutical Sciences, 16th ed. 1980, Mack Publishing Co. In addition, such compositions can be complexed with polyethylene glycol (PEG), metal ions, or incorporated into polymeric compounds such as polyacetic acid, polyglycolic acid, hydrogels, etc., or incorporated into liposomes, microemulsions, micelles, unilamellar or multilamellar vesicles, erythrocyte ghosts or spheroblasts. Suitable dosage forms for administration, *e.g.*, parenteral administration, include solutions, suspensions, and emulsions.

[0256] The components of the pharmaceutical formulation can be dissolved or suspended in a suitable solvent such as, for example, water, Ringer's solution, phosphate buffered saline (PBS), or isotonic sodium chloride. The formulation may also be a sterile solution, suspension, or emulsion in a nontoxic, parenterally acceptable diluent or solvent such as 1,3-butanediol.

[0257] In some cases, formulations can include one or more tonicity agents to adjust the isotonic range of the formulation. Suitable tonicity agents are well known in the art and include glycerin, mannitol, sorbitol, sodium chloride, and other electrolytes. In some cases, the formulations can be buffered with an effective amount of buffer necessary to maintain a pH suitable for parenteral administration. Suitable buffers are well known by those skilled in the art and some examples of useful buffers are acetate, borate, carbonate, citrate, and phosphate buffers.

[0258] In some embodiments, the formulation can be distributed or packaged in a liquid form, or alternatively, as a solid, obtained, for example by lyophilization of a suitable liquid formulation, which can be reconstituted with an appropriate carrier or diluent prior to administration. In some embodiments, the formulations can comprise a guide RNA and a Type II Cas protein in a pharmaceutically effective amount sufficient to edit a gene in a cell. The pharmaceutical compositions can be formulated for medical and/or veterinary use.

[0259] In some embodiments, the B-GEN endonuclease complexes may be introduced into host cells, e.g., iPSCs, to produce genetically modified cells that can be reintroduced into an individual. The iPSC-derived cells described herein may be provided in a pharmaceutical composition containing the cells and a pharmaceutically acceptable carrier. The pharmaceutically acceptable carrier may be cell culture medium that optionally does not contain any animal-derived component. For storage and transportation, the cells may be cryopreserved at $< -70^{\circ}\text{C}$ (e.g., on dry ice or in liquid nitrogen). Prior to use, the cells may be thawed, and diluted in a sterile cell medium that is supportive of the cell type of interest.

[0260] The cells may be administered into the patient systemically (e.g., through intravenous injection or infusion), or locally (e.g., through direct injection to a local tissue, e.g., the heart, the brain, and a site of damaged tissue). Various methods are known in the art for administering cells into a patient's tissue or organs, including, without limitation, intracoronary administration, intramyocardial administration, transendocardial administration, or intracranial administration.

[0261] A therapeutically effective number of iPSC-derived cells are administered to the patient. As used herein, the term "therapeutically effective" refers to a number of cells or amount of pharmaceutical composition that is sufficient, when administered to a human subject suffering from or susceptible to a disease, disorder, and/or condition, to treat, prevent, and/or delay the onset or progression of the symptom(s) of the disease, disorder, and/or condition. It will be appreciated by those of ordinary skill in the art that a therapeutically effective amount is typically administered via a dosing regimen comprising at least one-unit dose. In some embodiments, at least 10^3 (e.g., at least 10^4 , at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , at least 10^9 , at least 10^{10} , at least 10^{11} , or at least 10^{12}) cells are administered to a subject at a time in one or more sites. In some embodiments, 10^3 - 10^{18} (e.g., 10^3 - 10^4 , 10^3 - 10^5 , 10^3 - 10^6 , 10^3 - 10^7 , 10^3 - 10^8 , 10^3 - 10^9 , 10^3 - 10^{10} , 10^3 - 10^{11} , 10^3 -

10^{12} , 10^6 - 10^7 , 10^6 - 10^8 , 10^6 - 10^9 , 10^6 - 10^{10} , 10^6 - 10^{11} , 10^6 - 10^{12} , 10^9 - 10^{10} , 10^9 - 10^{11} , 10^9 - 10^{12}) cells are administered to a subject at a time in one or more sites. In some embodiments, more than 10^{12} (e.g., more than 10^{12} , more than 10^{13} , more than 10^{14} , more than 10^{15} , more than 10^{16} , more than 10^{17} , more than 10^{18} or more) cells are administered to a subject at a time at one or more sites.

7. NUMBERED EMBODIMENTS

[0262] While various specific embodiments have been illustrated and described, it will be appreciated that various changes can be made without departing from the spirit and scope of the disclosure(s). The present disclosure is exemplified by the numbered embodiments set forth below. Unless otherwise specified, features of any of the concepts, aspects and/or embodiments described in the detailed description above are applicable mutatis mutandis to any of the following numbered embodiments.

1. A polypeptide comprising an amino acid other than aspartic acid at the position corresponding to D504 of SEQ ID NO:1 (B-Gen.1) or D501 of SEQ ID NO:2 (B-Gen.1.2) or SEQ ID NO:3 (B-Gen.2).
2. The polypeptide of embodiment 1, which is an engineered Type V endonuclease polypeptide.
3. The polypeptide of embodiment 2, which is an engineered Bacillales Type V endonuclease polypeptide.
4. The polypeptide of any one of embodiments 1 to 3, which is a B-Gen polypeptide.
5. The polypeptide of any one of embodiments 1 to 4, which has an arginine at the position corresponding to D504 of SEQ ID NO:1 (B-Gen.1) or D501 of SEQ ID NO:2 (B-Gen.1.2) or SEQ ID NO:3 (B-Gen.2).
6. The polypeptide of any one of embodiments 1 to 5, which comprises a target-interacting sequence motif of any one of SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203; and SEQ ID NO:204.

7. The polypeptide of any one of embodiments 1 to 6, which comprises a RuvC I domain comprising an amino acid sequence having at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11.

8. The polypeptide of embodiment 7, wherein the RuvC I domain comprises an amino acid sequence having at least 40% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11.

9. The polypeptide of embodiment 7, wherein the RuvC I domain comprises an amino acid sequence having at least 70% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11.

10. The polypeptide of embodiment 7, wherein the RuvC I domain comprises an amino acid sequence having at least 80% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11.

11. The polypeptide of embodiment 7, wherein the RuvC I domain comprises an amino acid sequence having at least 90% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11.

12. The polypeptide of any one of embodiments 1 to 11, which comprises a RuvC II domain comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12.

13. The polypeptide of embodiment 12, wherein the RuvC II domain comprises an amino acid sequence having at least 60% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12.

14. The polypeptide of embodiment 12, wherein the RuvC II domain comprises an amino acid sequence having at least 70% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12.

15. The polypeptide of embodiment 12, wherein the RuvC II domain comprises an amino acid sequence having at least 80% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12.

16. The polypeptide of embodiment 12, wherein the RuvC II domain comprises an amino acid sequence having at least 90% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12.

17. The polypeptide of any one of embodiments 1 to 16, which comprises a RuvC III domain comprising an amino acid sequence having at least at least 80%, at least 85%, or at least 90% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13.

18. The polypeptide of embodiment 17, wherein the RuvC III domain comprises an amino acid sequence having at least 80% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13.

19. The polypeptide of embodiment 17, wherein the RuvC III domain comprises an amino acid sequence having at least 85% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13.

20. The polypeptide of embodiment 17, wherein the RuvC III domain comprises an amino acid sequence having at least 90% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13.

21. The polypeptide of any one of embodiments 1 to 20, which comprises the amino acid sequence of SEQ ID NO:8.

22. The polypeptide of any one of embodiments 1 to 21, which comprises the amino acid sequence of SEQ ID NO:9.

23. The polypeptide of any one of embodiments 1 to 22, which comprises the amino acid sequence of SEQ ID NO:10.

24. The polypeptide of any one of embodiments 1 to 20, which comprises the amino acid sequence of SEQ ID NO:11.

25. The polypeptide of any one of embodiments 1 to 20 and 24, which comprises the amino acid sequence of SEQ ID NO:12.

26. The polypeptide of any one of embodiments 1 to 20 and 24 to 25, which comprises the amino acid sequence of SEQ ID NO:13.

27. A polypeptide, which is optionally a polypeptide according to any one of embodiments 1 to 26, comprising an amino acid sequence which:

- (a) has a substitution at the position corresponding to D501 of SEQ ID NO:3, wherein the substitution:
 - (i) is the substitution D501R as compared to the amino acid sequence of SEQ ID NO:3; or
 - (ii) increases gene editing efficiency activity as compared to a corresponding amino acid sequence without the substitution at position D501;

and

- (b) has:
 - (i) at least 80%, at least 85%, at least 90% or at least 95% sequence identity to SEQ ID NO:1 (B-GEN.1), SEQ ID NO:2 (B-GEN.1.2), or SEQ ID NO:3 (B-GEN.2);
 - (ii) (A) a target-interacting sequence motif of any one of SEQ ID NO:201, SEQ ID NO:202, SEQ ID NO:203; and SEQ ID NO:204; (B) a RuvC I domain comprising an amino acid sequence having at least 40%, at least 45%, at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC I domain of SEQ ID NO:8 or SEQ ID NO:11; (C) a RuvC II domain comprising an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80% or at least 90% sequence identity to the RuvC II domain of SEQ ID NO:9 or SEQ ID NO:12; (D) a RuvC III domain comprising an amino acid sequence having at

least at least 80%, at least 85%, or at least 90% sequence identity to the RuvC III domain of SEQ ID NO:10 or SEQ ID NO:13; or (E) any combination of two, three, or all four of (A), (A), (C) and (D);

- (iii) up to 25 amino acid insertions, substitutions and/or deletions as compared to the amino acid sequence of SEQ ID NO:1 (B-GEN.1), SEQ ID NO:2 (B-GEN.1.2), or SEQ ID NO:3 (B-GEN.2);
- (iv) (b)(i) and (b)(ii);
- (v) (b)(i) and (b)(iii);
- (vi) (b)(ii) and (b)(iii); or
- (vii) (b)(i), (b)(ii), and (b)(iii).

28. The polypeptide of any one of embodiments 1 to 27, which has a gene editing efficiency that is at least 50% higher than the gene editing efficiency of a corresponding polypeptide lacking the amino acid substitution at the position corresponding to D504 of SEQ ID NO:1 (B-GEN.1) or D501 of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2).

29. The polypeptide of any one of embodiments 1 to 27, which has a gene editing efficiency that is at least 70% higher than the gene editing efficiency of a corresponding polypeptide lacking the amino acid substitution at the position corresponding to D504 of SEQ ID NO:1 (B-GEN.1) or D501 of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2).

30. The polypeptide of any one of embodiments 1 to 27, which has a gene editing efficiency that is at least 90% higher than the gene editing efficiency of a corresponding polypeptide lacking the amino acid substitution at the position corresponding to D504 of SEQ ID NO:1 (B-GEN.1) or D501 of SEQ ID NO:2 (B-GEN.1.2) or SEQ ID NO:3 (B-GEN.2).

31. The polypeptide of any one of embodiments 28 to 30, wherein the gene editing efficiency is assessed via *in cellula* gene editing assay, optionally wherein the gene editing assay is as describe in Example 1.

32. The polypeptide of any one of embodiments 27 to 31, wherein the sequence identity is to SEQ ID NO:1 and the amino acid insertions, substitutions and/or deletions are in relation to SEQ ID NO:1.

33. The polypeptide of embodiment 32, wherein the amino acid sequence has at least 98% sequence identity to the amino acid sequence of SEQ ID NO:1.

34. The polypeptide of embodiment 32, wherein the amino acid sequence has at least 99% sequence identity to the amino acid sequence of SEQ ID NO:1.

35. The polypeptide of embodiment 32, wherein the amino acid sequence has at least 99.5% sequence identity to the amino acid sequence of SEQ ID NO:1.

36. The polypeptide of any one of embodiments 27 to 31, wherein the sequence identity is to SEQ ID NO:2 and the amino acid insertions, substitutions and/or deletions are in relation to SEQ ID NO:2.

37. The polypeptide of embodiment 36, wherein the amino acid sequence has at least 98% sequence identity to the amino acid sequence of SEQ ID NO:2.

38. The polypeptide of embodiment 36, wherein the amino acid sequence has at least 99% sequence identity to the amino acid sequence of SEQ ID NO:2.

39. The polypeptide of embodiment 36, wherein the amino acid sequence has at least 99.5% sequence identity to the amino acid sequence of SEQ ID NO:2.

40. The polypeptide of any one of embodiments 27 to 31, wherein the sequence identity is to SEQ ID NO:3 and the amino acid insertions, substitutions and/or deletions are in relation to SEQ ID NO:3.

41. The polypeptide of embodiment 40, wherein the amino acid sequence has at least 98% sequence identity to the amino acid sequence of SEQ ID NO:3.

42. The polypeptide of embodiment 40, wherein the amino acid sequence has at least 99% sequence identity to the amino acid sequence of SEQ ID NO:3.

43. The polypeptide of embodiment 40, wherein the amino acid sequence has at least 99.5% sequence identity to the amino acid sequence of SEQ ID NO:3.

44. A polypeptide comprising the amino acid sequence of SEQ ID NO:4.

45. A polypeptide comprising the amino acid sequence of SEQ ID NO:5.

46. A polypeptide comprising the amino acid sequence of SEQ ID NO:6.

47. The polypeptide of any one of embodiments 1 to 46, which further comprises at least one nuclear localization signal ("NLS").

48. The polypeptide of embodiment 47, which comprises at least one NLS located C-terminal to the amino acid sequence, optionally wherein:

- (a) the polypeptide lacks any NLS N-terminal to the amino acid sequence; or
- (b) the polypeptide comprises at least one NLS located N-terminal to the amino acid sequence.

49. The polypeptide of embodiment 47 or embodiment 48, which lacks any NLS N-terminal to the amino acid sequence.

50. The polypeptide of any one of embodiments 47 to 49, which comprises a linker sequence between the amino acid sequence and each NLS.

51. The polypeptide of any one of embodiments 47 to 50 wherein each NLS comprises an amino acid sequence independently selected from an NLS sequence set forth in Table 2.

52. The polypeptide of any one of embodiments 47 to 51, wherein each linker sequence is independently selected from a linker sequence set forth in Table 3.

53. A nucleic acid comprising a nucleotide sequence encoding the polypeptide of any one of embodiments 1 to 52.

54. The nucleic acid of embodiment 53, wherein the nucleotide sequence encoding the polypeptide of any one of embodiments 1 to 52 is operably linked to a promoter.
55. The nucleic acid of embodiment 53 or embodiment 54, wherein the nucleic acid further encodes a guide RNA.
56. The nucleic acid of any one of embodiments 53 to 55, which is in the form of a vector.
57. The nucleic acid of embodiment 56, wherein the vector is an expression vector.
58. The nucleic acid of embodiment 57, wherein the expression vector is a production vector.
59. The nucleic acid of embodiment 57, wherein the expression vector is a delivery vector.
60. The nucleic acid of any one of embodiments 56 to 59, wherein the vector is an RNA vector.
61. The nucleic acid of any one of embodiments 56 to 59, wherein the vector is a DNA vector.
62. The nucleic acid of embodiment 61, wherein the DNA vector is a plasmid.
63. A cell comprising the nucleic acid of any one of embodiments 53 to 62.
64. A cell engineered to express a nucleotide sequence encoding the polypeptide of any one of embodiments 1 to 52.
65. The cell of embodiment 63 or embodiment 64, which is a eukaryotic cell.
66. The cell of embodiment 65, which is an insect cell.
67. The cell of embodiment 65, which is a plant cell.

68. The cell of embodiment 65, which is a mammalian cell.
69. The cell of embodiment 68, which is a human cell.
70. A method of producing a polypeptide according to any one of embodiments 1 to 52, comprising culturing the cell of any one of embodiments 63 to 69 under conditions in which the polypeptide is produced.
71. The method of embodiment 70, which further comprises isolating and/or purifying the polypeptide.
72. A composition comprising:
- (a) a polypeptide according to any one of embodiments 1 to 52; and
 - (b) a guide RNA.
73. The composition of embodiment 72, which is a ribonucleoprotein complex.
74. The composition of embodiment 72 or embodiment 73, in which the polypeptide:guide RNA molar ratio ranges from 1:1 to 1:4.
75. The composition of embodiment 72 or embodiment 73, in which the polypeptide:guide RNA molar ratio ranges from 1:1 to 1:3.
76. The composition of embodiment 72 or embodiment 73, in which the polypeptide:guide RNA molar ratio ranges from 1:1.5 to 1:2.5.
77. The composition of embodiment 72 or embodiment 73, in which the polypeptide:guide RNA molar ratio is 1:2.
78. A method of editing the genome of a cell, comprising introducing into the cell:
- (a) the polypeptide of any one of embodiments 1 to 52; and
 - (b) a guide RNA.
79. A method of editing the genome of a cell, comprising introducing into the cell one or more nucleic acids encoding:

- (a) the polypeptide of any one of embodiments 1 to 52; and
- (b) a guide RNA,

optionally wherein at least one of the one or more nucleic acids is a nucleic acid according to any one of embodiments 53 to 62.

80. A method of editing the genome of a cell, comprising introducing into the cell:

- (a) one or more nucleic acids encoding the polypeptide of any one of embodiments 1 to 52; and
- (b) a guide RNA.

81. The method of embodiment 80, which comprises contacting the cell a lipid nanoparticle comprising the one or more nucleic acids and the guide RNA.

82. A method of editing the genome of a cell, comprising introducing into the cell one or more nucleic acids encoding:

- (a) the polypeptide of any one of claims 1 to 52; and
- (b) a guide RNA,

optionally wherein at least one of the one or more nucleic acids is a nucleic acid according to any one of embodiments 53 to 62.

83. The method of embodiment 82, which comprises contacting the cell with one or more recombinant AAV particles comprising the one or more nucleic acids.

84. The method of embodiment 82, which comprises contacting the cell with one or more lipid nanoparticles comprising the one or more nucleic acids.

85. A method of editing the genome of a cell, comprising introducing into the cell the composition of any one of embodiments 72 to 77.

86. The method of any one of embodiments 78 to 80, wherein the cell is a mammalian cell.

87. The method of embodiment 86, wherein the mammalian cell is a human cell.

88. The method of embodiment 86 or embodiment 87, wherein the mammalian cell is an immune cell, optionally selected from a T cell, a T cell expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and an immunosuppressive macrophage.

89. The method of any one of embodiments 78 to 87, wherein the cell is a hematopoietic stem cell, erythroid progenitor cell, lymphoid progenitor cell, peripheral blood mononuclear cell, T lymphocyte, B lymphocyte, macrophage, monocyte, neutrophil, eosinophil, dendritic cell, or a cell reprogrammed therefrom.

90. The method of any one of embodiments 78 to 89, wherein the cell is a stem cell or a cell differentiated therefrom.

91. The method of embodiment 90, wherein the stem cell is a pluripotent stem cell (PSC) or a cell differentiated therefrom.

92. The method of embodiment 90 or embodiment 91, wherein the cell differentiated therefrom is a human immune cell, optionally selected from a T cell, a T cell expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and an immunosuppressive macrophage.

93. The method of embodiment 90 or embodiment 91, wherein the cell differentiated therefrom is a cell in the human nervous system, optionally selected from dopaminergic neuron, a microglial cell, an oligodendrocyte, an astrocyte, a cortical neuron, a spinal or oculomotor neuron, an enteric neuron, a Placode-derived cell, a Schwann cell, and a trigeminal or sensory neuron.

94. The method of embodiment 90 or embodiment 91, wherein the cell differentiated therefrom is a cell in the human cardiovascular system, optionally selected from a cardiomyocyte, an endothelial cell, and a nodal cell.

95. The method of embodiment 90 or embodiment 91, wherein the cell differentiated therefrom is a cell in the human metabolic system, optionally selected from a hepatocyte, a cholangiocyte, and a pancreatic beta cell.

96. The method of embodiment 90 or embodiment 91, wherein the cell differentiated therefrom is a cell in the human ocular system, optionally selected from a retinal pigment epithelial cell, a photoreceptor cone cell, a photoreceptor rod cell, a bipolar cell, or a ganglion cell

97. The method of any one of embodiments 78 to 80, wherein the cell is a plant cell.

98. A cell comprising:

- (a) the composition of any one of embodiments 72 to 77; or
- (b) a nucleic acid according to any one of any one of embodiments 53 to 62.

99. The cell of embodiment 98, which is a mammalian cell.

100. The cell of embodiment 99, wherein the mammalian cell is a human cell.

101. The cell of embodiment 99 or embodiment 100, wherein the mammalian cell is an immune cell, optionally selected from a T cell, a T cell expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and an immunosuppressive macrophage.

102. The cell of any one of embodiments 98 to 100, which is a hematopoietic stem cell, erythroid progenitor cell, lymphoid progenitor cell, peripheral blood mononuclear cell, T lymphocyte, B lymphocyte, macrophage, monocyte, neutrophil, eosinophil, dendritic cell, or a cell reprogrammed therefrom.

103. The cell of any one of embodiments 98 to 102, which is a stem cell or a cell differentiated therefrom.

104. The cell of embodiment 103, wherein the stem cell is a pluripotent stem cell (PSC) or a cell differentiated therefrom.

105. The cell of embodiment 103 or embodiment 104, wherein the cell differentiated therefrom is a human immune cell, optionally selected from a T cell, a T cell

expressing a chimeric antigen receptor (CAR) or recombinant TCR, a regulatory T cell, a myeloid cell, a dendritic cell, and an immunosuppressive macrophage.

106. The cell of embodiment 103 or embodiment 104, wherein the cell differentiated therefrom is a cell in the human nervous system, optionally selected from dopaminergic neuron, a microglial cell, an oligodendrocyte, an astrocyte, a cortical neuron, a spinal or oculomotor neuron, an enteric neuron, a Placode-derived cell, a Schwann cell, and a trigeminal or sensory neuron.

107. The cell of embodiment 103 or embodiment 104, wherein the cell differentiated therefrom is a cell in the human cardiovascular system, optionally selected from a cardiomyocyte, an endothelial cell, and a nodal cell.

108. The cell of embodiment 103 or embodiment 104, wherein the cell differentiated therefrom is a cell in the human metabolic system, optionally selected from a hepatocyte, a cholangiocyte, and a pancreatic beta cell.

109. The cell of embodiment 103 or embodiment 104, wherein the cell differentiated therefrom is a cell in the human ocular system, optionally selected from a retinal pigment epithelial cell, a photoreceptor cone cell, a photoreceptor rod cell, a bipolar cell, or a ganglion cell.

110. The cell of embodiment 98, which is a plant cell.

8. EXAMPLES

8.1. Materials and Methods

8.1.1. Structural Modeling of B-GEN.2

[0263] The structure of B-GEN.2 has not been characterized to date. AlphaFold2 software was used to predict the structure of B-GEN.2, and this was compared against the database of proteins with known crystal structures. Based on 14 lead crystal structure hits, a group of candidate structures for B-GEN.2 was generated. This data indicated that the predicted structure of B-GEN.2 aligned quite well with several crystal structures from the Bacillales order of Bacilli.

[0264] Based on the alignments, the OBD II and RuvC I domains of B-GEN.2 were used to carry out NCBI BLASTp searches within the order Bacillales. Approximately 20 Cas nucleases from related source organisms, having ~40 - 70 % sequence identity, were selected for multiple sequence alignment using the MUSCLE alignment algorithm. The phylogenetic relationships of these various nucleases are depicted in the tree of FIG. 2, whereas the alignment itself is shown in FIG. 4.

8.1.2. Designing B-GEN.2 Variants with Amino Acid Substitutions

[0265] Two parallel approaches were used to identify amino acid substitutions that might lead to enhanced indel activity of B-GEN.2. In the first approach, DNASTar Lasergene 17 software was used to identify amino acid residues in the AacC2c1 crystal structure that came within 3 angstroms of either the target DNA substrate, the guide spacer RNA, or the guide tracr RNA. The same exercise was conducted for the predicted structure of B-GEN.2, noting especially those instances where the selected amino acid residues differed from the corresponding residues of AacC2c1.

[0266] The second approach involved the use of Swiss-PdbViewer, aka Deepview (<https://spdbv.unil.ch/>) to identify amino acids within the crystal structure of AacC2c1 (RCSB entry 5U31) that interacted via predicted hydrogen bonding with either the target DNA substrate, the guide spacer RNA, or the guide tracr RNA, were identified. Then, by aligning the sequence of AacC2c1 with that of B-GEN.2, a search was conducted for instances where predicted protein-nucleic acid contacting residues within the AacC2c1 appeared to be different in B-GEN.2. To narrow the focus from potentially dozens of candidate mutants to a more manageable number for eventual protein production and purification, further evaluation was focused on Arg or Lys residues that formed close contacts with either the target or non-target DNA strand in the AacC2c1 crystal but did not correspond to Arg or Lys residues in the predicted B-GEN.2 structure.

8.1.3. Expression and Purification of B-GEN.2 Variants

[0267] Plasmids were constructed for expression of the selected point mutants of B-Gen.2. All constructs were expressed in BL21(DE3) cells by first chemically transforming the coding plasmids into these *E. coli* cells and plating on antibiotic-selection LB plates. Colonies were scraped and transferred to 250 mL MagicMedia (Thermo Scientific). Cultures were grown at 37 °C for 4 hours and then switched to 16 °C for 40 hours for protein expression. Cell pellets

were then harvested by spinning down at 5000 x g for 15 minutes at 4 °C and then frozen for later processing. To thaw, pellets were resuspending in lysis buffer (500 mM NaCl, 50 mM Tris, pH 8.0, 5 % glycerol, 5 mM EDTA, 0.5 mM TCEP), sonicated, and the resulting lysates were cleared by centrifugation at 50,000 x g for 30 minutes at 4 °C. Cleared lysates were loaded on Heparin Sepharose 6 FF columns in Loading buffer (100 mM NaCl, 50 mM Tris, pH 8.0, 5 % glycerol, 5 mM EDTA, 0.5 mM TCEP) followed by a wash step with the same buffer. Proteins were eluted using the same buffer with steps of 500 mM, 700, and 1000 mM NaCl. The protein in the main fraction was concentrated to ~20 mg/mL. Purity of proteins was estimated by band densitometry of Coomassie stained SDS-PAGE gels, and aliquots of purified protein were frozen at -80 °C for later use.

8.1.4. RNP Generation

[0268] Purified and concentrated variant B-GEN.2 nucleases were formed into RNPs with the guide RNA of the following sequence targeting a site within the B2M gene, wherein the spacer region is underlined:

*mU*mA*GCUAUAGGCUAAUAAGAUAGUUGUGUCAAGUGCUUCGGAGACCUAACACG
UCUCCAGUCACAACGGCUAAAAUAGCCAGCACAGUGUAGUACAAGAGAUAGA*mA*m
A*mG (SEQ ID NO:178)

[0269] RNPs were assembled by mixing the B2M-targeting guide RNA with nuclease in buffer containing 225 mM NaCl at a molar ratio of 2:1 and incubating at room temperature for ~ 30 minutes. The complexation efficiency of the RNPs was assessed via cation exchange UPLC on an Agilent Bio SCX (NP1.7, SS) column using a linear elution gradient of NaCl. Buffer A consisted of 10 mM Sodium Phosphate, 100 mM NaCl, pH 6.5, whereas Buffer B was 10 mM Sodium Phosphate, 1 M NaCl, pH 6.5. The percent efficiency of complexation was obtained by dividing the peak area of the RNP by the peak area of the total nuclease added (equivalent sample without guide RNA) and multiplying by 100. Values ranged from 70-90 % complexation, depending on the nuclease variant tested. The ability of the RNPs to cleave DNA in vitro was assessed using an in vitro plasmid cutting assay. Increasing quantities of RNPs were mixed in CutSmart Buffer (New England Biolabs) with a fixed quantity of B2M target site-containing plasmid that had been linearized using Xho-I enzyme. Following a 30 min incubation at 37 °C, the reaction was quenched for 10 min at 50 °C in the presence of Proteinase K. The samples were applied to a D5000 tape and

analyzed on an Agilent 4200 TapeStation instrument. Percent cleavage for each reaction was calculated as the peak area of the products over the peak area of the total substrate added (equivalent sample without RNP) and multiplying by 100. The EC₅₀ values, which ranged from 0.07 to 0.21 nM, were determined for each variant protein (or Cpf1 control) series using Prism software from GraphPad (Version 9.3.1).

8.1.5. *In Cellula* Gene Editing

[0270] iPSCs were cultured in substrate-coated T75 flasks with Essential 8 (E8) growth medium and maintained at 37 °C at 5% CO₂ level between passages. Passages were done at 75 to 80% confluency. On the day of nucleofection iPSCs were detached from the flask using Accutase™ cell detachment solution (Stem Cell Technologies) by incubating for 10 minutes at 37 °C and then quenching with equal volume of E8 medium. Cell pellets were harvested by centrifugation at 115xg for 3 minutes, followed by resuspension in Lonza P3 primary cell nucleofection buffer. Ribonucleoproteins (RNPs) were assembled with sgRNAs for each nuclease construct using 1:2 ratio of protein: sgRNA (IDT). Nucleofection of complexed RNP into P3-resuspended iPSCs was accomplished using LONZA 4D nucleofector. The nucleofected cells were then plated at 150,000 cells per well in substrate-coated 24-well Falcon flat bottom plates (Corning) and grown in E8 growth medium (with Rock inhibitor Y-27632 (Tocris)) for 72 to 96 hours. After harvesting, some of the cells were stained for flow cytometry using APC-conjugated antibody against B2M (for B2M-targeted experiments only) from BioLegend. The remaining cells were resuspended in 30 µL lysis buffer from BioRad's singleshots cell lysis kit and crude gDNA extraction done by incubation at room temperature for 10 minutes, then at 37 °C for 5 minutes, followed by proteinase K inactivation at 75 °C for 5 minutes. 1 µL of crude gDNA extract was directly used for every 25 µL of PCR reaction for amplicon sequencing first step, followed by end-prep and indexing and sequencing on an Illumina MiSeq sequencer.

8.2. Example 1: Design and Expression of Variant B-GEN.2 with Amino Acid Substitutions

[0271] A Type V endonuclease, B-GEN.2 (SEQ ID NO: 3) from a *Brevibacillus* species, was previously evaluated for gene editing activity in several proprietary iPSC lines. With the goal of improving indel formation, a plausible structure of the wild type B-GEN.2 was identified *in silico* as described on Section 8.1.1. Variant B-GEN.2 sequences with single amino acid point mutations were designed and generated as described in Section 8.1.2. Selected B-

GEn.2 variants were expressed in BL21(DE3) cells and purified as described in Section 8.1.3.

[0272] A lead plausible structure of B-GEn.2 was identified that aligned well with the previously published crystal structure of a Cas nuclease from *Alicyclobacillus acidoterrestris*, AacC2c1 (FIGs. 3A-3C) as well as a Cas nuclease from the organism *Geobacillus thermoleovorans* BthC2C1 (not shown).

[0273] Next, the amino acid sequences of AacC2c1 and B-GEn.2 were aligned to determine the amino acid differences in key locations. The sequence alignment between these two nucleases revealed about 37% amino acid sequence identity. Using the differences in AacC2c1 and B-GEn.2 amino acid sequences, non-Arg and non-Lys residues of the predicted B-GEn.2 structure corresponding to Arg or Lys residues in the AacC2c1 crystal structure that formed close contacts with target and non-target DNA were identified as mutation targets. An initial list of 19 mutation targets were identified using both DNA Star and DeepView as described in Section 8.1.2 from which 8 mutation targets were selected for further assessments, which were produced and purified as described in Section 8.1.3.

[0274] The expression profiles of all 8 B-GEn.2 variants were comparable to that of the wild type B-GEn.2 (Table 7). A representative Coomassie-stained SDS-PAGE image of the single-step heparin sulfate-based purification of one of the variants (B-GEn.2 D501R) is shown in FIG. 5.

TABLE 7			
B-GEn.2	Measured Concentration (mg/mL)	Estimated Purity (%)	Adjusted Concentration (mg/mL)
WT	9.0	65 %	5.8
Variant 1	6.0	61 %	3.7
Variant 2	8.9	74 %	6.6
Variant 3	6.4	74 %	4.7
Variant 4	10.0	59 %	5.9
Variant 5	9.0	76 %	6.8
Variant 6	17.0	72 %	12.2
Variant 7	12.3	59 %	7.3
Variant 8	9.0	65 %	5.9

8.3. Example 2: Characterization of RNPs with B-Gen.2 Variants

[0275] The purified B-Gen.2 variants were formed into RNPs with a guide RNA targeting a site within the B2M gene, as described in Section 8.1.4. Efficiency of RNP formation as well as in vitro cutting efficacy of the target B2M site were evaluated.

[0276] The results demonstrated that the B-Gen.2 variants were able to efficiently complex into RNPs with the target guide RNA, wherein the complexation efficiencies for the B-Gen.2 variants were found to be in the range of 71-92%.

[0277] The in vitro activity assay indicated that there were no significant differences detectable for B-Gen.2 variants compared to the WT B-Gen.2. Target plasmid cutting efficiencies for the B-Gen.2 variants, as estimated by EC_{50} , were found to be in the range of 0.07-0.17 nM. FIG. 6 shows the cutting efficiency for one of the B-Gen.2 variants, B-Gen.2 D501R, relative to cutting efficiencies of WT B-Gen.2 and Cpf1 for different nuclease:linearized plasmid ratios.

8.4. Example 3: High Efficacy *In Cellula* Gene Editing of B2M Locus in iPSCs with B-Gen.2 D501R

[0278] In cellula gene editing of B-Gen.2 D501R was evaluated by examining indel generation at the B2M target gene locus in iPSCs as described in Section 8.1.4., which was compared to the gene editing achieved by WT B-Gen.2 and AsCpf1 *Ultra* (IDT).

[0279] An indel formation efficiency of greater than 80% was achieved with B-Gen.2 D501R, which was greater than the efficiency achieved by Cpf1 (FIG. 7). This indel formation efficiency achieved by B-Gen.2 D501R was 2-3-fold higher than the indel formation achieved by the WT B-Gen.2 (FIG. 7). This result suggests that the substitution of Asp with Arg at D501, which in the crystal structure of AacC2c1 is shown to hydrogen-bond to the +1 phosphate backbone group of the target DNA strand, enhanced the ability of the B-Gen.2 variant to cleave its target substrate.

9. SEQUENCE LISTING

[0280] Exemplary sequences of the disclosure are provided in Table 8 below (where "SEQ" refers to the SEQ ID NO).

TABLE 8		
SEQ	Description	Sequence
1	WT B-GEEn.1 aa sequence	MTIRSMKLLKIYSGRSAPQLRQGLWRLHRLNNEGTAAYMDWLHMRQEA LPGKSKEEIRAELEERRVRQQQEKNGVQNDQVPMDEVLSALRQLYELLVPS AVNNSGDAQTLRSRKFLSPLVDPNSEGKGTSNAGAKPGWRKKQEAGDPSW EKDYERWLKRKQADPTAEILGKLETAGLKPLFPLYTNEVKDIRWMLPTSK QYVRNWD RDMFQQAI EHL LSWETWNRKVNEERAKLKETVRRFEEQHLANG KDWLSP LQAYEANREQALRDMAI SP SDRFRITRRQIKGWS ELYERWNKLA PTASVEAYMQEVRHVQKKLGGTFGDADLYRFLAKPENVHIWRDHQERLHY YAAYN DLHKRLMSAKEQA AFTLPDPVAHPLWVRFDARDGNLFTYILQADS SKQRSRRYVNF SRFLWPVEDGYFEETENVKVELALSQGFYRQVIVHDNPT GKQKITFDYSSKEILEGHLGGAKLQLDNRFLRKSGRDFETGDFGPAFLN VVLDLKPKEVKNGR LQSPLGQALLVKSRPNDIPKVYGYKPDALAAWLEQ ASGEESLGESLRQGFVRMSIDLGVRSAAAI SVFSVKGEKTRREGDKVCYP VGETGLFAVHDRSFLRLPGESSEKRVNVERDKRKTERMQIRYHIRTAR VLR LANKATPMDRIKAVQDVLNDIESTRFMNDHDDHHVYNHALET LRTYAP DHQGIWEEQVIAAHRQLEHHVGVI VGEWRKNWKGDRRGTVGLSMDNIEEL DEMRRLLI SWSRRARYPREAKPFQVNESNPVHLLRHLQNLKEDRLKQLAN LIVMTALGYVYDSKEKKWKAAYPACQLILFEDLQRYRFLHRSARENSQL MKWAHRSIPKYVWMQGEPLYGLQIGDVWAGFTSRYHAKTGAPGIRCKALTE KDFQQGRLLLES LVAEGMFTLQEVGTLKPGDIVPAEGGELFVTLADDSGDR IVITHADINAAQN VQKRFWLANSEFRVACRSVQIASQECFIPSSSEVAK KMGKGVFVRDFS FHKDMEVYHWNQVKLTAKNVPTDHSDDLQDLQDYQAI LEEARESSSSYKTLFRDPSGFFFFPDDVWVPQNIYWREVKKTTITALLRKRI MST
2	WT B-GEEn.1.2 aa sequence	MPIRSFKLKLVTNHGDS TYMDKLRRGLWKTHVI INRGIAAYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQRRNNWSEQTGTDELLSLSRVYELLV PSAIGEKDAQMLSRKFLSPLVDPNSEGGRGTAKSGRKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPLFPLFSDEQKEIRWLPPK KRQFVRTWDRDMFQQALERM LSWESWNRRAVEEYQKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEY LITRRQIRGWKEVYEKWSK TSENASQEQ LWRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQA AFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFS SLILPEAQSWVEIENVTVGIGNSLQLKQIRLDGYADKKQ KVKYDYDSSRFELTGVLGAKIQFDRKHLKKAHRLAEGETGPIFLNVVV DVEPFLEVKNRRLRTP LGQVLQVNTRDWPKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARIRSLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELLVSIQRKLELDQSEWISELEKLFYIDESAEEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISISWIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDRRLKQMANLFFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHR SIPRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNSYVVNQLIKDKFTQDNQTEILKAGQIVPWQGGELFVTFADRSGASLA VIHADINAAQN LQKRFWQHNS EFRVPCKVVGGLVPVYEKMRKLF GKGL FVNIDDPESKEVYRWEHSTKMKS KTT PVDLESEDI EHEELSDEWEDMQEG YKTL LRDPSGFFWSSDSWIPQKDFWIRVKSRIKSLREQIR
3	WT B-GEEn.2 aa sequence	MPIRSFKLKLVTNHGDS TYMDKLRRGLWKTHVI INRGIAAYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQRRNNWSEQTGTDELLSLSRVYELLV PSAIGEKDAQMLSRKFLSPLVDPNSEGGRGTAKSGRKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPLFPLFSDEQKEIRWLPPK KRQFVRTWDRDMFQQALERM LSWESWNRRAVEEYQKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEY LITRRQIRGWKEVYEKWSK

TABLE 8		
SEQ	Description	Sequence
		TSENASQEQWLWRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFSLLILPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVKYYDYSSRFELTGLGGAQIQFDRKHLKKAHRLAEGETGPIFLNVVV DVEPFLEVKNRRLRTPLGQVLQVNRDWPVKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLRLPGEEISSKKMIEKRKERARISLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELIVSIQKKLELDQSEWIELEKLFYDIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISWSIEELEET RKLIIAWSKHSREP GIPKRLEKEETFAPEHLQHIQNVKDDRLLQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHRSIPLRTYMQAELEFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNSYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVTFADRSGASLA VIHADINAAQNQLQKREFWQHNSSEVFRVPCKVVKGGLVVPEYKMRKLFQKGL FVNIDDPESKEVYRWEHSTKMKSKTTPVDLESEDIDHEELSDEWEDMQEG YKTLRLDPSGFFWSSDSWIPQKDFWIRVKSRIKSLREQIR
4	B-Gen.1 D504R aa sequence	MTIRSMKLLKIYSGRSAPQLRQGLWRLHRLNNEGTAAYMDWLVMRQEA LPGAESKEETRAELERRVRQQQEKNGVQNDQVPMDEVLSALRQLYELLVPS AVNNSGDAQTLRSKFLSPLVDPNSEGGKGTSNAGAKPGWRKKQEGADPSW EKDYERWLKRKQADPTAEILGKLETAGLKPLFPLYTNEVKDIRWMPITSK QYVRNWRDMFQQAIEHLLSWETWNRKVNEERAKLKETVRRFEEQHLANG KDWLSPLQAYEANREQALRDMATSPSDRFRITRRQIKGWSSELYERWNLKA PTASVEAYMQEVRHVQKKLGGTFGDADLYRFLAKPENVHIWRDHQERLHY YAAYNLDLHKRLMSAKEQAAFTLPDPVAHPLWVRFDARDGNLFTYILQADS SKQSRRYVNFSRFLWPVEDGYFEETENVKVELALSQFYRQVIVHDNPT GKQKITFDYSSKEILEGHLGGAKLQLDRNFLRKSGRDFETGDFGPAFLN VVLRLLKPKQEVKNRQLQSPLGQALLVKSRENIPKVYGYKPDALAAWLEQ ASGEESLGESLRQGRVMSIDLGVRSAAAI SVFSVKGEKTRREGDKVCYP VGETGLFAVHDSRFLRLPGESSEKRVNVERDKRKTTERMQIRYHIRTLLAR VLRLANKATPMDRIKAVQDVLNDIESTRFMNDHDDHVVYNHALETIRTYAP DHQGIWEEQVIAAHRQLEHHVGVI VGEWRKNWKGDRRGTVGLSDMNIEEL DEMRRLIISWSRRARYPREAKPFQVNESNPVHLLRHLQNLKEDRLKQLAN LIVMTALGYVYDSKEKKWKAAYPACQLILFEDLQRYRFHLDRSARENSQL MKWAHRSI PKYVWMQGEPIYGLQIGDVWAGFTSRYHAKTGAPGIRCKALTE KDFQQGRLLSLVAEGMFTLQEVGTLKPGDIVPAEGGELFVTLADDSGDR IVITHADINAAQNVQKRFWLANSERFRVACRSVQIASQECFIPSSSEVAK KMGKGVFVRDFS FHKDMEVYHWNQVKLTAKNVPTDHSDDLQDLQDYQAI LEEARESSSSYKTLFRDPSGFFFFPDDVWVPQNIYWREVKKTITALLRKRI MST
5	B-Gen.1.2 D501R aa sequence	MPIRSFKLLVTHNGDSTYMDKLRRGLWKTHVI INRGIAYYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQQRNNWSEQTGTDELLSLSRVYELLV PSAIGEKDAQMLSRKFLSPLVDPNSEGGRTAKSGRKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPLFPLFSDEQKEIRWLPPK KRQFVRTWRDMFQQALERMLSWESWNRRAVEEYQKLQAQRDEVYAKYLE DAGSWLNDLQTFEQREEELAEVSFEPNSEYLIITRRQIRGWKEVYEKWSK TSENASQEQWLWRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFSLLILPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVKYYDYSSRFELTGLGGAQIQFDRKHLKKAHRLAEGETGPIFLNVVV RVEPFLEVKNRRLRTPLGQVLQVNRDWPVKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT

TABLE 8		
SEQ	Description	Sequence
		DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARIRSLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELLVSIQKLELDQSEWISELEKLFYDIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISISWIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRMLMKW AHRSI PRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSN SYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVT FADRS GASLA VIHADINAAQN LQKRFWQHNS EVFRVPCKVVKGGLVPVYEKMRKLF GKGL FVNIDDPESKEVYRWEHSTKMKS KTT PVDLESEDI EHEELSDEWEDMQEG YKTL LRDP SGFFWSSDSWIPQKDFWIRVKS RIGKSLREQIR
6	B-Gen.2 D501R aa sequence	MPIRSFKLKLVTNHGDS TYMDKLRRLGWLKTHVI INRGIAYYMNTLALMRQ EPYGSKSREEVRLD LLSTLREQRRNNWSEQTGT DDELLSLSRRVYELLV PSAIGEKGDAQMLSRKFLSPLVDPNSEGGRGTA SGKRKPRWKKMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPLFPLFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERM LSWESWNNRRVAEEYKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQ LWRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVT FSSLI LPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVKYDYSSRFELTGVLGGAKIQFDRKHLKKAHRLAEGETGPIFLNVVV RVEPFLEVKNRRLRTP LGQVLQVNRDWPVKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARIRSLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELLVSIQKLELDQSEWISELEKLFYDIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISISWIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRMLMKW AHRSI PRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSN SYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVT FADRS GASLA VIHADINAAQN LQKRFWQHNS EVFRVPCKVVKGGLVPVYEKMRKLF GKGL FVNIDDPESKEVYRWEHSTKMKS KTT PVDLESEDI EHEELSDEWEDMQEG YKTL LRDP SGFFWSSDSWIPQKDFWIRVKS RIGKSLREQIR
7	B-Gen consensus I aa sequence	MPIRSFKLKLVTNHGDS TYMDKLRRLGWLKTHVI INRGIAYYMNTLALMRQ EPYGSKSREEVRLD LLSTLREQRRNNWSEQTGT DDELLSLSRRVYELLV PSAIGEKGDAQMLSRKFLSPLVDPNSEGGRGTA SGKRKPRWKKMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPLFPLFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERM LSWESWNNRRVAEEYXKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQ LWRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVT FSSLI LPEAQSWVEIENVTVGIGNSLQLKRQIRLDGYADKKQ KVKYDYSSRFELTGVLGGAKIQFDRKHLKKAHRLAEGETGPIFLNVVV DVEPFLEVKNRRLRTP LGQVLQVNRDWPVKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARIRSLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELLVSIQKLELDQSEWISELEKLFYDIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISISWIEELEET RKLLIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRMLMKW AHRSI PRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSN SYVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVT FADRS GASLA

TABLE 8

SEQ	Description	Sequence
		VIHADINAAQNLQKRFWQHNSSEVFRVPCKVVKGLVPVYEKMRKLFQKGL FVNIDDPESKEVYRWEHSTKMKSCTTPVDLESEDIXHEELSDEWEDMQEG YKTLRLRDPGFFWSSDSWIPQKDFWIRVKSRIKSLREQIRST
8	B-GEN.1 RuvC I domain aa sequence	DALAAWLEQASGEESLGSESLRQGFRVMSIDLGVRSAAAI SVFSVKGEKT REGDKVCYPVGETGLFAVHDRSFLRLPGESE
9	B-GEN.1 RuvC II domain aa sequence	CQLILFEDLQRYRFHLDRSARENSQLMKWAHRSIPKYVWMQGEPEYGLQIG DV
10	B-GEN.1 RuvC III domain aa sequence	ADINAAQNLQKRFWL
11	B-GEN.1.2 and B- GEN.2 RuvC I domain aa sequence	KELSVLMENTQIGNENGSTIEAGMRMSIDLQRTAAAVSIFEVISKKP DEKETKLFYPIADTDLYAVHRRSLLRLPGEEISS
12	B-GEN.1.2 and B- GEN.2 RuvC II domain aa sequence	CQVILFEDLSRYRFALDRPRENNRLMKWAHRSIPRLTYMQAELFGIQVG DV
13	B-GEN.1.2 and B- GEN.2 RuvC III domain aa sequence	ADINAAQNLQKRFWQ
14	Wild-type SV40 large T protein	PKKKRKV
15	Nucleoplasmin	KRPAATKKAGQAKKKK
16	c-myc	PAAKRVKLD
17	EGL-13	MSRRRKANPTKLS ENAKKLAK EVEN
18	TUS	KLKIKRPVK
19	Basic PY-NLS	(KR) -X (0, 2) - (KR) - (KR) -x (3, 10) - (RHK) -X (1, 5) -PY
20	SV40 large T protein, variant 1	PKKKRMV
21	SV40 large T protein, variant 2	PKKKRKWEDP
22	SV40 large T protein, variant 3	CGYGPKKKRKVGG
23	SV40 large T protein, variant 4	CGYGPKKKRKV
24	SV40 large T protein long NLS	CYDDEATADSQHSTPPKKKKRWEDPK DFESELLS
25	SV40 large T protein, variant 5	CGGPKKKRKWG
26	SV40 large T protein, variant 6	PKKKIKW

TABLE 8		
SEQ	Description	Sequence
27	SV40 large T protein, variant 7	KRTADGSEFESPKKKRKV
28	Nucleoplasmin min-NLS	TKKAGQAKKK
29	Nucleoplasmin NLS variant 1	TKKAGQAKKKKLD
30	Nucleoplasmin NLS variant 2	CGQAKKKKLD
31	c-myc NLS2	RQRRNELKRSP
32	Polyoma large T protein	PKKARED
33	Polyoma virus large T protein	CGYGWSRKRPRPG
34	SV40 VP1 capsid polypeptide	APTKRKGS
35	Polyoma virus major capsid protein VP1 (11 N-terminal aa)	APKRKSGVSKC
36	SV40 VP2 capsid protein (39 kD)	PNKKKRK
37	Polyoma virus capsid protein VP2	EEDGPQKKKRRL
38	Yeast histone H2B	GKKRSKA
39	Adenovirus E1a	KRPRP
40	Adenovirus type 2/5 E1a	CGGLSSKRPRP
41	Xenopus N1, NLS1	LVRKKRKTE ₃ SP
42	Xenopus N1, NLS2	LKDKDAKKSQEQ
43	V-Rel	GNKAKRQRST
44	NS1 protein of influenza A virus	PFLDRLRRDQK
45	Human lamin A	SVTKKRKLE
46	Xenopus lamin A	SASKRRRLE
47	Human c-myc	ACIDKRVKLD
48	Murine c-abl	SALIKKKKKMAP
49	Adenovirus 5 DBP	PPKKRMRRRIE
50	Rat glucocorticoid receptor	YRKCLQAGMNLEARKTKKKIKGIQQATA
51	Human glucocorticoid receptor	CGYGARKTKKKIK
52	Human glucocorticoid	RKCLQAGMNLEARKTKK

TABLE 8		
SEQ	Description	Sequence
	receptor NLS variant	
53	Rabbit progesterone receptor	RKFKKFNK
54	Human estrogen receptor	CGYGIRKDRRGGR
55	Human androgen receptor	CGYGARKLKKLGN
56	Chicken Ets1 core NLS	GKRKNKPK
57	c-myb	PLLKKIKQ
58	N-myc	PPQKKIKS
59	p53	PQPKKKP
60	p53 NLS variant	PQPKKKPL
61	c-erb-A	SKRWAKRKL
62	Yeast ribosomal protein L29 NLS	MTGSKTRKHRGSGA
63	Yeast ribosomal protein L29 NLS, variant 1	MTGSKHRKHPPGSGA
64	Yeast ribosomal protein L29 NLS, variant 2	RHRKHP
65	Yeast ribosomal protein L29 NLS, variant 3	KRRKHP
66	Yeast ribosomal protein L29 NLS, variant 4	KYRKHP
67	Yeast ribosomal protein L29 NLS, variant 5	KHRRHP
68	Yeast ribosomal protein L29 NLS, variant 6	KHKKHP
69	Yeast ribosomal protein L29 NLS, variant 7	RHLKHP
70	Yeast ribosomal protein L29 NLS, variant 8	KHRKYP
71	Yeast ribosomal protein L29 NLS, variant 9	KHRQHP

TABLE 8		
SEQ	Description	Sequence
72	Xenopus N1 NLS1	LVRKKRKTE ₃ SP
73	Xenopus N1 NLS2	LKDKDAKKSQ ₂ E
74	Viral Jun	ASKSRKRKL
75	Human T-cell leukemia virus Tax trans-activator protein	GGLCSARLHRHALLAT
76	Mouse nuclear MX1 protein (72 kD)	DTREKKKFLKRLLRLDE
77	Mouse nuclear MX1 protein NLS variant	REKKKFLKRR
78	Human retinoic acid receptor	CGYGDRNKKKKE
79	Human XPAC	RKRQRALMLRQAR
80	T-DNA -linked VirD2 endonuclease	EYLSRKGKLEL
206	Putative core NLS of yeast TRM1	KKS ₃ KKRC
81	Gag protein of human foamy retrovirus	QPQRYGGGRGRRW
82	SV40 Vp3 structural protein	NKKKRKLSRGSSQKT ₂ KGTSASAKARH KRRNRSSRS
83	Simian sarcoma virus v-sis gene product	RVTIRTWRWRPPKGKHRK
84	Putative bipartite NLS of Xenopus protein factor Xnf7	KRKIEEPEPEPKKAK
85	Yeast SWI5 gene product	KKYENVVIKRS ₂ PRKGRPRKD
86	Herpes simplex virus ICP8 protein	GRKRAFHGDDPFGE ₂ GPPDKKGD
87	Bipartite NLS of VirD2 endonuclease	KRPREDDDGEPSE ₂ RKRARDDR
88	IBB domain from importin-alpha	RMRIZFKNKGKDTAELRRRRVEVSVEL RKAKKDEQILKRRNV
89	hRNPAI M9 NLS	NQSSNFGPMKGGNFGGRSSG ₂ PYGGG GQYFAKPRNQGGY
90	Human poly(ADP-ribose) polymerase	KRKGDEV ₂ DGVDEVAKKKSKK

TABLE 8		
SEQ	Description	Sequence
91	Hepatitis virus delta antigen	RKLKKKIKKL
92	influenza virus NLS	PKQKKRK
93	mouse c-abl IV	SALIKKKKKMAP
94	myoma T protein NLS1	VSRKRPRP
95	myoma T protein NLS2	PPKKARED
96	myc proto-oncogene protein [Homo sapiens]	GPAAKRVKLD
97	heat shock factor protein HSF8 [Solanum lycopersicum]	KKRRIKQD
98	2x SV40, LrgT	PKKKRKVEDPKKKRKVD
99	3x SV40, LrgT	PKKKRKVDPKKRKVDPKKRKV
100	monopartite NLS consensus cluster 1 (Dissertation Tatyana Goldberg 2016)	KKGKKKGK
101	monopartite NLS consensus cluster 2 (Dissertation Tatyana Goldberg 2016)	PKRRRGVVL
102	monopartite NLS consensus cluster 3 (Dissertation Tatyana Goldberg 2016)	EQLFKRRNV
103	monopartite NLS consensus cluster 4 (Dissertation Tatyana Goldberg 2016)	KRRRR
104	monopartite NLS consensus cluster 5 (Dissertation Tatyana Goldberg 2016)	KKRRR
105	monopartite NLS consensus cluster 6 (Dissertation	EGAPPAKRPR

TABLE 8		
SEQ	Description	Sequence
	Tatyana Goldberg 2016)	
106	monopartite NLS consensus cluster 7 (Dissertation Tatyana Goldberg 2016)	MLRRRRRKRAR
107	monopartite NLS consensus cluster 8 (Dissertation Tatyana Goldberg 2016)	RRKRR
108	monopartite NLS consensus cluster 9 (Dissertation Tatyana Goldberg 2016)	RKRK
109	monopartite NLS minor cluster (Dissertation Tatyana Goldberg 2016)	FKAVLEDILGEL
110	NLSdb, protein origin P10152	KNRRL
111	NLSdb, protein origin Q09353	RKRHW
112	NLSdb, protein origin Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1	RRKKRR
113	NLSdb, protein origin Q99PU7, D3ZHS6, Q92560, A2VDM8	RRKRSR
114	NLSdb, protein origin Q14781, P30658	KRGRKP
115	NLSdb, protein origin P02545, P48678, P48679, Q3ZD69	KKRKLE
116	NLSdb, protein origin O35914, Q01954	PKKKSRK

TABLE 8		
SEQ	Description	Sequence
117	NLSdb, protein origin Q9FYS5, Q43386	PKRGRGR
118	NLSdb, protein origin E5RQA1	KEKRKKR
119	NLSdb, protein origin Q9Z1J1, Q9HCS4, Q924A0	KKKKRKR
120	NLSdb, protein origin Q80WE1, Q5R9B4, Q06787, P35922	RRGDGRRR
121	NLSdb, protein origin Q9Y261, P32182, P35583	LSPSLSPL
122	NLSdb, protein origin P07156	VNFSEFSK
123	NLSdb, protein origin Q96EB6	IVINILSE
124	NLSdb, protein origin Q6AZ28, O75928, Q8C5D8	PPAKRKCIF
125	NLSdb, SeqNLS prediction	QRPGPYDRP
126	NLSdb, protein origin Q8L7L5, A1L4X7, O80834, Q8LPN5	KRKRGRPRK
127	NLSdb, protein origin O88907, O75925	KIKELYRRR
128	NLSdb, SeqNLS prediction	MVQLRPRASR
129	NLSdb, protein origin Q5VK71	KKRREKQRRR
130	NLSdb, protein origin P0C6L6, P25880, Q81835, P29833, P25882, P06934, P0C6L3, P29997, P0C6M5, P0C6M9, P29996, P0C6M1, P0C6L8, P0C6M2, P0C6L7, P25881	EGAPPAKRAR

TABLE 8		
SEQ	Description	Sequence
131	NLSdb, protein origin Q45FA5	PKKGDKYDKTD
132	NLSdb, protein origin P97376, Q14331	KKKKSKDKKRK
133	NLSdb, protein origin P21827	AHRAKKMSKTHA
134	NLSdb, protein origin Q6ZN17	KKGPSVQKRKKT
135	NLSdb, protein origin Q4R8Y1	KGVKRKADTTTP
136	NLSdb, protein origin Q91Y44, D4A7T3	KGVKRRADTTTP
137	NLSdb, protein origin Q15397, Q8BKS9, Q562C7	KKPKWDDFKKKKK
138	NLSdb, protein origin Q96GM8	KRRRRRRREKRKR
139	NLSdb, protein origin P61635, Q6DV79, Q19S50, P52631	DVRKRVQDLEQKM
140	NLSdb, protein origin O60716	KKGKDEWFSGKK
141	NLSdb, protein origin P30999	KKGKDEWFSGKKP
142	NLSdb, SeqNLS prediction	ASPEYVNLPIGNG
143	NLSdb, protein origin Q7Z7C8, Q5ZMS1, Q9EQH4, A7MAZ4	YLRPVKKPKIRRK
144	NLSdb, protein origin O15381	KRKGLKNKGSKRKK
145	NLSdb, SeqNLS prediction	RRRGKNKVAAQNCRK
146	NLSdb, protein origin O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4	DKAKRVSRNKSEKRR
147	NLSdb, protein origin G5EFF5	EEQLRRRKNSRLNNTG

TABLE 8		
SEQ	Description	Sequence
148	NLSdb, protein origin P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0	HKKKHPDASVNFSEFSK
149	NLSdb, protein origin Q9Z301, O54943, Q8K3T2	KKTGKNRKLKSKRVKTR
150	NLSdb, protein origin Q38740, Q38741, Q700W2, Q9S7A9, Q6Z461, P93015, Q94JW8, Q9S758	KRSCRRRLAGHNERRRK
151	NLSdb, protein origin Q501B2	KRQRRKQSNRESARRSR
152	NLSdb, SeqNLS prediction	RGKGGKGLGKGGAKRHRK
153	NLSdb, protein origin Q63014, Q9DBR0	RRRGFERFGPDNMGRKRK
154	NLSdb, protein origin Q0IJ08, Q2TAE3, Q63470, Q13627, Q61214	RRHQGGQGGDSSHKKERK
155	NLSdb, protein origin Q8LAM0, O24454	KKKTGVIAPKRFVQRLKK
156	NLSdb, protein origin Q0E671	KRAMKDDSHGNSTSPKRRK
157	NLSdb, protein origin Q9P127	KVNFLDMSLDDIIYKELE
158	NLSdb, SeqNLS prediction	KKYENVVIKRSRKRGRPRK
159	NLSdb, SeqNLS prediction	KRGNSSIGPNDLKSRKQRKK
160	NLSdb, protein origin Q9BZZ5, Q5R644	KRASEDTTSGSPPKSSAGPKR
161	NLSdb, SeqNLS prediction	KRIHSVSLSQSQIDPSKKVKRAK

TABLE 8

SEQ	Description	Sequence
162	NLSdb, SeqNLS prediction	EVLKVI RTGKRKKKAWKRMVTKVC
163	NLSdb, SeqNLS prediction	I INGRKLKLKKSRRRSSQTSNNSFTSRRS
164	NLSdb, protein origin Q76IQ7	AHFKISGEKRPSTDPGKKAKNPKKKKKKDP
165	Flexible linker	(GGGGS) ₃
166	Mixed linker	GGG (EAAAK) ₃
167	Rigid linker	A (EAAAK) ₄ AEA (EAAAK) ₄ A
168	Flexible linker 2	GGGGS L VPRGSGGGGS
169	Proline rich, rigid, interdomain linker	GAAPAAAPAKQEAAAPAPAAKAEAPAAAPAAKA
170	Cleavable linker	TRHRQPRGWE
171	B-GEN.2-sgRNA_v4	guuuagcuauaggcuauaagauaguugugucaagugcuucggagaccuaacacgcucauccagucacacggcuaucauaaaauagauagguuagcac
172	B-GEN.2-sgRNA_v4.2	uuagcuauaggcuauaagauaguugugucaagugcuucggagaccuaacacgcucauccagucacacggcuaaaaauagguuagcac
173	B-GEN.2-sgRNA_v4.3	uuagcuauaggcuauaagauaguugugucaagugcuucggagaccuaacacgcucauccagucacacggcuaaaaauagccagcac
174	B-GEN.2-sgRNA_v4.4	uuagcuauaggcuauaagauaguugugucaagugcuucggagaccuaacacgcucauccagucacacggcuaaaaauagccagcac
175	B-GEN.2-sgRNA_v4.5	uuagcuauaggcuauaagauaguugugucaagugcuucggagaccuaacacgcucauccagucacacggcuaaaaauagccagcac
176	Tracr B-GEN.2	UAAUAAUAAU AUGAGUUACAGGCUCAUGUCAUUCGUGCGUGUGCUGCGAGGCUCUUAACAGUCUCAACACGGAUCACGUUAUUUCCA CUAAGUGAUUGGAC
177	Tracr B-GEN.1	CUGGAUUAUAAU AAGAUGUUUAGCUAUAGGCUAUAUAGAUAUUGUGUCAAGUGCUUCGGAGACCUAACACGUCAUCCAGUCACAACGGCUAUCUAUAUUUCCACUAAC
178	Guide RNA targeting B2M	*mU*mA*GCUAUAGGCUAUAUAGAUAUUGUGUCAAGUGCUUCGGA GACCUAACACGUCUCCAGUCACAACGGCUAAAAUAGCCAGCACAG UGUAGUACAAGAGAUAGA*mA*mA*mG
179	BkaCas12b aa sequence	MAVKSIKVKLRLSECPDILAGMWQLHRATNAGVRYTETWVSLMRQEILYS RGPDGGQQCYMTAEDCQRELLRRLNRQLHNGRQDQPGTDADLLAISRL YEILVLQSIGKRGDAQQIASSFLSPLVDPNSKGGRGEAKSGRKPWQKMR DQGDPRWVAAREKYEQRKAVDPSKEILNSLDALGLRPLFAVFTETYSRV DWKPLGKSQGVRTWDRDMFQQALERLMSWESWNRRVGEEYARLFQQKMKF EQEHFAEQSHLVKLARALEADMRAASQGFEAKRGTAHQITRRALRGADRV FEIWKSIPEEALFSQYDEVIRQVQAEKRRDFGSHDLFAKLAEPKYQPLWR ADETFLTRYALYNGVLRDLEKARQFATFTLPDACVNPITWTRFESSQGSNL HKYEFLFDHLGPRHAVRFQRLLVVESEGAKERDSVVPVAPSGQLDKLV LREEKSSVALHLHDTARPDGFMAEWAGAKLQYERSTLARKARRDKQGMRSWRRQPSMLMSAAQMLEDAKQAGDVYLNISVRVKSPSEVRGQRRPFPYAL FRIDDKQRRVTVNYNKLSAYLEEHPDKQIPGAPGLLSGLRVMSVDLGLRT SASISVFRVAKKEEVEALGDGRPPHYPIHGTDDLVAVHERSHLIQMPGE TETKQLRKLRERQAVLRPLFAQLALLRLLVRCGAADERIRTRSWQRLTK

TABLE 8

SEQ	Description	Sequence
		QGREFTKRLTPSWREALELELTRLEAYCGRVPDDEWSRIVDRTVIALWRR MGKQVRDWRKQVKSGAKVKVKG YQLDVVGNSLAQIDYLEQQYKFLRRWS FFARASGLVVRADRESHFAVALRQHIENAKRDLKKLADRI LMEALGYVY EASGPREGQWTAQHPPCQLII LEELSAYRFSDDRPPSENSKLMWGHGRI LEELVNQAQVHDLVGTVYAAFSSRFDARTGAPGVRCRRVPARFVGATVD DSLPLWLTEFLDKHRLDKNLLRPDDVIPTGEGEFLVSPCGEEAARVRQVH ADINAAQNLRRLWQNF DITELRLRCDVKMGEGT VLVPRVNNARAKQLF GKVLVSQDGVTFERSQTGGKPHSEKQTDLT DKELELIAEAD EARA KSV VLFRDPSGHI GKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSLD
180	BcyCas12b aa sequence	MAVKSIKVKLRRLNDMPEIQAGLWKLHTEVNAGVRYYTEWLSLLRQENLYR RSPSGDGAQEC DKTAEQCKAELLERLRARQLENGHRGPAGSDDELLQLAR QLYELLVPQARGAKGDAQQISRKFLSPLTDRDAVGGLGISKAGNKPRWVR MRDAGEPGWEEEEKAKAEARKSADVTADVLHALADFG LKPLMRVYTESEMS SVQWKPLRQGQAVRTWDRDMFQQAIERMMSWESWNRRVGEEYAKLTEQKN RFEQKNFVGQEHVLVHELQ LTMKEASPGLESKEQTAHYVTGRALRGAD KVFERWEKLAPDAPLDLYETEIKNVQRRNTRRFGSYDLFAKLAEPKYQAL WREDASFLTRYAAYNGILRKLNHAKMFATFTLPDPTVHPIWARFDKLGGN LHQYTFLEFNEFGEGRHAIRFQKLLKVENGVAKVDDVTVPISRSTQLDNL LPRGSTEPIALYFRDYGAEQPFEGEFGGAKIQYRRDQLEQLRRRKAGADV YLNF SVRVQSRSEARGERRPPYAAVFRLIGDNNRAVFHFDKLSDYLA EHP DEGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVAPRDELAPNSKGRVPF YFPIKGNLNLVAIHRSQLLKLPGETESKELRAIREERQRTLRLQRTQLM YLRLLVRCGSEDVRRRERSWAKLTEHPMDGSESITSDWREAFVNELOKLM PLHGVCSDKEWMEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYKK DVVGNSIEQIEYLERQYKFLKSWSFFGKVSQVIRADKGSRFATTLREH IDHAKEDRLKKLADRIIMEALGYVYALDDERGGKGVVAKYPPCQLILLEE LSEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLVGTMYAAFSS RFDARTGAPGIRCRRVPARCVKEHNPE SFPWWLNRVFAEQQLDPSLLRAD DLIPTGEGEVFVSPFSTKEGDFHQIHADLNAAQNLRRLWSDFDISQIRL RCDRGEVDGEPVLI PRLTGKRTAESYRNKVFSNTGVTYYERERGGKRG GLPLEELSEEEAE LLVEADEAREKSVVLMRDPSGIINRGDWTQREFWTL VNQRIEGLVKQIRSRGSACENTGEI
181	BacCas12b aa sequence	MAVKSMKVKLRRLNDMPEIRAGLWKLHTEVNAGVRYYTEWLSLLRQENLYR RSPNGDGEQECYKTAEECKAELLERLRARQVENGHCGPAGSDDELLQLAR QLYELLVPQAI GAKGDAQQIARKFLSPLADKDAVGGLGIAGAGNKPRWVR MREAGEPGWEEEEKAKAEARKSTDRTADVLRALADFG LKPLMRVYTDSDMS SVQWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQVRGEAYAKLVEQKS RFEQKNFVGQEHVLVQLVQLQDDMKEASHGLESKEQTAHYLTGRALRGSD KVFEKWEKLPDAPFDLYDTEIKNVQRRNTRRFGSHDLFAKLAEPKYQAL WREDASFLTRYAVYNSIVRKLNHAKMFATFTLPDATAHP IWTFRDKLGGN LHQYTFLEFNEFGEGRHAIRFQKLLTVEDGVAKVDDVTVPISMSAQLDDL LPRDPHELVALYFQDYGAEQHLAGEFEGGAKIQYRRDQNLHLHARRGARDV YLNL SVRVQSQSEARGERRPPYAAVFRLVGDNHRAVFHFDKLSDYLA EHP DDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNSEGRVPF CFP IEGNENLVAVHERSQLLKLPGETESKDLRAIREERQRTLRLQRTQLA YLRLLVRCGSEDVGRRERSWAKLIEQPM DANQMT PDWREAFEDLQKLKS LYGICGDREWTEAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYQKD VVGNSIEQIEYLERQYKFLKSWSFFGKVSQVIRAEKGSRFATTLREHI DHAKEDRLKKLADRIIMEALGYVYALDDERGGKGVVAKYPPCQLILLEE SEYQFNNDRPPSENNQLMQWSHRGVFQELLNQAQVHDLVGTMYAAFSS RFDARTGAPGIRCRRVPARCAREQNPEPFPWWLNKFVAEHKLDGCPLRAD

TABLE 8		
SEQ	Description	Sequence
		LIPTGEGEFFVSPFSAEEGDFHQIHADLNAAQNLRRLWSDFDISQIRLR CDWGEVDGEPVLIPTTGKRTADSYGNKVFTYKTGVTYERERGGKRRKV FAQEELSEEEAELLVEADEAREKSVVLMRDPSTGIINRGDWTRQKEFWSMV NQRIEGLVKQIRSRVRLQESACENTGDI
182	BmaCas12b aa sequence	MKSIKVKMLGLHLPETREGLWHLHEAVNLGVRYTTEWLALLRQGNLYRRG KDGAQECYMTAEQCRQELLVRLRDRQKRNHGTGDPGTDEELLGVARRLYE LLVPQSVGKKGQAQMLASGFLSLPADPKSEGGKGTSSKSRKPAWGMKEA GDSRWVEAKARYEANKAKDPTKQVIASLEMYGLRPLFDVFTETTYKTIRWM PLGKHQGVRAWDRDMFQQSLERLMSWESWNERVGAEFARLVDRDRDFREK HFTGQEHVLAQRLQEMKEASPGFESKSSQAHRTKRALRGADGIIDD WLKLSEGEVPDRFDELRLKRQAQNPFRFGSHDLFLKLAEPVFPQPLWREDP SFLSRWASVNEVLNKLEDAKQFATFTLPSPCSNPVWARFENAEGTNIFKY DPLFDHFGKGRHGVRFQRMIVMRDGVPTVEGIVVPIAPSRQLDALAPND AASPIDVFGDPAAPGAFRGQFGGAKIQYRRSALVRKGRREEKAYLCGFR LPSQRRGTGPADDAGEVFLNLSLRVESQSEQAGRRNPPYAAVFHISDQTR RVIVRYGEIERYLAEHPDTGIPGSRGLTSGLRVMSVDLGLRTSAAISVFR VAHRDELTPDAHGRQPFFFPIHGMHDLVALHERSHLIRLPGETESKKVRS IREQRLDRLNRLRSQMASLRLLVRTGVLDEQKDRDNWERLQSSMERGGER MPSDWWDLFQAQVRYLAQHRDASGEAWGRMVQAAVRTLWRQLAKQVRDWR KEVRRNADKVKIRGIARDVPGGHSQAQLDYLERQYRFLRSWSAFSVQAGQ VVRAERDSRFAVALREHIDNGKKDRLKKLADRI LMEALGYVYVTDGRRAG QWQAVYPPCQLVLLLEELSEYRFSNDRPPSENSQLMWVSHRGVLEELIHQA QVHDLVLGTIPAAFSSRFDARTGAPGIRCRRVPSIPLKDAPSIPIWLSHY LKQTERDAAALRPGELIPTGDGEFLVTPAGRGASGVRVVHADINAHNLQ RRLWENFDLSDIRVRCDRREGKDGTVVLI PRLTNQVRKERYSGVIFTS GVSTFTVGDAKTRRRSSASQGEEDLSDEEQELLAADARERSVVLFRDP SGFVNGGRWTAQRAFVGMVHNRIETLLAERFSVSGAAEKVVRG
183	BshCas12b aa sequence	MDALGLRPLFVVFTETETKSGVDWKPLGKSQGVRTWDRDMFQQALERLMSW ESWNRVRGEEYARLFQQKMKFEQEHFAEQSHLGELARALEADMRAASQGF EAKRGTAHQITRRALRGADRVFETWRRLPEDALFSQYDEVIRQVQAEKRR DFGSHDLFAKLAEPKYQPLWRADETFLTRYALYNGVLRDLEEAKQFATFT LPDACVNPWTRFESSQGSNLHKEYEFLFDHLGPGRHAVRFQRLLVVESEG AKEKDSVVVPVAPSGQLDKLVPREEEKSSVALHLHDTAQPDGFMMAEWAGA KLQYERSTLARKARDKQGMRLWRRQPSLMSAVQMLEDAKQAGDVYLN SVRVKSPSEVRGQRRPPYALFRIDDKQRRVTNKNLSAYLEEHPDKQI PGALGLLFGLRVMSVDLGLRTSASISVFRVAKREEVEALGDGHPHYPPI HGTDDLVAVHERSHLIQMPGETETKQLRKLRERQALRPLFAQLAVLRL LVRCGAADERIRTRSWQRLTKQGREFAKRLTPSWREALESELTRLEAYYG RVPDDEWSRLVDKTVTALWCRMKGQVRDWRKQVKSGAKVKVKRYQLDVVG GNSLAQIDYLEQQYRFLRRWSFFARASGLVVRADRDHFAVALRQHIENA KRDRLKKLADRI LMEALGYVYEASGPREGQWTAQHPPCQLIILEELSAYR FSDDRPPSENSKLMWGHGRIILEELVNQAQVHDLVLGTVYAAFSSRFDA TGAPGVRCRRVPARFVGATVDDSLPLWLTEFLDKHRLDKNLLRPDDVIPT GEGEFLVSPCGEEAARVRQVHADINAQNLRRLWQNFIDITELRLRCDVK MGGEKTVLVRVNNARAKQLFGKKVLVSQDGVTFERSQTGGKPHSEKQT DLTDKELELIAEADARAKSVVLFDRDPSTGHIGKGHWIRQREFWSLVKQRI ESHAAATIRVRGVGSSLD
184	BagCas12b aa sequence	MAIRSIKCLKLTKHTGPEAQNLRKGIWRTHRLNNEGVAAYMKMLLLFRQES TGERPKQELQEELICHIREQQQRNQADKNTQALPLDKALEALRQLYELLV PSSVGQSGDAQIISRKFLSPLVDENSEGGKGTSSKAGAKPTWQKKKEANDP TWEQDYEKWKRRREEDPTASVITTLLEYGIRPIFPLYTNTVTDIAWLPLQ

TABLE 8		
SEQ	Description	Sequence
		SNQFVRTWDRDMLQQAIERLLSWESWNKRVQEEYAKLKEKMAQLNEQLEG GQEWISLLEQYEENRERELRENMTAANDKYRITKRQMKGNELYELWSTF PASASHEQYKEALKRVQQRLRGRFGDAHFFQYLMEEKNRLIWKNPQRIH YFVARNELTKRLEEAKQSATMTLPNARKHPLWVRFDARGGNLQDYILTAE ADKPRSRRFVTFSQLIWPSESGWMEKKDVEVELALSQRQFYQQVKLLKNDK GKQKIEFKDKGSGSTFNGHLGGAKLQLERGDLEKEEKNFEDGEIGSVYLN VVIDFEPLQEVKNGRVQAPYGGVQLIIRRPNEFPKVTYKSEQLVWEIKA SPQHSAGVESLASGFRVMSIDLGLRAAAATSI FSVEESSDKNAADF SYWI EGTPLVAVHQRSYMLRLPGEQVEKQVMEKRDERFQLHQRVKFQIRVLAQI MRMANQYGDWRDELDSLQAVEQKKSPLDQTDRTFWEGIVCDLTQVLP NEADWEQAVVQIHRKAEYVVGKAVQAWRKRFADDERKGIAGLSMWNIEEL EGLRKLISWSRRTNRPQEVNRFERGHSTHQRLLLTHIQNVKEDRLKQLSH AIVMTALGYVYDERKQEWCAEYPACQVILFENLSQYRSNLDSTKENSTL MKWAHRSIPKYVHMQAEPYGIQIGDVRAEYSSRFYAKTGTGPIRCKKVRG QDLQGRRFENLQKRLVNEQFLTEEQVKQLRPGDIVPDDSGELFMTLTDGS GSKEVVFLQADINAAHNLQKRWFQRYNELFKVSCRIVRDEEEYVLPKTK SVQAKLGKGLFVKKSDTAWKDVYVWDSQAKLKGKTTFTTESESPEQLEDF QEIEEEAEAKGTYRTLFDRPSGVFFPESVWYPQKDFWGEVKKRKYGLR ERFLTAKAR
185	AacC2c1 aa sequence	MAVKSIKVKLRLLDDMPEIRAGLWKLHKEVNAGVRYTTEWLSLLRQENLYR RSPNGDGEQECDKTAECKAELLERLRARQVENGHRGPAGSDDELLQLAR QLYELLVPQAI GAKGDAQQIARFSLPLADKDAVGGGLIAKAGNKPRWVR MREAGEPGWEEKEKAETRKSADRTADVLRALADFGKLPLMRVYTDSEMS SVEWKPLRKGQAVRTWDRDMFQQAIERMMSWESWNQVRVQGEYAKLVEQKN RFEQKNFVGQEHVLVNLQQLQDMKEASPGLESKEQTAHYVTGRALRGSD KVFEKWGLAPDAPFDLYDAEIKNVQRRNTRRFGSHDLFAKLAEPEYQAL WREDASFLTRYAVYNSILRKLNHAKMFATFTLPDATAHPIWTRFDKLGGN LHQYTFLFNEFGERRHAI RFHKLKLVENGVAREVDDVTVPISMSEQLDNL LPRDPNEPIALYFRDYGAEQHFTGEFGGAKIQCRDQLAHMHRRRGARDV YLNVSVRVQSQSEARGERRPPYAAVFRVLVDGNHRAFFVHFDKLSDYLAEHP DDGKLGSEGLLSGLRVMSVDLGLRTSASISVFRVARKDELKPNKGRVPF FFPIKGNLNLVAVHERSOLLKLPGETESKDLRAIREERQRTLRLQTLA YLRLLVRCGSEDVGRRRERSWAKLIEQPVDAANHMTPDWREAFENELQKLK SLHGICSDKEWMDAVYESVRRVWRHMGKQVRDWRKDVRSGERPKIRGYAK DVVGGNSIEQIEYLERQYKFLKSWFFGKVSQVIRAEKGSRFATLREH IDHAKEDRLKKLADRI MEALGYVYALDERGKGKWKVAKYPPCQLILLEEL SEYQFNDRPPSENNQLMQWSHRGVFQELINQAQVHDLVGTMYAAFSSR FDARTGAPGIRCRVPARCTQEHNPFPFPWWLNKFVVEHTLDACPLRADD LIPTGEGEIFVSPFSAEEGDFHQIHADLNAAQNLQQLRWSDFDISQIRLR CDWGEVDGELVLIPLRTGKRTADSYSNKVFTNTGTYYERERGGKRRKV FAQEKLSSEEAELLVEADEAREKSVVLMRDPGSI INRGNWTRQKEFWSMV NQRIEGLVKQIRSRVPLQDSACENTGDI
186	BsyCas12b aa sequence	MPVRSFKVLVTRSGDAEHMLQLRRGLWKTHEIVNQGIAYYMNKLALMRQ EPYAGKSREVVRLLELLHSLRAQQKRNNTGDAAGTDEILNLSRRLYELLV PSAIGEKDAQMLSRKFLSPLVDENSEGGKGTAKSGRKPRWMMKREEGHP DWEAEREKDRAKKAADPTASILNDLEAFGLRPLFLFTDEQKGIQWLPKQ KRQFVRTFDRDMFQQALERMLSWESWNRRVAEEYQKLQAQRDELYAKYLA DGGAWLEALQSFEKQREVELAEESFAAKSEYLITRRQIRGWKQYKESQ LPEHAAQEQFWQVADVQTSPLGAFGDPKVYQFLSQPEHHHIWRGYPNRL FHYSYDNGVRKKLQRAHHDATFTLPDPVEHPLWIRFDARGGNIHDEYISQ NGKQYQVTFSRLLWPENETWVERENVTVAGASQQLKRQIRLDGYADKKQ

TABLE 8		
SEQ	Description	Sequence
		KVRYRDYSSGIELTGVLGGAKIQFDRRHRLKASNRLADGETGPVYLVNVV DIEPFLAMRNGRLQTPIGQVLQVNTKDWPKVTGYKPAELISWIQNSPLAV GTGVNTIEAGMRVMSVDLGQRSAAAVSIFEVMRQKPAEQETKLFYPIAVT GLYAVHRRSLLRLPGEKISDEIEQQRKIRAHARSLVRYQIRLLADVRL HTRGTAEQRRAKLDELLATLQTKQELDQKLWQTELEKLFDYIHEPAERWQ QALVAAHRTLEPVIQAVRHWRKSLRIDRKGLAGMSMWNI EEELETRKLL IAWSKHSRVPGEPNRLDKEETFAPQQLQHIQNVKDDRLKQMANLLVMTAL GYKYDEAEKQWKEAYPACQMILFEDLSRYRFALDRPRENNRLMKWAHRS I PRLVYLQGELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALKEEDLQPN YVVKQLIKDGFIREDTGSLKPGQIVPWSGGELFVTLADRSGLSLAVIHA DINAAQNLQKRFWQQNTEIFRVPCVTTSGLIPAYDKMKKLFKGKYFAKI NQTDTSSEVYVWEHSAMKMGKTTTPADPAEEGVFDES LTDEMEELDSQEGY KTLFRDP SGFFWSSDRWLPQKEFWFWKRRIEKKLREQLO
187	Bv3Cas12b aa sequence	MAIRSIKLMKMTNSGTD SIYLRKALWRTHQLINEGIAYYMNLLTYRQEA IGDKTKEAYQAEINII RNQQRNNGSSEEHGSDQEI LALLRQLYELI IPS S I GESGDANQLGNKFLYPLVDPNSQSGKGT SNAGRKPRWKRLKEEGNPDW ELEKKKDEERKAKDPTVKI FDNLNKYGLLPFLPFTNIQKDI EWLPKGKR QSVRKWDKDMFIQAIERLLSWESWNRRAVEYKQLKEKTESYKHELTGG EEWIEKIRKFEKERNMELEKNAFAPNDGYFITSRQIRGWDRVYEKWSKL ESASPEELWKVVAEQQNMSEGFGDPKVF SFLANRENRIWRGHSERIYH IAAYNGLQKKLSRTKEQATFTLPDAIEHPLWIRYESPGGTNLNLFKLEEK QKKNYYVTLSKI IWPSEEKWI EKENIEI PLAPSIQFNRIKLGQHVKGKQ EISFSDYSSRISLDGVLGGSRIQFNRKYIKNHKELLGEGDIGPVFFNLV DVAPLQETRNGRLQSPIGKALKVISDFSKVIDYKPKELMDWMNTGSASN SFGVASLLEGMRVMSIDMGQRTSASVSI FEVVKELPKDQEQKLFYSINDT ELFAIHKRSFLLNLPGEVVTKNNKQQRQERRKKRQFVRSQIRMLANVLRL ETKKTTPDERKKA I HKLMEIVQSYDSWTASQKEVWEKELNLLTNMAAFNDE IWKESLVELHHRIEPIYVGQIVSKWRKGLSEGRKNLAGISMWNI DELEDTR RLISWSKRSRTPGEANRIETDEPFGSSLLQHIQNVKDDRLKQMANLIIM TALGFKYDKEEKDRYKRWKETYPACQII LFNENLNRYLFNLD RSRENSRL MKWAHRSIPRTVSMQGEMFGLQVGDVRSEYSSRFHAKTGAPGIRCHALTE EDLKAGSNTLKR LI EDGFINESLAYLKKGDI IPSQGGELFVTL SKRYKK DSDNNELTVIHADINAAQNLQKRFWQQNSEVYRVPCQLARMGEDKLYIPK SQTETIKKYFGKGSFVKNNTEQEYKWEKSEKMKIKTDTT FDLQDLGFE DISKTI ELAQEQKKYLT MFRDP SGYFFNNETWRPQKEYWSIVNNI KSC LKKKILSNKVEL
188	PteCas12b aa sequence	MVVRSIKLMKMTNSGTD SIYLRKALWRTHRLINEGIVYMSLLTYRQDT LGDRTKEEIQSELIKKIREQQRNNGLSEELGSDQEI LSLRHLYELI IPS INGESGDANQLGNKFLYPLVDPNSQSGKGT SNAGRKSKWKRMKEEGNPDW EVEFKKDEERKANDPTIKVFDHLKKYSLLPFLPFTNNQKDI EWLPMGKR QSVRKWDKDMFIQAIERLLSWESWNRRLGAEREKLEEKIENFYKHELSGG QIWI EKIREFEVRDRELGETSFSSNDGYLITSRQIRGWDRVYEKWSKIS ESASKEELWRVVAEQQSKMREGFGDPKVF SFLAELENMDIWRKHPERIYH IATYNGLLKKLSHTKAQATFTLPDAVKHPLWIRYEAQGGTNLNLFKLEET TKKNCKVILSKI IWPTEGWF EKENVEVDLAPSKQFYRQIKLQDHIKGKQ EISFSDYSSGISLKGVLGGSRIQFDRKYIENHQELLPSGDIGPVFFNLVI DLLPIQETRNGRLKSPIGKTLKVVSS EFPKVIDYKPKELTEWINVSVTG KVGVESTIEGMRVMSIDLQRTSASVSI FEVVKELPKDKEKMLYNIKDT ELFALHKRSFLLNLPGEVTKRNRQKRKDRRKLLFIRSQIRMLASVLKL ETKNTTPDERKKAINKLVEIVDSYEWTESEKEIWNSEFEYLTNKAVFKQEI WRESLIKSHHRMEPHVGQLVSEWRKSLNEGRNLAGITMWNIEELEDTRR

TABLE 8		
SEQ	Description	Sequence
		LLISWSKRSRTPGEANRINNDPEFGAKLLEHIQNVKDDRLKQLANLIVMT ALGYKYDKEEKSRDKRWKEKYPACQVILFENLNRYLFSLDRSKRENSKLM KWAHRISIPRTVWMQGEMFGLQVGDVRSEYSSRFHAKSGAPGIRCHSLNDE DLKEESFKLKHLEIETGFISEEEISSLKKGDIVPWPGGELFVTLSPYKKG KDIELTVIHADINAAQNLRWFQQNSEVYRIPCQLEKAGDDEFFIPKSQ TEIVKKYFGKGRFVKINDKKEVYNWEESEKMKIKTESTITLQDLEGFEDI FQTLLELAQEQQKYSTLFRDPSGYFFNEKNWRPQKEFWSIVNNIIRSSLK KKILKNKVEV
189	BcoCas12b aa sequence	MAIKSMKLLKTKSGSNALSIRKGLWKTHEMNAGIAYMEWLTLRQES IGLRSKEELRLELILRLKRQQQNGYVGDSSNIPEEVFTVLRELYECIVP SSVRQSGDAQVLSRKYLSPLVDPKSKGEGESKAGRKPGWLLKQEAAGDPS WEQDYEKAKKRKESDPTARLMQKLREYGLKPLFPLFNEQKEIQWLPLKE NQYVRTWDRDMFQQAIERLLSWESWNRLRLKDERDELQKAVRFEQNYLID ADEWMEPLKQYELARAKELAQVAEAPVTDPMITKRQIRGWKQLEKWKGL DKNASEEDFIAIIAEVQSSMPKEFGDPIFRFLARPENHWIWRDHQDRLF LFQTYNELKRRLAQVKEQATFTLPDPVNHPLWIRFDARGGNLHDYDLWQE SRKRSRQRTVTFFSLIMPSDQGWEEQADVEVEIALSKQFYRQVRIQDHTK GKQEIIFYDYSVHSGKPANIPHLGGLGAKIQFDRKHLEKNRDKVALGEI GSVFLNVTVDIEPFQPLKNGRLQTPLGQVLKVLPEWPKVIEYKPSLEN WWKETLDAQILSTEQKKGIESLSAGMRIMTVDMGIRSSAAVSI FEIATER PTDSSKLCFRLPDNDLYAVHCRSLLVNLPGEKPKDRI REARELRTNQRYG VRQLIRMLSNIQRLHSRETEAERLKAVTDLEQALWQENVTQVERDQLIP VLRELLQRTADPDVWTEQIEKTYRELERLVGAALTWKWKSFGPGRNLA GLSMWNI EELESRLRMLISWSKRSRRPQEKHNLQEKEQFAQGILLTHIQEV KDNRLKQMANLIVMTALGYKYVDKHAKWVASYPACQIILFEDLSRYRMKQ DRSRMENSLLMKWAHRISIPRHTWMQGEFGLQVGDVRSEFSSRFHARTGA PGIRCHVVTEKDINNPLFKDQLLRKNFLKEEQFYLPQPGDIVPMQGGELF VTLSPNSQDVILIHADINAAQNLRWFWRTRNQEIFRIVCQAVEYEGNVA FVPKYEKRLGKGLLVKRFADQVYKWDQAQAKLSKKALPDSSYSEEDGEE SFEGLEEAEVRGEYKTLFRDPSGTFFPSDTWRPQVEFWGIVKARLEKLL REKILTGR
190	LseCas12b aa sequence	MSIRSFKLKIKTKSGVNAEELRRGLWRTHQLINDGIAYMNWLVLLRQED LFIRNEETNEIEKRSKEEIQGELLERVHKQQQRNQWSGEVDDQTLLQTLR HLYEEIVPSVIGKSGNASLKARFFLGPLVDPNNKTTKDVSKSGPTPKWKK MKDAGDPNWVQYKYMAERQTLVRLEEMGLIPLFPMYTDEVGDIHWLPQ ASGYTRTWDRDMFQQAIERLLSWESWNRRVRERRAQFEKKTDFASRFSE SDVQWMNKLREYEAQQEKSLEENAFAPNEPYALTKKALRGWERVYHSWMR LDSAASEEAYWQEVATCQTAMRGEFGDPAIYQFLAQKENHDIWRGYPERV IDFAELNHLQRELRRAKEDATFTLPDSVDHPLWVRYEAPGGTNIHGYDLV QDTKRNLTLILDKFILPDENGSWHEVKKVPFSLAKSKQFHRQVWLQEEQK QKKREVVFYDYSTNLPHLGTLAGAKLQWDRNFLNKRTOQQIEETGEIGKV FFNISVDVRPAVEVKNGRLQNLGKALTVLTHPDGTKIVTGWKAQLEKW VGESGRVSSLGLDSLSEGLRVMSIDLQRTSATVSVFEITKEAPDNPYKF FYQLEGTELFVHQRSLALPGENPPQKIQMREIRWKERNRIKQOVDQ LSAILRLHKKVNEDERIQAIKLLQKVASWQLNEEIATAWNQALSQLYSK AKENDLQWNQAIKNAHHQLEPVVGKQISLWRKDLSTGRQGIAGLSLWSIE ELEATKKLLTRWSKRSREPGVVKRIERFETFAKQIQHHINQVKENRLKQL ANLIVMTALGYKYDQEQKKWIEVYPACQVVLNLSYRFSYERSRRENK KLMWESHRSIPKLVMQGEFGGLQVADVAAAYSSRYHGRTGAPGIRCHAL TEADLRNETNIIHELIEAGFIKEEHRPYLQQGDLVPWSGGELFATLQKPY DNPRIILTLHADINAAQNIQKRFWHPMSWFRVNCESVMEGEIVTYVPKNKT

TABLE 8		
SEQ	Description	Sequence
		VHKKQGKTRFRVVKVEGSDVYEWAKWSKNRNKNTFSSITERKPPSSMILFR DPSGTFFKEQEWEVQKTFWGKVQSMIQAYMKKTIVQRMEE
191	BheCas12b aa sequence	MTVRSIRVKLAVGSPQYRDVRRGLWKTHEIMNQGVRYICEWLVMRQEP YDEDEHGLTVVQRTREDIQAELL SRLRTLQSAHQHSGDMGTDELLSLMR QLYEQLVPSVDKNKSGDARMIARNFFNPLTNPNSSQGLGISNAGRKPWK LLKKLSGDPTWEEDYKKAMEQKQESSVSFLLELRRFGLHPIFLPYTDTV LEVSWAPKKARQWVRKWDYDLFQQSIERMLSWESWTRRVKERFEKLVESE KKFYDENFATDPEFIKLAETLEGELOASSQGFVAVDEHAFQIRPRSMRGF DRVADEWCKLADDAPIEEYEAAIKRVQARLGRNFGSYVLFHAHLAKPEYWS LWRSDPTKILRFARLRALQRAVARAKRHARLTLPDAIHHPWIWIRYDAKGG NIYSYRLLIPEKRSKRYVEFSSLIMPDGENRWAHEHNRIRVPLAFSRQWE RLHFSIMEDGSLCVYRDPGVDEPLRAELGGAKIQFDRRYLIRRSSTLSA GECGPVYLVNSVDVNPAPHRPDVQVLSAKLVSVSRDNTRIYLRPENLSAY WKSQGDGTLPLRVMSVDLGVRSSAAVVICRLEHRDSVVSSGRRTATYRI AGTDEFVAVQERAFLLRLPGEGKGTNEDAPLRDVYAQLGTIRQGIQILRS LLRLCDTKTPDERQEALHGLAQSLPSGAWKDELHPLHVLMLQGVVHDSVD NWKQKVISVHRQMERILGHAVREWKVARKNAGKPPIRRAGGLSLRRIRQ LEQERRTLVAWSNHAREPGQVVRIRKGTQVAQWLVERVNHKLKEDRLKKLA DLLIMTALGYVYDETKPSGHKWDKRYPPCQIILMEDLSRYRFQSDRPPSE NSQLMAWSHRRLLEILKLQADLHKLIVGTVPFPAFSSRFDAQSGAPGVRRCR SVKKQDIENAAQGGKWLARELQRLNWTLEWLQPNDLIPTGDGELFVTPAC CDRQKGIKIVHADLNAAQNLQRRFEWGGHAESLCRVTCDVVERDGRYAVP RISNAFADSFYKVFQGVFVSTDEEDVYRWVGEKISSRGRSRGRTSDEE AEAETWIDEAREQQGKVIALFRDASGQIHGGDWLVAKVFWGWERLVTAR LLSRMSEREAAHKE
192	BmeCas12b aa sequence	MAVRSMKLMKMTNTGQDSIHLRKALWRTHRLVNEGIAAYMDLLTLFRQDV IGDKTKEETQSELLQKIREQQRRNNGSLKEIGSDQEILTLLRQLYELIIPS AIGESGDANQLGNKFLYPLVDPNSSQAGKGTSKAGRKPWRKRMQEEGDPNW ETEKQKDDERKAKDRTVKILNNLNNLGLPLFTLFTSKQKDI EWLPMGKR QSVRKWDKDMFIQAIERLLSWESWNRVADEREKLEKVENYYKENLSGG EKWIEKIREFEKDRDKELAQNAFASNDGYLITSRQIRGWDRVYEKWSKLP ESASKEELWKVVAEQSKMSEGFDPKVFSLAELENRDIWRGHPERIYR IAVYNGLLKRLSQTREQATFTLPDAIEHPLWVRYEAQGGTNLNLFKLEQN SKKELKVLISKMIWPKEEGWFEKEKVEITLAPSKQSIRQIILEEHIKKGQ EISFSDYSSGIPKGVLGGSRIQFNKHLNHNREFLTAGDIGPVFLNLVI DVTPLQEVKNRQLQSPIGKALKIVPSEYPKVIDYKPKELIEWINNSSATS KVGVESLTEGMRIMSIDMGQRTSASVIFEVVKELPKDKERRLFYNIKDT DLYALHRRSFLLNLPGEVVTKRISQEREERRKNRLIVRAHIRMLASVLKL GTKKTPDERKKAMNKLENIELNNLWSVADKKAWIREFNELLNMAFYNEE IWKENVIQLHRRMEPYVGQIVSKWRRGLSDGRNLAGISMNIDELEDTR RLISWSKRSRTPGEANRIENDEPFGTNLLEHIQNVKEDRLKQMANLIVM TALGYKYDEGEPNKDKRWKETYPACQVILFENLNRYLFNLDRSRRNSKL MKWAHRSIPKTVYMQGEMFGLQVGDVRSFSSRFHAKTGAPGIRCRALTE EDLDGGSYVLNQLIESNFINEINHLRKGDIVPWEGGELFATLSKPYKK ESDDNELTVIHADINAAQNLQKRFWQQNSEVFRVSCQLAKVGDDDEVYVPK SESVKKYLKGKGFVKVSDKQEVYNLEKSDKIKIKIKIKTDSTIELQDLDELG ETIAQTIELAEQQKKYVTLFRDPSGYFFNNGTWRPQKEFWSIVNNIIRNC LKKKILKNKVEV
193	BcaCas12b aa sequence	MATKSFILMKMTKNNPQLRLSLWKTHELFNFGVAYYMDLLSLFRQKDLYM HNDEDPDHPVVLKKEETIQLRLWMKVRETQQKNGFHVSKDEVLETLRAL YEELVPSAVGKSGEANQISNKYLYPLTDPASQSGKGTANSGRKPWRKKLK

TABLE 8		
SEQ	Description	Sequence
		EAGDPSWKDAYEKWEKERQEDPKLKILAALQSFGILPLFRPFTENDHKAV ISVKWMPKSKNQSVRKFDKDMFNQAIERFLSWESWNEKVAEDYEKTVSIY ESLQKELKGISTKAFEIMERVEKAYEAHLREITFSNSTYRIGNRAIRGWT EIVKKWMKLDPSAPQGNLYLDVVKDYQRRHPRESGDFKLFELLSRPENQAA WREYPEFLPLYVKYRHAEQRMKTAKKQATFTLCDPIRHPLWVRYEERSGT NLNKYRLIMNEKEKVQFDRILCLNADGHYEEQEDVTVPPLAPSQQFDDQI KFSSSEDTKGKHNFSYYHKGINYELKGTGGARIQFDREHLRRQGVKAG NVGRIFLNVTLNIEPMQPFSSRGNLQTSVGKALKVYVDGYPKVVNFKPKE LTEHIKESEKNTLTGVESLPTGLRVMSVDLGQRQAAASIFEVVEKPD DNKLFYPVKDSDLFAVHRTSFNIKLPGEKRTERRMLEQQKRDQAIRDLRS KLKFLKNVLMQKLEKTDEREKRVNRWIKDREREENPVYVQEFEMISKV LYSPHSVWVDQLKSIHRKLEEQLGKEISKWRQSSISQGRQGVYGISLKNIE DIEKTRRLFLRWSMRPENPGEVKQLQPGERFAIDQQNHLNHLKDDRIKKL ANQIVMTALGYRYDGKRKKWIAKHPACQLVLFEDLSRYAFYDERSRLNR NLMRWSRREIPKQVAQIGGLYGLLVGEVGAQYSSRFHAKSGAPGIRCRVV KEHELYITEGGQKVRNQKFLDSLVENNIIEPDDARRLEPGDLIRDQGGDK FATLDERGELVITHADINAAQNLOKRFWTRTHGLYRIRCESREIKDAVVL VPSDKDQKEKMNLFYGYLQPFKQENDVYKWKGEIKGKKTSSQSDDK ELVSEILQEASVMADELKGNRKTFRDPSGYVFPKDRWYTGGRYFGTLEH LLKRKLAERRLFDGGSSRRGLFNGTDSNTNVE
194	BpaCas12b aa sequence	MAIRSIKLLKTRTGPEAQNLKGIWRTHRLNNEGVAHYMKMLLLFRQES TGGQTKKELQEELVRHIREQQQKNRADKNTQALPLDKALEALRQLYELLV PSSIGQSGDAQIISRKFLSPLVDPNSEGGKGTSGKAGAKPTWQKKKEANDP TWEQDYEKWKRRREEDPTASVITLLEEYGRPIFFPLYTNTVADIAWLPLQ SNQFVRTWDRDMLQQAIERLLSWESWNKRVQEEYSKLQEKMTQLNEQLEG GQEWISLLEQYEEQREQELIENMTAANDKYRITKRQMKGNWELYEQWSTV LPNASHEQYREALKRVQORLRGRFGDAHFFQYLMKEEHHLIWKGNPQRIH YFVARNELKKRLEEAKQATMTLPDARKHPLWVRFDARGGNLQDYLLTAE ADNPRSRRFVTFSQLIWPNESGWMEKQDVEVELALSRYFYQVYLTQKNDK GKQEIFKDKGSGSTFSGHLGGAKLQLERGDLEKEEKDFEGGETIGSVYLN IVIDFEPLQEVKNGRLQSPYGGVLQLVRRPNEFPKVTYTKSEELVEWMKA SQNHSSGVESLESGRVMSIDILGLRTAAATSIIFSVEESNDANAAGFSYWI EGTFLVAVHKRSYMLKLPGEQVEKQVREKRDERQDQQRVRVFIIRLSQV IRMAKKQNRERADELHLSQALEKQKSLDQTDRTFWNGIVCDLTDALRE KEGGWEQAVVQIHRKAEHVGVVQAWRKRFADERKGIAGLSMWSIEEL DSLRLKLLISWSRRTNRNPQEIINRFEQGHTSHQRLTHIQNVKEDRLKQLSH AIVMTALGYVYDEKKLEWFAKYPACQVILFENLSQYRSHMDRSTKENSTL MKWAHRSIPKYVHMQAEPYGIQIGDVRAEYSSRFHAKTGTGPIRCKMVSG HDLQGRRFENLQKRLISEQFLTEEQVKQLRPGDIVPDDSGEWFVTLTNGS GDKEVLLQADINAAQNLOKRFWQRYNELFKVSCRVLIRGEEYLIPIKTK SVQAKLGKGLFVKKTDTVIKDVYVWDSQAKLKGKTTTTEESPEQLEDF QEIEEEAEAKGTYRTLFRDPSGVFFPEFVWNTQKDFWSEVKRRLYGKLR ERFLMKTR
195	BpuCas12b aa sequence	MAVRSFKLKIMAGQNEKLRKALWKTHELHNYGVAYYMEWLSLLRQEDLFE LNEEITTQDTPSRTKERLQEELWTRVREAQHRNEFTSVVNKQEVLETLR LYEQLVPSAAGESGEANQICNKYLYPLTDANSQSGKGTASSGRKPRWKNL KGAGDPSWEEKKKWAQKQKDPKLQIMNRLDSYGLPLFPLFTDSEDFF VRDITWLPKSKKQSVRKWDKDMFNQAIERFLSWESWNQVKTEYEELASK YKSLKATLIQMDSKAFDALGSFEEKRIEELKNITTFHNSTYYLGTRELGR WKVIVDKWIRFSENKTFADYIEVYKDYQRSHSRESGDFEVYNFLSHPENH FIWRNNKEFFFLYAKYSETKLKLMNAKKQATFTLSDPIEHPLWVRFEERS

TABLE 8		
SEQ	Description	Sequence
		GTNLNKYKMITSDQEKESEKRRVPLTVEVDRFIVPNGEDGYLEEAKYKLQ LAPSRQFYNQVLFSEDEGKGKHQFKYVDEATGMELNGYLGGAIRIQFDRN YIRRHNSQVAKANVGKIYFNMTLNIIVPLQEIGRTGRLQTAVGKALSTYND DYLKVVNFKPKELTELISQSKKLPLVKGPDLSLKVGLRIMSVDLGGQRQAAA VSFFEVS DIK PENKLYPIKDTLFAVHNRSNLKLAGEKRTTKREKIQN KRDERIRELSRKLTLFRNINLNLQLVENVEERKKKVGRWLDREDSTQKELY EENQSKLKNVLYSPQDVWIKTLKEIYSKLEHSIGREIHEWRSTISDDREG VYGISLKNIEEIERSRRLLSWSNRSTEPGQPKRLEKGRFAIDQQVHLN DLKDDRIKKMANLLVMTALGYKYRGKHKRWVAERPACQVVLFDLSEYGF REERSRQENSKLMRWSRREIPRQVALQGEYGLQVGDIGAQQFSSRFHAKT GAPGIRCHKLTMDMQNDWMKKDLIQRGFTKEEQIELIKAGDYIPSKGGE KFATLSRDRSLILTADINAAQNLRKFRWTRNHGFFRISCYVIQSDDGQI LVPKEYTKKRLQELYGSSSKGYFIMIDDNKGEVYKVVSRDKLKQKVSLSKSK RTSEETEAMNDIFEIAEEISGESITLYRDPGQMFRSDLWYTGGRYFGTI EGRIKKQLKQRIQGGGLKRPDIAEDDEEWDLFL
196	BhiCas12b aa sequence	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNIKLIRQEAIYEH EQDPKNPKKVSKAIEQAELEWDFVLKMQKCNSTHEVDKDEVFNILRELYE ELVPSSVEKKGEANQLSNKFYPLVDPNSQSGKGTASSGRKPRWYNLKIA GDPSWEEKKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPTDSNEPIVKE IKWMEKSRNQSVRRLDKDMFIQALERFLSWESWNLKVKEEYKVEKEYKT LEERIKEDIQALKALEQYEKERQEQLLRDTLNTNEYRLSKRGLRGWREII QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRN HPEYPYLYATFCEIDKKKKDAKQATFTLADPINHPLWVRFEERSGSNLN KYRILTEQLHTEKLLKKLTVQLDRLIYPTESGGWEEKGVDIVLLPSRQF YNQIFLDIEEKKGHAFTYKDESIKFPLKGTGGARVQFDRDHLRRYPHKV ESGNVGRIYFNMTVNIPTESPVSKSLKIHRDDFPKVVNFKPKELTEWIK DSKGKLLSGIESLEIGLRVMSIDLQGRQAAAASIFEVVDQKPDIEGKLF FPIKGTELYAVHRASFNIKLPGETLVKSREVLKAREDNKLKMNQKLNFL RNVLHFQQFEDITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPY KDWAFLKQLHKKRLEVEIGKEVKHWRKSLSDGRKGLYGISLKNIDEIDRT RKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKKMANTII MHALGYCYDVRKKKQAKNPACQIILFEDLSNYPYERSRFENSKLMKW SRREIPRQVALQGEIYGLQVGEVGAQQFSSRFHAKTGSPGIRCSVVTKEKL QDNRRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKCVTTH ADINAAQNLRKFRWTRTHGFYKVYCKAYQVDGQTVYIPESKDQKQKIEE FGEGYFILKDGVEWVNAGKLKIKKGSSKQSSSELVSDILKDSFDLASE LKGEKLMLYRDPGSGNVFSPDKWMAAGVFFGKLERILISKLTNQYSISTIE DDSSKQSM
197	BgeCas12b aa sequence	MAIRSIKLLKLTNTGPEAQNLKGIWRTHRLNNEGVAYYMNMILLFRQES TDKKTQEIHEELIRHIRAQQRNHADEKTQALPLEKALEALRKLYELLV PSSVGQSGDSQIISRKFLSPLVDPNSEGGKGTSKAGAKPAWQKKKEANDP TWKQDYEKWKKRREEDPTASVITTL EEYGI RPLFPLYTNTVAEIAWLPLK SGQFVRTWDRDMFQQAIEGMLSWESWNRVQEEYAKLEGKMAQLNEQLEG GEEWIRLLEQYEEKREQELRENMTAANDKFRITKRQMKGWKELYEVWSTF LPSASQEQYKEAIKRQQRLRGKFGDFHFFQYLS EENRLIWKGNPQRIH YFVARNELTKKLEKAKQSARRTLDPANKHPLWVRYDARGGNLQDYLYTAE SDKPRSRRFVTFSQLIWPSESGWLEKKDVQAEALASRQFYQQVTFKNDK GKQEIFKDKSGSTTFSGHLGGAKLQLERSVLENKERKFEEGEIGKAYLN VAIDFKPLQEVKNRQVQAPYGGVQLLIRLPNAPFKVPTYKSEELVEWIK SPQHLSGVESLASGFRVMSIDLGLRAAAATSI FSVEESSDKNATKLAYWI EGTPLVAVHQRSYMLRLPGEQVEQHVWEKRD ERGDQHKRVRFQIRRLAEI

TABLE 8		
SEQ	Description	Sequence
		IRLANKQYGRWDELNRLDEAVSKEKSPLDQADRTFWEGIVSDLTALPL NDADWTEAVVQIHRKAELYVGKVVQAWRKRFNADERKGIAGLSMWSIEEL DGLRKLISWSRRTRNPQEVNRFEPHTGHRLLTHIQNVKKDRLKQVSH AIVMTALGYIYDEKKQKCAKYPACQVILFENLSQYRSNLDRAKENSTL MKWAHR SIPKYVHMQAEPYGIQIGDVRAEYSSRYAKTGTGPIRCKKLRE HDVKGWRLDHLKKRLVNEQFLTEAQVEQLKAGDIIIPDDSGELFMTMTDGS GGKEIVFLQADINAAQNLOKRFWQRNSELFKVSCRIVRDEAEYLPQAK KVQEKLGKGVFVKKSDTAWKEVYVWDSQAKLKGGKTTTTEESSEPEQLEDL QEMI EEAEAEAKGTYRTLFRDP SGVFFPDEFVWNTPKDFWGEVKKRLYGLR ERLLTKVR
198	BkaCas12b aa sequence	MAVKSIKVKLRLSECPDILAGMWQLHRATNAGVRYTETWVSLMRQEILYS RGPDGGQQCYMTAEDCQRELLRLRLNRQLHNGRQDQPGTDADLLAISRL YEILVLQSIGKRGDAQQIASSFLSPLVDPNSKGGGEAKSGRKPWQKMR DQGDPRWVAAREKYEQRKAVDPSKEILNSLDALGLRPLFAVFTETYSRGV DWKPLGKSQGVRTWDRDMFQQALERLMSWESWNRRVGEYARLFQQKMKF EQEHFAEQSHLVKLARALEADMRAASQGF EAKRGTAHQITRRALRGADRV FEIWKSIPEEALFSQYDEVIRQVQAEKRRDFGSHDLFAKLAEPKYQPLWR ADETFLTRYALYNGVLRDLEKARQFATFTLPDACVNIWTRFESSQGSNL HKYEFLLFDHLGPGRHAVRFQRLLVVESEGAKERDSVVVPVAPSGQLDKLV LREEKSSVALHLHDTPARPDGFMAEWAGAKLQYERSTLARKARRDKQGM SWRRQPSMLMSAAQMLEDAKQAGDVYLNISVRVKSPSEVRGQRPPYAAL FRIDDKQRRVTVNYNKLSAYLEEHPDKQIPGAPGLLSGLRVMSVDLGLRT SASISVFRVAKKEEVEALGDGRPPHYPIHGTDDLVAVERSHLIQMPGE TETKQLRKLREERQAVLRPLFAQLALLRLVRCGAADERIRTRSWQRLTK QGREFTKRLTPSWREALELELTRLEAYCGRVPDDEWSRIVDRVTIALWRR MGKQVRDWRKQVKSGAKVKVKGQYQLDVVGGNSLAQIDYLEQQYKFLRRWS FFARASGLVVRADRESHFAVALRQHIENAKRDRLLKKLADRIIMEALGYVY EASGPREGQWTAQHPPCQLIIIEELSAYRFSDDRPPSENSKLMWGHGRG LEELVNQAQVHDVLVGTVYAAFSSRFDARTGAPGVRCRRVPARFVGATVD DSLPLWLTEFLDKHRLDKNLLRPDDVIPTGEGEFLVSPCGEEAARVRQVH ADINAAQNLOQRLWQNFDTITELRLCDVKMGEGTVLVRVNNARAKQLF GKKVLVSQDGVTFEERSQTGGKPHSEKQTDLTDKELIAEADARAKSV VLFDRPDSGHIGKGHWIRQREFWSLVKQRIESHTAERIRVRGVGSSLD
199	BthCas12b aa sequence	MATRSFILKIEPNEEVKKGLWKTHEVLNHGIAYYMNILKLIRQEAIEYHH EQDPKNPKKVSKAIEQAELWDFVLKMQKCSFTHEVDKDEVFNILRELYE ELVPSSVEKKGEANQLSNKFLYPLVDPNSQSGKGTASSGRKPRWYNLKIA GDPSWEEKKKKWEEDKKKDPLAKILGKLAEYGLIPLFIPTDSNEPIVKE IKWMEKSRNQSVRRLLDKDMFIQALERFLSWESWNLVKVEEYKVEKEYKT LEERIKEDIQALKALEQYEKERQEQLLRDTLNANEYRLSKRGLRGWREIV QKWLKMDENEPSEKYLEVFKDYQRKHPREAGDYSVYEFLSKKENHFIWRN HPEYPYLYATFCEIDKKKKDAKQATFTLADPINHPLWVRFEERSGSNLN KYRILTEQLHTEKLLKLTQVLDRLIYPTESGGWEEKGKVDIVLLPSRQF YNQIFLDIEEKKGHAFTYKDESIKFPLKGT LGGARVQFDRDHLRRYPHKV ESGNVGRIYFNMTVNIEPTESPVSKSLKIHRRDDFPKVVNFKPKELTWIK DSKGKLLSGIESLEIGLRVMSIDLQGRQAAAASI FEVVDQKPDIEGKLF FPIKGT ELYAVHRASFNIKLPGETLVKSREVLARKAREDNLKL MNQKLNFL RNVLHFQQFEDITEREKRVTKWISRQENS DVPLVYQDELIQIRELMYKPY KDWVAFLLQHLHRLEVEIGKEVKHWRKSLSDGRKGGLYGISLKNIDEIDRT RKFLLRWSLRPTEPGEVRRLEPGQRFAIDQLNHLNALKEDRLKMKMANTII MHALGYCYDVRKKKWKQAKNPACQIILFEDLSNYPYEERSRFENSKLMKW SRREIPRQVALQGEIYGLQVGEVGAQFSSRFHAKTGSPGIRCSVVTKEKL

TABLE 8		
SEQ	Description	Sequence
		QDNRRFFKNLQREGRLTLDKIAVLKEGDLYPDKGGEKFISLSKDRKLVTTTH ADINAAQNLOKRFWTRTHGFYKVYCKAYQLDQGTVYIPESKDQKQKIIIEE FGEGYFILKDGVIYEWGNAGKLKIKKGSSKQSSSELVSDILKDSFDLASE LKGEKLMLYRDPSPGNVFPSPDKWMAAGVFFGKLERILISKLTNQYSISTIE DDSSKQSM
200	B-Gen.2(D501R)- NLS aa sequence	MPIRSFKLKLVTHNGDSTYMDKLRRLWKTTHVIINRGIAYYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQQRNNWSEQTGTDDLLSLSRVYELLV PSAIGEKGDAQMLSRKFLSPLVDPNSEGGRGTAQSGRKPFWKKMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLPLPLFPLFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERMLSWESWNRRAVEEYLKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQLRMVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFFSSILPEAQSWVEIENVTVGIGNSLQKQIRLDGYADKKQ KVKYDYSSRFELTGVLGGAKIQFDRKHLKKAHRLAEGETGPIFLNVVV <u>R</u> VEPFLEVKNRRLRTPLGQVLQVNTRDWPKVVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLRLPGEIESSKKMIEKRKERARIRSLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELLVSIQKKLELDQSEWISELEKLFDYIDESAEKW KEALVVAHRTLEPIVVEAVRNWKSLSKENKDRRRIAGISISWIEELEET RKLIIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHRSIPLRTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNYSVVNQLIKDKFIQDNQTEILKAGQIVPWQGGELFVTFADRSGLA VIHADINAAQNLOKRFWQHNSVFRVPCKVVKGLVPVYEKMRKLFQKGL FVNIDDPESKEVYRWEHSTKMKSCTTPVDLESEDIDHEELSDEWEDMQEG YKTLRLRDPGFFWSSDSWIPQKDFWIRVKSRIKSLREQIRGGSGSGSGS <u>GP</u> <u>PKKKRKV</u>
201	B-Gen.1 (D504R) target-interacting strand aa sequence	GPAFLNVVLRL
202	B-Gen.1.2 (D501R) target-interacting strand aa sequence	GPIFLNVVVRV
203	B-Gen.2 (D501R) target-interacting strand aa sequence	GPIFLNVVVRV
204	Consensus target- interacting strand aa sequence	GX ₁ X ₂ X ₃ X ₄ NX ₅ X ₆ X ₇ DX ₈ X ₁ : any amino acid; in some embodiments X ₁ is D, E, S, P, K, or R. X ₂ : any amino acid; in some embodiments X ₂ is V, I, or A. X ₃ : any amino acid; in some embodiments X ₃ is Y or F. X ₄ : any amino acid; in some embodiments X ₄ is L or F. X ₅ : any amino acid; in some embodiments X ₅ is I, L, F, V, or M. X ₆ : any amino acid; in some embodiments X ₆ is S, V, T, or A.

TABLE 8		
SEQ	Description	Sequence
		X₇: any amino acid; in some embodiments X₇ is V, L, or I. X₈: any amino acid; in some embodiments X₈ is V, F, L, or I.
205	B-GE consensus II aa sequence	MPIRSFKLKLVTHNGDSTYMDKLRRLGWLKTHVIINRGIAYYMNTLALMRQ EPYGSKSREEVRLDLLSTLREQRRNNWSEQTGTDELLSLSRVYELLV PSAIGEKGAQMLSRKFLSPLVDPNSEGGRGTAAGSRKPRWKKMMEEGHP DWEKEKEKDAAKKAEDPTASILADLEAVGLLPFLPFSDEQKEIRWLPKK KRQFVRTWDRDMFQQALERMMLSWESWNNRRVAEEYXKLQAQRDEVYAKYLE DAGSWLNDLQTFEKQREEELAEVSFEPNSEYLITRRQIRGWKEVYEKWSK TSENASQEQLRWVADVQTAMAGAFGDPKVYQFLSQPKHHHIWREHPNRL FYYSKYNEVREKLNRAKKQAFTLPDPVEHPLWTRFDARGGNIHDYEISK VGKQYHVTFSLLPEAQSWVEIENVTVGIGNSLQLKQRIRLDGYADKKQ KVKYDYDSRFELTGLGGAKIQFDRKHLKKAHRLAEGETGPIFLNVVV DVEPFLEVKNRRLRTPGLQVLQVNRDWPVKVDYKAKELSVLMENTQIGN ENGVSTIEAGMRIMSIDLGQRTAAAVSIFEVISKKPDEKETKLFYPIADT DLYAVHRRSLLLRLPGEEISSKKMIEKRKERARISLVRYQIRLLSEVLR LHTQGTAEQRRFKLDELVSIXKLELDQSEWISELEKLFYDIDESAEKW KEALVVAHRTLEPIVVEAVRNWKKSLSKENKDRRRIAGISWISIEELEET RKLIIAWSKHSREPGIPKRLEKEETFAPEHLQHIQNVKDDRLKQMANLFV MTALGYKYDEGNKRWVEAYPACQVILFEDLSRYRFALDRPRENNRLMKW AHRISIRLTYMQAELFGIQVGDVYSAYTSRFHAKTGAPGIRCHALTEADL QSNSYVNVQLIKDKFIQDNQTEILKAGQIVPWQGGELFVTFADRSGASLA VIHADINAAQNQLQKRFWQHNSVEFRVPCKVVKGLVPVYEKMRKLFKGGL FVNIDDPESKEVYRWEHSTKMKSKTTPVDLESEDIXHEELSDWEDMQEG YKTLRLDPGSGFFWSDSWIPQKDFWIRVKSRIKSLREQIR
207	Bacillales Cas nuclease consensus aa sequence	MAXRSXKLKXXXXXXXXXXLRXGLWXTHTXXXNXGAXYMXLXLXLRQEX XXXXXXXXXXXXXXXXXTEEXQXELXXXXXXQXNXXXXXXXXDXEXLXLX XLYELLVPSXXGXSGDAQXXRKFLSPLVDPNSXGGKGTSKAGRKFXWXX XKEAGDPXWEXEXXKXKXDXPTAXXLXGLXPLFXXXTXXXXX IXWXPLXKXQXVRTWDRDMFQQAIERXLSWESWNNRRVXEEYXKLXXXXX XXXXXXXXXXXXXXXXXXEXXRXXLXXXXXXXXXXYXITXRXRGWXX VXEKWXKLXXXAXEXYXXXXXXVQXXXXXFGDXXXFXFLAXPENXXIW RXXPXXXXYAXYNXXXXKLXAKXXATFTLPDXXXHPLWXRFXFGGNL XXYXXXXXXXXXXRXVXFXXLIXPXXGXXEXXVXVXLAXSQFXXQX XXXXXXXXXGKXXXXFXDXXXXXLGXGLGGAKIQXXRXXLXXXXXXXXXGX XGXVYLNXXVDXXPXXEXXGRXXXPXGXXLXXXXXXXXPKVXXXKXXELX XWXXXXXXXXXXGXESLXXGLRVMSXDLGXRTSAAXSIFVXXXXXPXXXX XXXXYPIXGTXLXAVHXRSLXLPGEXXXXXXXRXRXRXRXRXRXQ XXXLXXXLRXXXXXXXXXXRXRXRXLXXXXXXXXXXXXXXXXXXELXX LXXXXXXXXXXWXXVXXHRLXXXXXGKXVXXWRKXXXXGRXXXXGX XXIEELEXXRKLXSWXRXRXPGEVXRXEXXXFAXXLLXHXIXXKXDR LXXLANXIVMTALGYXDXXXXXWAXYPACQXILFEXLSXYRFXDRSX XENSXLMKWXHRI PXVXXQXEXXGLQVGDVXAXFSRFHAKTGAPGIR CXXVXXDXLXXXXXXXXXXLXXXFXEXXXXXLXXGDXPXXGGEXFVTL XXXXXXXXXXHADINAAQNQLQKRFWXXXXXXXXRVXCCCCCCCCXXLP XXXXXXXXXXGKGXVXXXXXXVYXWXXXXKXKXXXXXXXXXXXXX EXXXXXXEAXEXXXXXTLFRDPGXXFXXXWXXQKXFWXXVXXIXX XLXXXIXXXXXXXXXX

10. INCORPORATION BY REFERENCE

[0281] All publications, patents, patent applications and other documents cited in this application are hereby incorporated by reference in their entireties for all purposes to the same extent as if each individual publication, patent, patent application or other document were individually indicated to be incorporated by reference for all purposes. In the event that there are any inconsistencies between the teachings of one or more of the references incorporated herein and the present disclosure, the teachings of the present specification are intended.

WHAT IS CLAIMED IS:

1. An engineered Type V endonuclease comprising an amino acid other than aspartic acid at the position corresponding to D504 of SEQ ID NO:1 (B-Gen.1) or D501 of SEQ ID NO:2 (B-Gen.1.2) or SEQ ID NO:3 (B-Gen.2).
2. The engineered Type V endonuclease of claim 1, which is an engineered B-Gen polypeptide.
3. The engineered Type V endonuclease of claim 1 or claim 2, which has an arginine at the position corresponding to D504 of SEQ ID NO:1 (B-Gen.1) or D501 of SEQ ID NO:2 (B-Gen.1.2) or SEQ ID NO:3 (B-Gen.2).
4. An engineered Type V endonuclease, which is optionally an engineered Type V endonuclease according to any one of claims 1 to 3, comprising an amino acid sequence which:
 - (a) has a substitution at the position corresponding to D501 of SEQ ID NO:3, wherein the substitution:
 - (i) is the substitution D501R as compared to the amino acid sequence of SEQ ID NO:3; or
 - (ii) increases gene editing efficiency activity as compared to a corresponding amino acid sequence without the substitution at position D501; and
 - (b) has at least 90% sequence identity to SEQ ID NO:1 (B-Gen.1), SEQ ID NO:2 (B-Gen.1.2), or SEQ ID NO:3 (B-Gen.2).
5. An engineered Type V endonuclease comprising the amino acid sequence of SEQ ID NO:4.
6. An engineered Type V endonuclease comprising the amino acid sequence of SEQ ID NO:5.

7. An engineered Type V endonuclease comprising the amino acid sequence of SEQ ID NO:6.

8. The engineered Type V endonuclease of any one of claims 1 to 7, which comprises at least one NLS, optionally located C-terminal to the amino acid sequence, further optionally wherein:

- (a) the engineered Type V endonuclease lacks any NLS N-terminal to the amino acid sequence; or
- (b) the engineered Type V endonuclease comprises at least one NLS located N-terminal to the amino acid sequence.

9. A nucleic acid comprising a nucleotide sequence encoding the engineered Type V endonuclease of any one of claims 1 to 8.

10. The nucleic acid of claim 9, wherein the nucleotide sequence encoding the engineered Type V endonuclease of any one of claims 1 to 8 is operably linked to a promoter.

11. The nucleic acid of claim 9 or claim 10, wherein the nucleic acid further encodes a guide RNA.

12. The nucleic acid of any one of claims 9 to 11, which is in the form of a vector.

13. The nucleic acid of claim 12, wherein the vector is a recombinant adeno-associated virus (AAV) vector.

14. A cell comprising the nucleic acid of any one of claims 9 to 13.

15. A cell engineered to express a nucleotide sequence encoding the engineered Type V endonuclease of any one of claims 1 to 8.

16. The cell of claim 14 or claim 15, which is a eukaryotic cell, optionally wherein the eukaryotic cell is a human cell.

17. A method of producing an engineered Type V endonuclease according to any one of claims 1 to 8, comprising culturing the cell of any one of claims 14 to 16 under conditions in which the engineered Type V endonuclease is produced, the method optionally further comprising isolating and/or purifying the engineered Type V endonuclease.

18. A composition comprising:

- (a) an engineered Type V endonuclease according to any one of claims 1 to 8; and
- (b) a guide RNA.

19. The composition of claim 18, which is a ribonucleoprotein complex.

20. The composition of claim 18 or claim 19, in which the molar ratio of the engineered Type V endonuclease:guide RNA molar ratio ranges from 1:1 to 1:4.

21. A method of editing the genome of a cell, comprising introducing into the cell:

- (a) the engineered Type V endonuclease of any one of claims 1 to 8; and
- (b) a guide RNA.

22. A method of editing the genome of a cell, comprising introducing into the cell:

- (a) one or more nucleic acids encoding the engineered Type V endonuclease of any one of claims 1 to 8; and
- (b) a guide RNA.

23. The method of claim 22, which comprises contacting the cell a lipid nanoparticle comprising the one or more nucleic acids and the guide RNA.

24. A method of editing the genome of a cell, comprising introducing into the cell one or more nucleic acids encoding:

- (a) the engineered Type V endonuclease of any one of claims 1 to 8; and
- (b) a guide RNA,

optionally wherein at least one of the one or more nucleic acids is a nucleic acid according to any one of claims 9 to 13.

25. The method of claim 24, which comprises contacting the cell with one or more recombinant AAV particles comprising the one or more nucleic acids.

26. The method of claim 24, which comprises contacting the cell with one or more lipid nanoparticles comprising the one or more nucleic acids.

27. A method of editing the genome of a cell, comprising introducing into the cell the composition of any one of claims 18 to 20.

28. The method of claim 24, wherein the composition is a ribonucleoprotein complex.

29. The method of claim 26 or claim 27, wherein the composition is introduced into the cell via a lipid nanoparticle.

30. The method of any one of claims 21 to 29, wherein the cell is a mammalian cell, optionally a human cell.

31. The method of claim 30, wherein the cell is a stem cell or a cell differentiated therefrom, optionally wherein the stem cell is a pluripotent stem cell (PSC), or a cell differentiated therefrom.

32. A cell comprising:

- (a) the composition of any one of claims 18 to 20; or
- (b) a nucleic acid according to any one of claims 9 to 13.

33. The cell of claim 32, which is a mammalian cell, optionally a human cell.