



US 20250250554A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0250554 A1**

Bowen et al.

(43) **Pub. Date:** **Aug. 7, 2025**

(54) **DNA MODIFYING ENZYMES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF AND METHODS OF USE**

(71) Applicant: **Life Edit Therapeutics, Inc.**, Durham, NC (US)

(72) Inventors: **Tyson D. Bowen**, Morrisville, NC (US); **Alexandra Briner Crawley**, Cary, NC (US); **Tedd D. Elich**, Durham, NC (US)

(73) Assignee: **Life Edit Therapeutics, Inc.**, Durham, NC (US)

(21) Appl. No.: **19/055,281**

(22) Filed: **Feb. 17, 2025**

#### Related U.S. Application Data

(60) Division of application No. 18/631,568, filed on Apr. 10, 2024, now Pat. No. 12,252,718, which is a division of application No. 17/929,162, filed on Sep. 1, 2022, now Pat. No. 11,981,940, which is a continuation of application No. 17/851,880, filed on Jun. 28, 2022, which is a continuation of application No. PCT/US2021/049853, filed on Sep. 10, 2021.

(60) Provisional application No. 63/146,840, filed on Feb. 8, 2021, provisional application No. 63/077,089, filed on Sep. 11, 2020.

#### Publication Classification

(51) **Int. Cl.**  
*CI2N 9/22* (2006.01)  
*CI2N 15/63* (2006.01)

(52) **U.S. Cl.**  
CPC ..... *CI2N 9/22* (2013.01); *CI2N 15/63* (2013.01); *CI2N 2310/20* (2017.05); *CI2N 2800/80* (2013.01)

#### ABSTRACT

Compositions and methods comprising novel deaminase polypeptides for targeted editing of nucleic acids are provided. Compositions comprise deaminase polypeptides. Also provided are fusion proteins comprising a DNA-binding polypeptide and a deaminase of the invention. The fusion proteins include RNA-guided nucleases fused to deaminases, optionally in complex with guide RNAs. Compositions also include nucleic acid molecules encoding the deaminases or the fusion proteins. Vectors and host cells comprising the nucleic acid molecules encoding the deaminases or the fusion proteins are also provided.

**Specification includes a Sequence Listing.**

## DNA MODIFYING ENZYMES AND ACTIVE FRAGMENTS AND VARIANTS THEREOF AND METHODS OF USE

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a divisional of U.S. application Ser. No. 18/631,568, filed Apr. 10, 2024, which is divisional of U.S. application Ser. No. 17/929,162, filed Sep. 1, 2022, which is a continuation of U.S. application Ser. No. 17/851,880 filed Jun. 28, 2022, which is a continuation of International Application No. PCT/US2021/049853, filed Sep. 10, 2021, which claims priority to U.S. Provisional Application Nos. 63/077,089, filed Sep. 11, 2020, and 63/146,840, filed Feb. 8, 2021, each of which application is incorporated by reference herein in its entirety.

### STATEMENT REGARDING THE SEQUENCE LISTING

[0002] The Sequence Listing associated with this application is provided in ST.26 (XML) format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The XML copy named L103438\_1230US\_Seq\_List.xml is 1,159,840 bytes in size, was created on Jun. 28, 2024, and is being submitted electronically via EFS-Web.

### FIELD OF THE INVENTION

[0003] The present invention relates to the field of molecular biology and gene editing.

### BACKGROUND OF THE INVENTION

[0004] Targeted genome editing or modification is rapidly becoming an important tool for basic and applied research. Initial methods involved engineering nucleases such as meganucleases, zinc finger fusion proteins or TALENs, requiring the generation of chimeric nucleases with engineered, programmable, sequence-specific DNA-binding domains specific for each particular target sequence. RNA-guided nucleases (RGNs), such as the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-associated (Cas) proteins of the CRISPR-Cas bacterial system, allow for the targeting of specific sequences by complexing the nucleases with guide RNA that specifically hybridizes with a particular target sequence. Producing target-specific guide RNAs is less costly and more efficient than generating chimeric nucleases for each target sequence. Such RNA-guided nucleases can be used to edit genomes through the introduction of a sequence-specific, double-stranded break that is repaired via error-prone non-homologous end-joining (NHEJ) to introduce a mutation at a specific genomic location.

[0005] Additionally, RGNs are useful for targeted DNA editing approaches. Targeted editing of nucleic acid sequences, for example targeted cleavage, to allow for introduction of a specific modification into genomic DNA, enables a highly nuanced approach to studying gene function and gene expression. RGNs may also be used to generate chimeric proteins which use the RNA-guided activity of the RGN in combination with a DNA modifying enzyme, such as a deaminase, for targeted base editing. Targeted editing may be deployed for targeting genetic diseases in humans or for introducing agronomically ben-

eficial mutations in the genomes of crop plants. The development of genome editing tools provides new approaches to gene editing-based mammalian therapeutics and agrobiotechnology.

### BRIEF SUMMARY OF THE INVENTION

[0006] Compositions and methods for modifying a target DNA molecule are provided. The compositions find use in modifying a target DNA molecule of interest. Compositions provided comprise deaminase polypeptides. Also provided are fusion proteins comprising a nucleic acid molecule-binding polypeptide (e.g., DNA-binding polypeptide) and a deaminase polypeptide, and ribonucleoprotein complexes comprising a fusion protein comprising an RNA-guided nuclease and a deaminase polypeptide and ribonucleic acids. Compositions provided also include nucleic acid molecules encoding the deaminase polypeptides or the fusion proteins, and vectors and host cells comprising the nucleic acid molecules. The methods disclosed herein are drawn to binding a target sequence of interest within a target DNA molecule of interest and modifying the target DNA molecule of interest.

### DETAILED DESCRIPTION

[0007] Many modifications and other embodiments of the inventions set forth herein will come to mind to one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions. Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.

#### I. Overview

[0008] This disclosure provides novel adenine deaminases and fusion proteins that comprise a nucleic acid molecule-binding polypeptide, such as a DNA-binding polypeptide, and a novel deaminase polypeptide. In certain embodiments, the DNA-binding polypeptide is a sequence-specific DNA-binding polypeptide, in that the DNA-binding polypeptide binds to a target sequence at a greater frequency than binding to a randomized background sequence. In some embodiments, the DNA-binding polypeptide is or is derived from a meganuclease, zinc finger fusion protein, or TALEN. In some embodiments, the fusion protein comprises an RNA-guided DNA-binding polypeptide and a deaminase polypeptide. In some embodiments, the RNA-guided DNA-binding polypeptide is an RNA-guided nuclease, such as a Cas9 polypeptide domain that binds to a guide RNA (also referred to as gRNA), which, in turn, binds a target nucleic acid sequence via strand hybridization.

[0009] The deaminase polypeptides disclosed herein can deaminate a nucleobase, such as, for example, adenine. The deamination of a nucleobase by a deaminase can lead to a point mutation at the respective residue, which is referred to herein as "nucleic acid editing", or "base editing". Fusion proteins comprising an RNA-guided nuclease (RGN) polypeptide and a deaminase can thus be used for the targeted editing of nucleic acid sequences.

[0010] Such fusion proteins are useful for targeted editing of DNA in vitro, e.g., for the generation of genetically modified cells. These genetically modified cells may be plant cells or animal cells. Such fusion proteins may also be useful for the introduction of targeted mutations, e.g., for the correction of genetic defects in mammalian cells ex vivo, e.g., in cells obtained from a subject that are subsequently re-introduced into the same or another subject; and for the introduction of targeted mutations, e.g., the correction of genetic defects or the introduction of deactivating mutations in disease-associated genes in a mammalian subject. Such fusion proteins may also be useful for the introduction of targeted mutations in plant cells, e.g., for the introduction of beneficial or agronomically valuable traits or alleles.

[0011] The terms "protein," "peptide," and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, etc. A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof.

[0012] Any of the proteins provided herein may be produced by any method known in the art. For example, the proteins provided herein may be produced via recombinant protein expression and purification, which is especially suited for fusion proteins comprising a peptide linker. Methods for recombinant protein expression and purification are well known, and include those described by Green and Sambrook, *Molecular Cloning: A Laboratory Manual* (4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2012)), the entire contents of which are incorporated herein by reference.

## II. Deaminases

[0013] The term "deaminase" refers to an enzyme that catalyzes a deamination reaction. The deaminases of the invention are nucleobase deaminases and the terms "deaminase" and "nucleobase deaminase" are used interchangeably herein. The deaminase may be a naturally-occurring deaminase enzyme or an active fragment or variant thereof. A deaminase may be active on single-stranded nucleic acids, such as ssDNA or ssRNA, or on double-stranded nucleic acids, such as dsDNA or dsRNA. In some embodiments, the deaminase is only capable of deaminating ssDNA and does not act on dsDNA.

[0014] The presently disclosed methods and compositions comprise an adenine deaminase. In some embodiments, the deaminase is an ADAT family deaminase or a variant thereof. Deamination of adenine, adenosine, or deoxyadenosine yields inosine, which is treated as guanine by polymerases. To date there are no known naturally occurring adenine deaminases that deaminate adenine in DNA. Sev-

eral methods have been employed to evolve and optimize adenine deaminase acting on tRNA (ADAT) proteins to be active on DNA molecules in mammalian cells (Gaudelli et al, 2017; Koblan, L. W. et al, 2018, *Nat Biotechnol* 36, 843-846; Richter, M. F. et al, 2020, *Nat Biotechnol*, doi: 10.1038/s41587-020-0562-8, each of which are incorporated by reference in their entirety herein). One such method uses a bacterial selection assay where only cells with the ability to activate antibiotic resistance through A: T>G:C conversions are able to survive.

[0015] The present invention relates to novel adenine deaminase polypeptides which were produced through evolution and optimization of bacterial deaminases. Novel adenine deaminases are presently disclosed and set forth as SEQ ID NOs: 1-10 and 399-441. The deaminases of the invention may be used for editing of DNA or RNA molecules. In some embodiments, the deaminases of the invention may be used for editing of ssDNA or ssRNA molecules. The adenine deaminases described herein are useful as deaminases alone or as components in fusion proteins. A fusion protein comprising a DNA-targeting polypeptide and an adenine deaminase polypeptide is referred to herein as an "A-based editor", "adenine base editor", or an "ABE" and can be used for the targeted editing of nucleic acid sequences.

[0016] "Base editors" are fusion proteins comprising a DNA-targeting polypeptide, such as an RGN, and a deaminase. Adenine base editors (ABEs) comprise a DNA-targeting protein, such as an RGN, and an adenine deaminase. ABEs function through the deamination of adenine into inosine on a DNA target molecule (Gaudelli, N. M. et al. 2017). Inosine is recognized as a guanine by polymerases and allows for the incorporation of a cytosine on the complementary DNA strand across from the inosine. After a round of replication post-deamination, there is a resulting A: T to G:C base pair change in the genome. In some embodiments, the presently disclosed adenine deaminases or active variants or fragments thereof introduce A>N mutations in a DNA molecule, wherein N is C, G, or T. In further embodiments, they introduce A>G mutations in a DNA molecule.

[0017] In those embodiments wherein the deaminase has been targeted to a specific region of a nucleic acid molecule via fusion with a DNA-binding polypeptide, the mutation rate of adenines within or adjacent to the target sequence to which the DNA-binding polypeptide binds can be measured using any method known in the art, including polymerase chain reaction (PCR), restriction fragment length polymorphism (RFLP), or DNA sequencing.

[0018] The presently disclosed novel deaminases or active variants or fragments thereof that retain deaminase activity may be introduced into the cell as part of a deaminase-DNA-binding polypeptide fusion, and/or may be co-expressed with a DNA-binding polypeptide-deaminase fusion, to increase the efficiency of introducing the desired A>G mutation in a target DNA molecule. The presently disclosed deaminases have the amino acid sequence of any of SEQ ID NOs: 1-10 and 399-441 or a variant or fragment thereof retaining deaminase activity. In some embodiments, the deaminase has an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least

97%, at least 98%, or at least 99% identity to the amino acid sequence of any of SEQ ID NOS: 1-10 and 399-441. In particular embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. For example, the deaminase comprises an amino acid sequence having at least about 80% identity, at least about 90% identity, at least about 95% identity, at least about 96% identity, at least about 97% identity, at least about 98% identity, at least about 99% identity, at least about 99.5% identity, or at least about 99.9% identity to SEQ ID NO: 407. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% identity, at least 90% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, at least 99.5% identity, or at least 99.9% identity to SEQ ID NO: 407. In some embodiments, the deaminase comprises the amino acid sequence of SEQ ID NO: 407. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 399. For example, the deaminase comprises an amino acid sequence having at least about 80% identity, at least about 90% identity, at least about 95% identity, at least about 96% identity, at least about 97% identity, at least about 98% identity, at least about 99% identity, at least about 99.5% identity, or at least about 99.9% identity to SEQ ID NO: 399. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% identity, at least 90% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, at least 99.5% identity, or at least 99.9% identity to SEQ ID NO: 399. In some embodiments, the deaminase comprises the amino acid sequence of SEQ ID NO: 399. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405. For example, the deaminase comprises an amino acid sequence having at least about 80% identity, at least about 90% identity, at least about 95% identity, at least about 96% identity, at least about 97% identity, at least about 98% identity, at least about 99% identity, at least about 99.5% identity, or at least about 99.9% identity to SEQ ID NO: 405. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% identity, at least 90% identity, at least 95% identity, at least 96% identity, at least 97% identity, at least 98% identity, at least 99% identity, at least 99.5% identity, or at least 99.9% identity to SEQ ID NO: 405. In some embodiments, the deaminase comprises the amino acid sequence of SEQ ID NO: 405.

### III. Nucleic Acid Molecule-Binding Polypeptides

**[0019]** Some aspects of this disclosure provide fusion proteins that comprise a nucleic acid molecule-binding polypeptide and a deaminase polypeptide. While binding to and targeted editing of RNA molecules is contemplated by the present invention, in some embodiments, the nucleic acid molecule-binding polypeptide of the fusion protein is a DNA-binding polypeptide. Such fusion proteins are useful for targeted editing of DNA in vitro, ex vivo, or in vivo. These novel fusion proteins are active in mammalian cells and are useful for targeted editing of DNA molecules.

[0020] The term “fusion protein” as used herein refers to a hybrid polypeptide which comprises protein domains from

at least two different proteins. A fusion protein may comprise more than one different domain, for example, a DNA-binding domain and a deaminase. In some embodiments, a fusion protein is in a complex with, or is in association with, a nucleic acid, e.g., RNA.

**[0021]** In some embodiments, the presently disclosed fusion proteins comprise a DNA-binding polypeptide. As used herein, the term "DNA-binding polypeptide" refers to any polypeptide which is capable of binding to DNA. In certain embodiments, the DNA-binding polypeptide portion of the presently disclosed fusion proteins binds to double-stranded DNA. In particular embodiments, the DNA-binding polypeptide binds to DNA in a sequence-specific manner. As used herein, the terms "sequence-specific" or "sequence-specific manner" refer to the selective interaction with a specific nucleotide sequence.

**[0022]** Two polynucleotide sequences can be considered to be substantially complementary when the two sequences hybridize to each other under stringent conditions. Likewise, a DNA-binding polypeptide is considered to bind to a particular target sequence in a sequence-specific manner if the DNA-binding polypeptide binds to its sequence under stringent conditions. By "stringent conditions" or "stringent hybridization conditions" is intended conditions under which the two polynucleotide sequences (or the polypeptide binds to its specific target sequence) will bind to each other to a detectably greater degree than to other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. Typically, stringent conditions will be those in which the salt concentration is less than 1.5 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3, and the temperature is at least 30° C. for short sequences (e.g., 10 to 50 nucleotides) and at least 60° C. for long sequences (e.g., greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringency conditions include hybridization with a buffer solution of 30 to 35% formamide, 1 M NaCl, 1% SDS (sodium dodecyl sulfate) at 37° C., and a wash in 1 $\times$  to 2 $\times$ SSC (20 $\times$ SSC=3.0 M NaCl/0.3 M trisodium citrate) at 50 to 55° C. Exemplary moderate stringency conditions include hybridization in 40 to 45% formamide, 1.0 M NaCl, 1% SDS at 37° C., and a wash in 0.5 $\times$  to 1 $\times$ SSC at 55 to 60° C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37° C., and a wash in 0.1 $\times$ SSC at 60 to 65° C. Optionally, wash buffers may comprise about 0.1% to about 1% SDS. Duration of hybridization is generally less than about 24 hours, usually about 4 to about 12 hours. The duration of the wash time will be at least a length of time sufficient to reach equilibrium.

**[0023]** The Tm is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched sequence. For DNA-DNA hybrids, the Tm can be approximated from the equation of Meinkoth and Wahl (1984) Anal. Biochem. 138:267-284:  $T_m = 81.5^\circ C + 16.6 (\log M) + 0.41 (\% GC) - 0.61 (\% form) - 500/L$ ; where M is the molarity of monovalent cations, % GC is the percentage of guanosine and cytosine nucleotides in the DNA, % form is the percentage of formamide in the hybridization solution, and L is the length of the hybrid in base pairs. Generally, stringent conditions are selected to be about  $5^\circ C$  lower than the thermal melting point (Tm) for the specific sequence and its complement at

a defined ionic strength and pH. However, severely stringent conditions can utilize a hybridization and/or wash at 1, 2, 3, or 4° C. lower than the thermal melting point (Tm); moderately stringent conditions can utilize a hybridization and/or wash at 6, 7, 8, 9, or 10° C. lower than the thermal melting point (Tm); low stringency conditions can utilize a hybridization and/or wash at 11, 12, 13, 14, 15, or 20° C. lower than the thermal melting point (Tm). Using the equation, hybridization and wash compositions, and desired Tm, those of ordinary skill will understand that variations in the stringency of hybridization and/or wash solutions are inherently described. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes, Part I, Chapter 2* (Elsevier, New York); and Ausubel et al., eds. (1995) *Current Protocols in Molecular Biology, Chapter 2* (Greene Publishing and Wiley-Interscience, New York). See Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual* (2d ed., Cold Spring Harbor Laboratory Press, Plainview, New York).

[0024] In certain embodiments, the sequence-specific DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide (RGDBP). As used herein, the terms “RNA-guided, DNA-binding polypeptide” and “RGDBP” refer to polypeptides capable of binding to DNA through the hybridization of an associated RNA molecule with the target DNA sequence.

[0025] In some embodiments, the DNA-binding polypeptide of the fusion protein is a nuclease, such as a sequence-specific nuclease. As used herein, the term “nuclease” refers to an enzyme that catalyzes the cleavage of phosphodiester bonds between nucleotides in a nucleic acid molecule. In some embodiments, the DNA-binding polypeptide is an endonuclease, which is capable of cleaving phosphodiester bonds between nucleotides within a nucleic acid molecule, whereas in certain embodiments, the DNA-binding polypeptide is an exonuclease that is capable of cleaving the nucleotides at either end (5' or 3') of a nucleic acid molecule. In some embodiments, the sequence-specific nuclease is selected from the group consisting of a meganuclease, a zinc finger nuclease, a TAL-effector DNA binding domain-nuclease fusion protein (TALEN), and an RNA-guided nuclease (RGN) or variants thereof wherein the nuclease activity has been reduced or inhibited.

[0026] As used herein, the term “meganuclease” or “homing endonuclease” refers to endonucleases that bind a recognition site within double-stranded DNA that is 12 to 40 bp in length. Non-limiting examples of meganucleases are those that belong to the LAGLIDADG family that comprise the conserved amino acid motif LAGLIDADG (SEQ ID NO: 49). The term “meganuclease” can refer to a dimeric or single-chain meganuclease.

[0027] As used herein, the term “zinc finger nuclease” or “ZFN” refers to a chimeric protein comprising a zinc finger DNA-binding domain and a nuclease domain.

[0028] As used herein, the term “TAL-effector DNA binding domain-nuclease fusion protein” or “TALEN” refers to a chimeric protein comprising a TAL effector DNA-binding domain and a nuclease domain.

[0029] As used herein, the term “RNA-guided nuclease” or “RGN” refers to an RNA-guided, DNA-binding polypeptide that has nuclease activity. RGNs are considered “RNA-guided” because guide RNAs form a complex with the RNA-guided nucleases to direct the RNA-guided nuclease

to bind to a target sequence and in some embodiments, introduce a single-stranded or double-stranded break at the target sequence. The RGN may be a CasX, a CasY, a C2cl, a C2c2, a C2c3, a GeoCas9, aSpCas9, a SaCas9, a Nme2Cas9, aCjCas9, aCasl2a (formerly known as Cpf1), a Cas12b, a Cas12g, a Cas12h, a Cas12i, aLbCas12a, a AsCas12a, a CasMINI, a Cas13b, a Cas13c, a Cas13d, a Cas14, aCsn2, anxCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, or a Spy-macCas9, a Spy-macCas9 domain, or a RGN with an amino acid sequence set forth in any one of SEQ ID NOs: 41, 60, 366, or 368. In some embodiments, as described below, the RGNs provided herein are RGN nickases.

[0030] According to the present invention, an RGN protein that has been mutated to become nuclease-inactive or “dead”, such as for example dCas9, can be referred to as an RNA-guided, DNA-binding polypeptide or a nuclease-inactive RGN or nuclease-dead RGN. Additionally, suitable nuclease-inactive Cas9 domains of other known RNA guided nucleases (RGNs) can be determined (for example, a nuclease-inactive variant of the RGN APG08290.1 disclosed in U.S. Patent Publication No. 2019/0367949, the entire contents of which are incorporated herein by reference herein).

[0031] In some embodiments, the fusion protein comprises an RGN fused to a deaminase described herein. In those embodiments of fusion proteins described above, the deaminase is selected from deaminases comprising an amino acid sequence having at least 80% sequence identity to any one of SEQ ID NOs: 1-10 and 399-441. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 399. In some embodiments, the deaminase comprises an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405. In those embodiments of fusion proteins described above, the RGN is selected from a CasX, a CasY, a C2cl, a C2c2, a C2c3, a GeoCas9, aSpCas9, a SaCas9, a Nme2Cas9, aCjCas9, aCasl2a (formerly known as Cpf1), a Cas12b, a Cas12g, a Cas12h, a Cas12i, aLbCas12a, a AsCas12a, a CasMINI, a Cas13b, a Cas13c, a Cas13d, a Cas14, aCsn2, anxCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or an RGN with an amino acid sequence set forth in any one of SEQ ID NOs: 41, 60, 366, or 368. In particular embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. In some embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 399. In particular embodiments, the fusion protein comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405. The Cas9 nickase, can be any Cas9 nickase disclosed in PCT Patent Publication No. WO2020181195, the entire contents of which is incorporated herein by reference herein.

[0032] The term “RGN polypeptide” encompasses RGN polypeptides that only cleave a single strand of a target nucleotide sequence, which is referred to herein as a nickase.

Such RGNs have a single functioning nuclease domain. RGN nickases can be naturally-occurring nickases or can be RGN proteins that naturally cleave both strands of a double-stranded nucleic acid molecule that have been mutated within one or more nuclease domains such that the nuclease activity of these mutated domains is reduced or eliminated, to become a nickase. In some embodiments, the nickase RGN of the fusion protein comprises a mutation (e.g., a D10A mutation) which renders the RGN capable of cleaving only the non-base edited, target strand (the strand which comprises the PAM and is base paired to a gRNA) of a nucleic acid duplex. This D10A mutation mutates the first aspartic acid residue in the split RuvC nuclease domain of the RGN. The present application discloses several D10A nickase variants or homologous nickase variants of described RGNs (see Example 4). nAPG07433.1 and nAPG08290.1 (set forth as SEQ ID NOS: 42 and 61, respectively) are nickase variants of APG07433.1 and APG08290.1, which are set forth as SEQ ID NO: 41 and 60, respectively, and are described in WO 2019/236566 (incorporated by reference in its entirety herein). nAPG00969 (set forth as SEQ ID NO: 52) and nAPG09748 (set forth as SEQ ID NO: 54) are nickase variants of APG00969 and APG09748, respectively, which are described in WO 2020/139783 (incorporated by reference in its entirety herein). nAPG06646 (set forth as SEQ ID NO: 53) and nAPG09882 (set forth as SEQ ID NO: 55) are nickase variants of APG06646 and APG09882, respectively, which are described in PCT publication WO 2021/030344 (incorporated by reference in its entirety herein). nAPG03850, nAPG07553, nAPG055886, and nAPG01604 are set forth as SEQ ID NOS: 56-59, respectively, and are nickase variants of APG03850, APG07553, APG055886, and APG01604 which are described in the pending PCT Application No. PCT/US2021/028843 (incorporated by reference in its entirety herein). Various RGN nickases, their variants and their sequences are disclosed in PCT Patent Publication No. WO2020181195, the entire contents of which are incorporated herein by reference herein. One exemplary suitable nuclease-inactive Cas9 is the D10A/H840A Cas9 mutant (see, e.g., Qi et al., *Cell.* 2013; 152 (5): 1173-83, the entire contents of which are incorporated herein by reference).

**[0033]** In some embodiments, the nickase RGN of the fusion protein comprises a mutation (e.g., a H840A mutation), which renders the RGN capable of cleaving only the base-edited, non-targeted strand (the strand which does not comprise the PAM and is not base paired to a gRNA) of a nucleic acid duplex. The H840A mutation mutates the first histidine of the HNH nuclease domain. A nickase RGN comprising an H840A mutation, or an equivalent mutation, has an inactivated HNH domain. A nickase RGN with an H840A mutation cleaves the non-targeted strand. A nickase comprising a D10A mutation, or an equivalent mutation, has an inactivated RuvC nuclease domain and cleaves the targeted strand. D10A nickases are not able to cleave the non-targeted strand of the DNA, i.e., the strand where base editing is desired.

**[0034]** Other additional exemplary suitable nuclease inactive Cas9 domains include, but are not limited to, D10A/D839A/H840A, and D10A/D839A/H840A/N863A mutant domains (See, e.g., Mali et al., *Nature Biotechnology*. 2013; 31 (9): 833-838, the entire contents of which are incorporated herein by reference). Additional suitable RGN proteins mutated to be nickases will be apparent to those of skill in

the art based on this disclosure and knowledge in the field (such as for example the RGNs disclosed in PCT Publication Nos. WO 2019/236566, WO2020181195, which are herein incorporated by reference in their entirety) and are within the scope of this disclosure. In preferred embodiments, an RGN which has nickase activity on the target strand nicks the target strand, while the complementary, non-target strand is modified by the deaminase. Cellular DNA-repair machinery may repair the nicked, target strand using the modified non-target strand as a template, thereby introducing a mutation in the DNA.

**[0035]** In some embodiments the RGN nickase retaining nickase activity comprises an amino acid sequence that has at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% identity to SEQ ID NO: 42 or any one of SEQ ID NOS: 52-59, 61, 397, and 398.

**[0036]** Any method known in the art for introducing mutations into an amino acid sequence, such as PCR-mediated mutagenesis and site-directed mutagenesis, can be used for generating nickases or nuclease-dead RGNs. See, e.g., U.S. Publ. No. 2014/0068797 and U.S. Pat. No. 9,790,490; each of which is incorporated herein by reference in its entirety. RNA-guided nucleases (RGNs) allow for the targeted manipulation of a single site within a genome and are useful in the context of gene targeting for therapeutic and research applications. In a variety of organisms, including mammals, RNA-guided nucleases have been used for genome engineering by stimulating either non-homologous end joining or homologous recombination. RGNs include CRISPR-Cas proteins, which are RNA-guided nucleases directed to the target sequence by a guide RNA (gRNA) as part of a Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) RNA-guided nuclease system, or active variants or fragments thereof.

**[0037]** Further provided herein are RGN polypeptides (and nucleic acid molecules encoding RGN polypeptides) that comprise the amino acid sequence set forth as SEQ ID NO: 41 or 60, but lacking amino acid residues 590 to 597 of SEQ ID NO: 41 or 60, or an active variant or fragment thereof. In certain embodiments, the RGN polypeptide comprises the amino acid sequence set forth as SEQ ID NO: 366, 368, 397, or 398 or an active variant or fragment thereof.

**[0038]** Some aspects of this disclosure provide fusion proteins that comprise an RNA-guided DNA-binding polypeptide and a deaminase polypeptide, specifically an adenine deaminase polypeptide. In some embodiments, the RNA-guided DNA-binding polypeptide is an RNA-guided nuclease. In further embodiments, the RNA-guided nuclease is a naturally-occurring CRISPR-Cas protein or an active variant or fragment thereof. CRISPR-Cas systems are classified into Class 1 or Class 2 systems. Class 2 systems comprise a single effector nuclease and include Types II, V, and VI. The Class 1 and 2 systems are subdivided into types (Types I, II, III, IV, V, VI), with some types further divided into subtypes (e.g., Type II-A, Type II-B, Type II-C, Type V-A, Type V-B).

**[0039]** In certain embodiments, the CRISPR-Cas protein is a naturally-occurring Type II CRISPR-Cas protein or an active variant or fragment thereof. As used herein, the term “Type II CRISPR-Cas protein,” “Type II CRISPR-Cas effector protein,” or “Cas9” refers to a CRISPR-Cas effector protein that requires a trans-activating RNA (tracrRNA) and

comprises two nuclease domains (i.e., RuvC and HNH), each of which is responsible for cleaving a single strand of a double-stranded DNA molecule. In some embodiments, the present invention provides a fusion protein comprising a presently disclosed deaminase fused to *Streptococcus pyogenes* Cas9 (SpCas9) or a SpCas9 nickase, the sequences of which are set forth as SEQ ID NOs: 555 and 556, respectively, and are described in U.S. Pat. Nos. 10,000,772 and 8,697,359, each of which is herein incorporated by reference in its entirety. In some embodiments, the present invention provides a fusion protein comprising a presently disclosed deaminase fused to *Streptococcus thermophilus* Cas9 (StCas9) or a StCas9 nickase, the sequences of which are set forth as SEQ ID NOs: 557 and 558, respectively, and are disclosed in U.S. Pat. No. 10,113,167, which is herein incorporated by reference in its entirety. In some embodiments, the present invention provides a fusion protein comprising a presently disclosed deaminase fused to *Streptococcus aureus* Cas9 (SaCas9) or a SaCas9 nickase, the sequences of which are set forth as SEQ ID NOs: 559 and 560, respectively, and are disclosed in U.S. Pat. No. 9,752,132, which is herein incorporated by reference in its entirety.

[0040] In some embodiments, the CRISPR-Cas protein is a naturally-occurring Type V

[0041] CRISPR-Cas protein or an active variant or fragment thereof. As used herein, the term “Type V CRISPR-Cas protein,” “Type V CRISPR-Cas effector protein,” or “Cas12” refers to a CRISPR-Cas effector protein that cleaves dsDNA and comprises a single RuvC nuclease domain or a split-RuvC nuclease domain and lacks an HNH domain (Zetsche et al 2015, *Cell* doi: 10.1016/j.cell.2015.09.038; Shmakov et al 2017, *Nat Rev Microbiol* doi: 10.1038/nrmicro.2016.184; Yan et al 2018, *Science* doi: 10.1126/science.aav7271; Harrington et al 2018, *Science* doi: 10.1126/science.aav4294). It is to be noted that Cas12a is also referred to as Cpf1, and does not require a tracrRNA, although other Type V CRISPR-Cas proteins, such as Cas12b, do require a tracrRNA. Most Type V effectors can also target ssDNA (single-stranded DNA), often without a PAM requirement (Zetsche et al 2015; Yan et al 2018; Harrington et al 2018). The term “Type V CRISPR-Cas protein” encompasses the unique RGNs comprising split RuvC nuclease domains, such as those disclosed in U.S. Provisional Appl. Nos. 62/955,014 filed Dec. 30, 2019 and 63/058,169 filed Jul. 29, 2020, and PCT International Appl. No. PCT/US2020/067138 filed Dec. 28, 2020, the contents of each of which are incorporated herein by reference in its entirety. In some embodiments, the present invention provides a fusion protein comprising a presently disclosed deaminase fused to *Francisella novicida* Cas12a (FnCas12a), the sequence of which is set forth as SEQ ID NOs: 561 and is disclosed in U.S. Pat. No. 9,790,490, which is herein incorporated by reference in its entirety, or any of the nuclease-inactivating mutants of FnCas12a disclosed within U.S. Pat. No. 9,790,490.

[0042] In some embodiments, the CRISPR-Cas protein is a naturally-occurring Type VI CRISPR-Cas protein or an active variant or fragment thereof. As used herein, the term “Type VI CRISPR-Cas protein,” “Type VI CRISPR-Cas effector protein,” or “Cas13” refers to a CRISPR-Cas effector protein that does not require a tracrRNA and comprises two HEPN domains that cleave RNA.

[0043] The term “guide RNA” refers to a nucleotide sequence having sufficient complementarity with a target

nucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of an associated RGN to the target nucleotide sequence. For CRISPR-Cas RGNs, the respective guide RNA is one or more RNA molecules (generally, one or two), that can bind to the RGN and guide the RGN to bind to a particular target nucleotide sequence, and in those instances wherein the RGN has nickase or nuclease activity, also cleave the target nucleotide sequence. A guide RNA comprises a CRISPR RNA (crRNA) and in some embodiments, a trans-activating CRISPR RNA (tracrRNA).

[0044] A CRISPR RNA comprises a spacer sequence and a CRISPR repeat sequence. The “spacer sequence” is the nucleotide sequence that directly hybridizes with the target nucleotide sequence of interest. The spacer sequence is engineered to be fully or partially complementary with the target sequence of interest. In various embodiments, the spacer sequence comprises from about 8 nucleotides to about 30 nucleotides, or more. For example, the spacer sequence can be about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In some embodiments, the spacer sequence is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In some embodiments, the spacer sequence is about 10 to about 26 nucleotides in length, or about 12 to about 30 nucleotides in length. In some embodiments, the spacer sequence is 10 to 26 nucleotides in length, or 12 to 30 nucleotides in length. In particular embodiments, the spacer sequence is about 30 nucleotides in length. In particular embodiments, the spacer sequence is 30 nucleotides in length. In some embodiments, the degree of complementarity between a spacer sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is between 50% and 99% or more, including but not limited to about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a spacer sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more. In particular embodiments, the spacer sequence is free of secondary structure, which can be predicted using any suitable polynucleotide folding algorithm known in the art, including but not limited to mFold (see, e.g., Zuker and Stiegler (1981) *Nucleic Acids Res.* 9:133-148) and RNAfold (see, e.g., Gruber et al. (2008) *Cell* 106 (1): 23-24).

[0045] The CRISPR RNA repeat sequence comprises a nucleotide sequence that forms a structure, either on its own or in concert with a hybridized tracrRNA, that is recognized by the RGN molecule. In various embodiments, the CRISPR RNA repeat sequence comprises from about 8 nucleotides to about 30 nucleotides, or more. In particular embodiments, the CRISPR RNA repeat sequence comprises from 8 nucleotides to 30 nucleotides, or more. For example, the CRISPR repeat sequence can be about 8, about 9, about 10, about 11,

about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In particular embodiments, the CRISPR repeat sequence is 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In some embodiments, the degree of complementarity between a CRISPR repeat sequence and its corresponding tracrRNA sequence, when optimally aligned using a suitable alignment algorithm, is between 50% and 99%, or more, including but not limited to about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a CRISPR repeat sequence and its corresponding tracrRNA sequence, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more.

[0046] In some embodiments, the guide RNA further comprises a tracrRNA molecule. A trans-activating CRISPR RNA or tracrRNA molecule comprises a nucleotide sequence comprising a region that has sufficient complementarity to hybridize to a CRISPR repeat sequence of a crRNA, which is referred to herein as the anti-repeat region. In some embodiments, the tracrRNA molecule further comprises a region with secondary structure (e.g., stem-loop) or forms secondary structure upon hybridizing with its corresponding crRNA. In particular embodiments, the region of the tracrRNA that is fully or partially complementary to a CRISPR repeat sequence is at the 5' end of the molecule and the 3' end of the tracrRNA comprises secondary structure. This region of secondary structure generally comprises several hairpin structures, including the nexus hairpin, which is found adjacent to the anti-repeat sequence. There are often terminal hairpins at the 3' end of the tracrRNA that can vary in structure and number, but often comprise a GC-rich Rho-independent transcriptional terminator hairpin followed by a string of Us at the 3' end. See, for example, Briner et al. (2014) *Molecular Cell* 56:333-339, Briner and Barrangou (2016) *Cold Spring Harb Protoc*; doi: 10.1101/pdb.top090902, and U.S. Publication No. 2017/0275648, each of which is herein incorporated by reference in its entirety.

[0047] In various embodiments, the anti-repeat region of the tracrRNA that is fully or partially complementary to the CRISPR repeat sequence comprises from about 6 nucleotides to about 30 nucleotides, or more. For example, the region of base pairing between the tracrRNA anti-repeat sequence and the CRISPR repeat sequence can be about 6, about 7, about 8, about 9, about 10, about 11, about 12, about 13, about 14, about 15, about 16, about 17, about 18, about 19, about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, or more nucleotides in length. In particular embodiments, the region of base pairing between the tracrRNA anti-repeat sequence and the CRISPR repeat sequence is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, or more nucleotides in length. In particular embodiments, the anti-repeat region of the tracrRNA that is fully or

partially complementary to a CRISPR repeat sequence is about 10 nucleotides in length. In particular embodiments, the anti-repeat region of the tracrRNA that is fully or partially complementary to a CRISPR repeat sequence is 10 nucleotides in length. In some embodiments, the degree of complementarity between a CRISPR repeat sequence and its corresponding tracrRNA anti-repeat sequence, when optimally aligned using a suitable alignment algorithm, is between 50% and 99% or more, including but not limited to about or more than about 50%, about 60%, about 70%, about 75%, about 80%, about 81%, about 82%, about 83%, about 84%, about 85%, about 86%, about 87%, about 88%, about 89%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more. In particular embodiments, the degree of complementarity between a CRISPR repeat sequence and its corresponding tracrRNA anti-repeat sequence, when optimally aligned using a suitable alignment algorithm, is 50%, 60%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more.

[0048] In various embodiments, the entire tracrRNA comprises from about 60 nucleotides to more than about 210 nucleotides. In particular embodiments, the entire tracrRNA comprises from 60 nucleotides to more than 210 nucleotides. For example, the tracrRNA can be about 60, about 65, about 70, about 75, about 80, about 85, about 90, about 95, about 100, about 105, about 110, about 115, about 120, about 125, about 130, about 135, about 140, about 150, about 160, about 170, about 180, about 190, about 200, about 210 or more nucleotides in length. In particular embodiments, the tracrRNA is 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 150, 160, 170, 180, 190, 200, 210 or more nucleotides in length. In particular embodiments, the tracrRNA is about 100 to about 210 nucleotides in length, including about 95, about 96, about 97, about 98, about 99, about 100, about 105, about 106, about 107, about 108, about 109, and about 100 nucleotides in length. In particular embodiments, the tracrRNA is 100 to 110 nucleotides in length, including 95, 96, 97, 98, 99, 100, 105, 106, 107, 108, 109, and 110 nucleotides in length.

[0049] Guide RNAs form a complex with an RNA-guided, DNA-binding polypeptide or an RNA-guided nuclease to direct the RNA-guided nuclease to bind to a target sequence. If the guide RNA complexes with an RGN, the bound RGN introduces a single-stranded or double-stranded break at the target sequence. After the target sequence has been cleaved, the break can be repaired such that the DNA sequence of the target sequence is modified during the repair process. Provided herein are methods for using mutant variants of RNA-guided nucleases, which are either nuclease inactive or nickases, which are linked to deaminases to modify a target sequence in the DNA of host cells. The mutant variants of RNA-guided nucleases in which the nuclease activity is inactivated or significantly reduced may be referred to as RNA-guided, DNA-binding polypeptides, as the polypeptides are capable of binding to, but not necessarily cleaving, a target sequence. RNA-guided nucleases only capable of cleaving a single strand of a double-stranded nucleic acid molecule are referred to herein as nickases.

[0050] A target nucleotide sequence is bound by an RNA-guided, DNA-binding polypeptide and hybridizes with the guide RNA associated with the RGDBP. The target sequence

can then be subsequently cleaved if the RGDBP possesses nuclease activity (i.e., is an RGN), which encompasses activity as a nickase.

[0051] The guide RNA can be a single guide RNA or a dual-guide RNA system. A single guide RNA comprises the crRNA and optionally tracrRNA on a single molecule of RNA, whereas a dual-guide RNA system comprises a crRNA and a tracrRNA present on two distinct RNA molecules, hybridized to one another through at least a portion of the CRISPR repeat sequence of the crRNA and at least a portion of the tracrRNA, which may be fully or partially complementary to the CRISPR repeat sequence of the crRNA. In some of those embodiments wherein the guide RNA is a single guide RNA, the crRNA and optionally tracrRNA are separated by a linker nucleotide sequence.

[0052] In general, the linker nucleotide sequence is one that does not include complementary bases in order to avoid the formation of secondary structure within or comprising nucleotides of the linker nucleotide sequence. In some embodiments, the linker nucleotide sequence between the crRNA and tracrRNA is at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, or more nucleotides in length. In particular embodiments, the linker nucleotide sequence between the crRNA and tracrRNA is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more nucleotides in length. In particular embodiments, the linker nucleotide sequence of a single guide RNA is at least 4 nucleotides in length. In particular embodiments, the linker nucleotide sequence of a single guide RNA is 4 nucleotides in length.

[0053] In certain embodiments, the guide RNA can be introduced into a target cell, organelle, or embryo as an RNA molecule. The guide RNA can be transcribed in vitro or chemically synthesized. In some embodiments, a nucleotide sequence encoding the guide RNA is introduced into the cell, organelle, or embryo. In some embodiments, the nucleotide sequence encoding the guide RNA is operably linked to a promoter (e.g., an RNA polymerase III promoter). The promoter can be a native promoter or heterologous to the guide RNA-encoding nucleotide sequence.

[0054] In various embodiments, the guide RNA can be introduced into a target cell, organelle, or embryo as a ribonucleoprotein complex, as described herein, wherein the guide RNA is bound to an RNA-guided nuclease polypeptide.

[0055] The guide RNA directs an associated RNA-guided nuclease to a particular target nucleotide sequence of interest through hybridization of the guide RNA to the target nucleotide sequence. A target nucleotide sequence can comprise DNA, RNA, or a combination of both and can be single-stranded or double-stranded. A target nucleotide sequence can be genomic DNA (i.e., chromosomal DNA), plasmid DNA, or an RNA molecule (e.g., messenger RNA, ribosomal RNA, transfer RNA, micro RNA, small interfering RNA). The target nucleotide sequence can be bound (and in some embodiments, cleaved) by an RNA-guided, DNA-binding polypeptide in vitro or in a cell. The chromosomal sequence targeted by the RGDBP can be a nuclear, plastid or mitochondrial chromosomal sequence. In some embodiments, the target nucleotide sequence is unique in the target genome.

[0056] In some embodiments, the target nucleotide sequence is adjacent to a protospacer adjacent motif (PAM). A PAM is generally within about 1 to about 10 nucleotides

from the target nucleotide sequence, including about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, or about 10 nucleotides from the target nucleotide sequence. In particular embodiments, a PAM is within 1 to 10 nucleotides from the target nucleotide sequence, including 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides from the target nucleotide sequence. The PAM can be 5' or 3' of the target sequence. In some embodiments, the PAM is 3' of the target sequence. Generally, the PAM is a consensus sequence of about 2-6 nucleotides, but in particular embodiments, is 1, 2, 3, 4, 5, 6, 7, 8, 9, or more nucleotides in length.

[0057] The PAM restricts which sequences a given RGDBP or RGN can target, as its PAM needs to be proximal to the target nucleotide sequence. Upon recognizing its corresponding PAM sequence, the RGN can cleave the target nucleotide sequence at a specific cleavage site. As used herein, a cleavage site is made up of the two particular nucleotides within a target nucleotide sequence between which the nucleotide sequence is cleaved by an RGN. The cleavage site can comprise the 1st and 2nd, 2nd and 3rd, 3rd and 4th, 4th and 5th, 5th and 6th, 7th and 8th, or 8th and 9th nucleotides from the PAM in either the 5' or 3' direction. As RGNs can cleave a target nucleotide sequence resulting in staggered ends, in some embodiments, the cleavage site is defined based on the distance of the two nucleotides from the PAM on the positive (+) strand of the polynucleotide and the distance of the two nucleotides from the PAM on the negative (-) strand of the polynucleotide.

[0058] RGDBPs and RGNs can be used to deliver a fused polypeptide, polynucleotide, or small molecule payload to a particular genomic location.

[0059] In those embodiments wherein the DNA-binding polypeptide comprises a meganuclease, a target sequence can comprise a pair of inverted, 9 basepair "half sites" which are separated by four basepairs. In the case of a single-chain meganuclease, the N-terminal domain of the protein contacts a first half-site and the C-terminal domain of the protein contacts a second half-site. Cleavage by a meganuclease produces four basepair 3' overhangs. In those embodiments wherein the DNA-binding polypeptide comprises a compact TALEN, the recognition sequence comprises a first CNNNGN sequence that is recognized by the I-TevI domain, followed by a non-specific spacer 4-16 basepairs in length, followed by a second sequence 16-22 bp in length that is recognized by the TAL-effector domain (this sequence typically has a 5' T base). In those embodiments wherein the DNA-binding polypeptide comprises a zinc finger, the DNA binding domains typically recognize an 18-bp recognition sequence comprising a pair of nine basepair "half-sites" separated by 2-10 basepairs and cleavage by the nuclease creates a blunt end or a 5' overhang of variable length (frequently four basepairs).

#### IV. Fusion Proteins

[0060] In some embodiments, a DNA-binding polypeptide (e.g., nuclease-inactive or a nickase RGN) is operably linked to a deaminase of the invention. In some embodiments, a DNA-binding polypeptide (e.g., nuclease inactive RGN or nickase RGN) fused to a deaminase of the invention can be targeted to a particular location of a nucleic acid molecule (i.e., target nucleic acid molecule), which in some embodiments is a particular genomic locus, to alter the expression of a desired sequence. In some embodiments, the binding of a fusion protein to a target sequence results in deamination

of a nucleobase, resulting in conversion from one nucleobase to another. In some embodiments, the binding of this fusion protein to a target sequence results in deamination of a nucleobase adjacent to the target sequence. The nucleobase adjacent to the target sequence that is deaminated and mutated using the presently disclosed compositions and methods may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 base pairs from the 5' or 3' end of the target sequence (bound by the gRNA) within the target nucleic acid molecule. Some aspects of this disclosure provide fusion proteins comprising (i) a DNA-binding polypeptide (e.g., a nuclease-inactive or nickase RGN polypeptide); (ii) a deaminase polypeptide; and optionally (iii) a second deaminase. The second deaminase may be the same deaminase as the first or may be a different deaminase. In some embodiments, both the first and the second deaminase are adenine deaminases of the invention.

[0061] The instant disclosure provides fusion proteins of various configurations. In some embodiments, the deaminase polypeptide is fused to the N-terminus of the DNA-binding polypeptide (e.g., RGN polypeptide). In some embodiments, the deaminase polypeptide is fused to the C-terminus of the DNA-binding polypeptide (e.g., RGN polypeptide).

[0062] In some embodiments, the deaminase and DNA-binding polypeptide (e.g., RNA-guided, DNA-binding polypeptide) are fused to each other via a peptide linker. The linker between the deaminase and DNA-binding polypeptide (e.g., RNA-guided, DNA-binding polypeptide) can determine the editing window of the fusion protein, thereby increasing deaminase specificity and reducing off-target mutations. Various linker lengths and flexibilities can be employed, ranging from very flexible linkers of the form (GGGGS)<sub>n</sub> and (G)<sub>n</sub> to more rigid linkers of the form (EAAAK)<sub>n</sub> and (XP)<sub>n</sub>, to achieve the optimal length and rigidity for deaminase activity for the specific applications. The term “linker,” as used herein, refers to a chemical group or a molecule linking two molecules or moieties, e.g., a binding domain and a cleavage domain of a nuclease. In some embodiments, a linker joins an RNA guided nuclease and a deaminase. In some embodiments, a linker joins a dead or inactive RGN and a deaminase. In further embodiments, a linker joins two deaminases. Typically, the linker is positioned between, or flanked by, two groups, molecules, or other moieties and connected to each one via a covalent bond, thus connecting the two. In some embodiments, the linker is an amino acid or a plurality of amino acids (e.g., a peptide or protein). In some embodiments, the linker is an organic molecule, group, polymer, or chemical moiety. In some embodiments, the linker is 3-100 amino acids in length, for example, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 30-35, 35-40, 40-45, 45-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, or 150-200 amino acids in length. Longer or shorter linkers are also contemplated. In some embodiments, a shorter linker is preferred to decrease the overall size or length of the fusion protein or its coding sequence.

[0063] In some embodiments, the linker comprises a (GGGGS)<sub>n</sub>, a (G)<sub>n</sub>, an (EAAAK)<sub>n</sub>, or an (XP)<sub>n</sub> motif, or a combination of any of these, wherein n is independently an integer between 1 and 30. In some embodiments, n is independently 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30,

or, if more than one linker or more than one linker motif is present, any combination thereof. Additional suitable linker motifs and linker configurations will be apparent to those of skill in the art. In some embodiments, suitable linker motifs and configurations include those described in Chen et al., 2013 (*Adv Drug Deliv Rev.* 65 (10): 1357-69, the entire contents of which are incorporated herein by reference). Additional suitable linker sequences will be apparent to those of skill in the art. In some embodiments, the linker sequence comprises the amino acid sequence set forth as SEQ ID NO: 45 or 442.

[0064] In some embodiments, the general architecture of exemplary fusion proteins provided herein comprises the structure: [NH<sub>2</sub>]-[deaminase]-[DBP]-[COOH]; [NH<sub>2</sub>]-[DBP]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[DBP]-[deaminase]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[DBP]-[deaminase]-[COOH]; or [NH<sub>2</sub>]-[deaminase]-[deaminase]-[DBP]-[COOH], wherein DBP is a DNA-binding polypeptide, NH<sub>2</sub> is the N-terminus of the fusion protein and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion protein comprises more than two deaminase polypeptides.

[0065] In certain embodiments, the general architecture of exemplary fusion proteins provided herein comprises the structure: [NH<sub>2</sub>]-[deaminase]-[RGN]-[COOH]; [NH<sub>2</sub>]-[RGN]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[RGN]-[deaminase]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[RGN]-[deaminase]-[COOH]; or [NH<sub>2</sub>]-[deaminase]-[deaminase]-[RGN]-[COOH], wherein NH<sub>2</sub> is the N-terminus of the fusion protein and COOH is the C-terminus of the fusion protein. In some embodiments, the fusion protein comprises more than two deaminase polypeptides.

[0066] In some embodiments, the fusion protein comprises the structure: [NH<sub>2</sub>]-[deaminase]-[nuclease-inactive RGN]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[deaminase]-[nuclease-inactive RGN]-[COOH]; [NH<sub>2</sub>]-[nuclease-inactive RGN]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[nuclease-inactive RGN]-[deaminase]-[COOH]; or [NH<sub>2</sub>]-[nuclease-inactive RGN]-[deaminase]-[deaminase]-[COOH]. It should be understood that “nuclease-inactive RGN” represents any RGN, including any CRISPR-Cas protein, which has been mutated to be nuclease-inactive. In some embodiments, the fusion protein comprises more than two deaminase polypeptides.

[0067] In some embodiments, the fusion protein comprises the structure: [NH<sub>2</sub>]-[deaminase]-[RGN nickase]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[deaminase]-[RGN nickase]-[COOH]; [NH<sub>2</sub>]-[RGN nickase]-[deaminase]-[COOH]; [NH<sub>2</sub>]-[deaminase]-[RGN nickase]-[deaminase]-[COOH]; or [NH<sub>2</sub>]-[RGN nickase]-[deaminase]-[deaminase]-[COOH]. It should be understood that “RGN nickase” represents any RGN, including any CRISPR-Cas protein, which has been mutated to be active as a nickase.

[0068] In some embodiments, the “-” used in the general architecture above indicates the presence of an optional linker sequence. In some embodiments, the fusion proteins provided herein do not comprise a linker sequence. In some embodiments, at least one of the optional linker sequences are present.

[0069] Other exemplary features that may be present are localization sequences, such as nuclear localization sequences, cytoplasmic localization sequences, export sequences, such as nuclear export sequences, or other localization sequences, as well as sequence tags that are useful

for solubilization, purification or detection of the fusion proteins. Suitable localization signal sequences and sequences of protein tags that are provided herein, and include, but are not limited to, biotin carboxylase carrier protein (BCCP) tags, myc-tags, calmodulin-tags, FLAG-tags, hemagglutinin (HA)-tags, polyhistidine tags, also referred to as histidine tags or His-tags, maltose binding protein (MBP)-tags, nus-tags, glutathione-S-transferase (GST)-tags, green fluorescent protein (GFP)-tags, thioredoxin-tags, S-tags, Softags (e.g., Softag 1, Softag 3), streptags, biotin ligase tags, FLASH tags, V5 tags, and SBP-tags. Additional suitable sequences will be apparent to those of skill in the art.

[0070] In certain embodiments, the presently disclosed fusion proteins comprise at least one cell-penetrating domain that facilitates cellular uptake of the fusion protein. Cell-penetrating domains are known in the art and generally comprise stretches of positively charged amino acid residues (i.e., polycationic cell-penetrating domains), alternating polar amino acid residues and non-polar amino acid residues (i.e., amphipathic cell-penetrating domains), or hydrophobic amino acid residues (i.e., hydrophobic cell-penetrating domains) (see, e.g., Milletti F. (2012) *Drug Discov Today* 17:850-860). A non-limiting example of a cell-penetrating domain is the trans-activating transcriptional activator (TAT) from the human immunodeficiency virus 1.

[0071] In some embodiments, deaminases or fusion proteins provided herein further comprise a nuclear localization sequence (NLS). The nuclear localization signal, plastid localization signal, mitochondrial localization signal, dual-targeting localization signal, and/or cell-penetrating domain can be located at the amino-terminus (N-terminus), the carboxyl-terminus (C-terminus), or in an internal location of the fusion protein.

[0072] In some embodiments, the NLS is fused to the N-terminus of the fusion protein or deaminase. In some embodiments, the NLS is fused to the C-terminus of the fusion protein or deaminase. In some embodiments, the NLS is fused to the N-terminus of the deaminase of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the deaminase of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the DNA-binding polypeptide (e.g., RGN polypeptide) of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the DNA-binding polypeptide (e.g., RGN polypeptide) of the fusion protein. In some embodiments, the NLS is fused to the N-terminus of the deaminase polypeptide of the fusion protein. In some embodiments, the NLS is fused to the C-terminus of the deaminase polypeptide of the fusion protein. In some embodiments, the NLS is fused to the fusion protein via one or more linkers. In some embodiments, the NLS is fused to the fusion protein without a linker. In some embodiments, the NLS comprises an amino acid sequence of any one of the NLS sequences provided or referenced herein. In some embodiments, the NLS comprises an amino acid sequence as set forth in SEQ ID NO: 43 or SEQ ID NO: 46. In some embodiments, the fusion protein or deaminase comprises SEQ ID NO: 43 on its N-terminus and SEQ ID NO: 46 on its C-terminus.

[0073] In some embodiments, fusion proteins as provided herein comprise the full-length sequence of a deaminase, e.g., any one of SEQ ID NO: 1-10 and 399-441. In some embodiments, however, fusion proteins as provided herein do not comprise a full-length sequence of a deaminase, but

only a fragment thereof. For example, in some embodiments, a fusion protein provided herein further comprises a DNA-binding polypeptide (e.g., an RNA-guided, DNA-binding) domain and a deaminase domain.

[0074] In some embodiments, a fusion protein of the invention comprises a DNA-binding polypeptide (e.g., an RGN) and a deaminase, wherein the deaminase has an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to any of SEQ ID NOs: 1-10 and 399-441. Examples of such fusion proteins are described in the Examples section herein.

[0075] In some embodiments, the fusion protein comprises one deaminase polypeptide. In some embodiments, the fusion protein comprises at least two deaminase polypeptides, operably linked either directly or via a peptide linker. In some embodiments, the fusion protein comprises one deaminase polypeptide, and a second deaminase polypeptide is co-expressed with the fusion protein.

[0076] Also provided herein is a ribonucleoprotein complex comprising a fusion protein comprising a deaminase and an RGDBP and the guide RNA, either as a single guide or as a dual guide RNA (also collectively referred to as gRNA).

#### V. Nucleotides Encoding Deaminases, Fusion Proteins, and/or gRNA

[0077] The present disclosure provides polynucleotides (SEQ ID NOs: 11-20 and 443-485) encoding the presently disclosed deaminase polypeptides. The present disclosure further provides polynucleotides encoding for fusion proteins which comprise a deaminase and DNA-binding polypeptide, for example a meganuclease, a zinc finger fusion protein, or a TALEN. The present disclosure further provides polynucleotides encoding for fusion proteins which comprise a deaminase domain and an RNA-guided, DNA-binding polypeptide. Such RNA-guided, DNA-binding polypeptides may be an RGN or RGN variant. The protein variant may be nuclease-inactive or a nickase. The RGN may be a CRISPR-Cas protein or active variant or fragment thereof. SEQ ID NOs: 41 and 42 are non-limiting examples of an RGN and a nickase RGN variant, respectively. Examples of CRISPR-Cas nucleases are well-known in the art, and similar corresponding mutations can create mutant variants which are also nickases or are nuclease inactive.

[0078] An embodiment of the invention provides a polynucleotide encoding a fusion protein which comprises an RGDBP and a deaminase described herein (SEQ ID NO: 1-10 and 399-441, or a variant thereof). In some embodiments, a second polynucleotide encodes the guide RNA required by the RGDBP for targeting to the nucleotide sequence of interest. In some embodiments, the guide RNA and the fusion protein are encoded by the same polynucleotide.

[0079] The use of the term "polynucleotide" is not intended to limit the present disclosure to polynucleotides comprising DNA, though such DNA polynucleotides are contemplated. Those of ordinary skill in the art will recognize that polynucleotides can comprise ribonucleotides (RNA) and combinations of ribonucleotides and deoxyribonucleotides. Such deoxyribonucleotides and ribonucleotides include both naturally occurring molecules and synthetic analogues. The polynucleotides disclosed herein also encompass all forms of sequences including, but not limited

to, single-stranded forms, double-stranded forms, stem-and-loop structures, circular forms (e.g., including circular RNA), and the like.

[0080] An embodiment of the invention is a nucleic acid molecule comprising a sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to any of SEQ ID NOs: 11-20 and 443-485, wherein the nucleic acid molecule encodes a deaminase having adenine deaminase activity. The nucleic acid molecule may further comprise a heterologous promoter or terminator. The nucleic acid molecule may encode a fusion protein, where the encoded deaminase is operably linked to a DNA-binding polypeptide, and optionally a second deaminase. In some embodiments, the nucleic acid molecule encodes a fusion protein, where the encoded deaminase is operably linked to an RGN and optionally a second deaminase.

[0081] In some embodiments, nucleic acid molecules comprising a polynucleotide which encodes a deaminase of the invention are codon optimized for expression in an organism of interest. A "codon-optimized" coding sequence is a polynucleotide coding sequence having its frequency of codon usage designed to mimic the frequency of preferred codon usage or transcription conditions of a particular host cell. Expression in the particular host cell or organism is enhanced as a result of the alteration of one or more codons at the nucleic acid level such that the translated amino acid sequence is not changed. Nucleic acid molecules can be codon optimized, either wholly or in part. Codon tables and other references providing preference information for a wide range of organisms are available in the art (see, e.g., Campbell and Gowri (1990) *Plant Physiol.* 92:1-11 for a discussion of plant-preferred codon usage). Methods are available in the art for synthesizing plant-preferred genes. See, for example, U.S. U.S. Pat. Nos. 5,380,831, and 5,436,391, and Murray et al. (1989) *Nucleic Acids Res.* 17:477-498, herein incorporated by reference.

[0082] In some embodiments, polynucleotides encoding the deaminases, fusion proteins, and/or gRNAs described herein are provided in expression cassettes for in vitro expression or expression in a cell, organelle, embryo, or organism of interest. The cassette may include 5' and 3' regulatory sequences operably linked to a polynucleotide encoding a deaminase and/or a fusion protein comprising a deaminase, an RNA-guided DNA-binding polypeptide and optionally a second deaminase, and/or gRNA provided herein that allows for expression of the polynucleotide. The cassette may additionally contain at least one additional gene or genetic element to be cotransformed into the organism. Where additional genes or elements are included, the components are operably linked. The term "operably linked" is intended to mean a functional linkage between two or more elements. For example, an operable linkage between a promoter and a coding region of interest (e.g., a region coding for a deaminase, RNA-guided DNA-binding polypeptide, and/or gRNA) is a functional link that allows for expression of the coding region of interest. Operably linked elements may be contiguous or non-contiguous. When used to refer to the joining of two protein coding regions, by operably linked is intended that the coding regions are in the same reading frame. In some embodiments, the additional gene(s) or element(s) are provided on multiple expression cassettes. For example, the nucleotide sequence encoding a

presently disclosed deaminase, either alone or as a component of a fusion protein, can be present on one expression cassette, whereas the nucleotide sequence encoding a gRNA can be on a separate expression cassette. Another example may have the nucleotide sequence encoding a presently disclosed deaminase alone on a first expression cassette, a second expression cassette encoding a fusion protein comprising a deaminase, and a nucleotide sequence encoding a gRNA on third expression cassette. Such an expression cassette is provided with a plurality of restriction sites and/or recombination sites for insertion of the polynucleotides to be under the transcriptional regulation of the regulatory regions. Expression cassettes which comprise a selectable marker gene may also be present.

[0083] The expression cassette may include in the 5'-3' direction of transcription, a transcriptional (and, in some embodiments, translational) initiation region (i.e., a promoter), a deaminase-encoding polynucleotide of the invention, and a transcriptional (and in some embodiments, translational) termination region (i.e., termination region) functional in the organism of interest. The promoters of the invention are capable of directing or driving expression of a coding sequence in a host cell. The regulatory regions (e.g., promoters, transcriptional regulatory regions, and translational termination regions) may be endogenous or heterologous to the host cell or to each other. As used herein, "heterologous" in reference to a sequence is a sequence that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus by deliberate human intervention. As used herein, a chimeric gene comprises a coding sequence operably linked to a transcription initiation region that is heterologous to the coding sequence.

[0084] Convenient termination regions are available from the Ti-plasmid of *A. tumefaciens*, such as the octopine synthase and nopaline synthase termination regions. See also Guerineau et al. (1991) *Mol. Gen. Genet.* 262:141-144; Proudfoot (1991) *Cell* 64:671-674; Sanfacon et al. (1991) *Genes Dev.* 5:141-149; Mogen et al. (1990) *Plant Cell* 2:1261-1272; Munroe et al. (1990) *Gene* 91:151-158; Ballas et al. (1989) *Nucleic Acids Res.* 17:7891-7903; and Joshi et al. (1987) *Nucleic Acids Res.* 15:9627-9639.

[0085] Additional regulatory signals include, but are not limited to, transcriptional initiation start sites, operators, activators, enhancers, other regulatory elements, ribosomal binding sites, an initiation codon, termination signals, and the like. See, for example, U.S. Pat. Nos. 5,039,523 and 4,853,331; EPO 0480762A2; Sambrook et al. (1992) *Molecular Cloning: A Laboratory Manual*, ed. Maniatis et al. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), hereinafter "Sambrook 11"; Davis et al., eds. (1980) *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory Press), Cold Spring Harbor, N.Y., and the references cited therein.

[0086] In preparing the expression cassette, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Toward this end, adapters or linkers may be employed to join the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like. For this purpose, in vitro mutagenesis, primer repair, restriction, annealing, resubstitutions, e.g., transitions and transversions, may be involved.

[0087] A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, inducible, growth stage-specific, cell type-specific, tissue-preferred, tissue-specific, or other promoters for expression in the organism of interest. See, for example, promoters set forth in WO 99/43838 and in U.S. Pat. Nos. 8,575,425; 7,790,846; 8,147,856; 8,586832; 7,772,369; 7,534,939; 6,072,050; 5,659,026; 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611; herein incorporated by reference.

[0088] For expression in plants, constitutive promoters also include CaMV 35S promoter (Odell et al. (1985) *Nature* 313:810-812); rice actin (McElroy et al. (1990) *Plant Cell* 2:163-171); ubiquitin (Christensen et al. (1989) *Plant Mol. Biol.* 12:619-632 and Christensen et al. (1992) *Plant Mol. Biol.* 18:675-689); pEMU (Last et al. (1991) *Theor. Appl. Genet.* 81:581-588); and MAS (Velten et al. (1984) *EMBO J.* 3:2723-2730).

[0089] Examples of inducible promoters are the Adh1 promoter which is inducible by hypoxia or cold stress, the Hsp70 promoter which is inducible by heat stress, the PPDK promoter and the pepcarboxylase promoter which are both inducible by light. Also useful are promoters which are chemically inducible, such as the In2-2 promoter which is safener induced (U.S. Pat. No. 5,364,780), the Axig1 promoter which is auxin induced and tapetum specific but also active in callus (PCT US01/22169), the steroid-responsive promoters (see, for example, the ERE promoter which is estrogen induced, and the glucocorticoid-inducible promoter in Schena et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:10421-10425 and McNellis et al. (1998) *Plant J.* 14 (2): 247-257) and tetracycline-inducible and tetracycline-repressible promoters (see, for example, Gatz et al. (1991) *Mol. Gen. Genet.* 227:229-237, and U.S. Pat. Nos. 5,814,618 and 5,789,156), herein incorporated by reference.

[0090] In some embodiments, tissue-specific or tissue-preferred promoters are utilized to target expression of an expression construct within a particular tissue. In certain embodiments, the tissue-specific or tissue-preferred promoters are active in plant tissue. Examples of promoters under developmental control in plants include promoters that initiate transcription preferentially in certain tissues, such as leaves, roots, fruit, seeds, or flowers. A “tissue specific” promoter is a promoter that initiates transcription only in certain tissues. Unlike constitutive expression of genes, tissue-specific expression is the result of several interacting levels of gene regulation. As such, promoters from homologous or closely related plant species can be preferable to use to achieve efficient and reliable expression of transgenes in particular tissues. In some embodiments, the expression comprises a tissue-preferred promoter. A “tissue preferred” promoter is a promoter that initiates transcription preferentially, but not necessarily entirely or solely in certain tissues.

[0091] In some embodiments, the nucleic acid molecules encoding a deaminase described herein comprise a cell type-specific promoter. A “cell type specific” promoter is a promoter that primarily drives expression in certain cell types in one or more organs. Some examples of plant cells in which cell type specific promoters functional in plants may be primarily active include, for example, BETL cells, vascular cells in roots, leaves, stalk cells, and stem cells. The nucleic acid molecules can also include cell type preferred promoters. A “cell type preferred” promoter is a promoter

that primarily drives expression mostly, but not necessarily entirely or solely in certain cell types in one or more organs. Some examples of plant cells in which cell type preferred promoters functional in plants may be preferentially active include, for example, BETL cells, vascular cells in roots, leaves, stalk cells, and stem cells.

[0092] In some embodiments, the nucleic acid sequences encoding the deaminases, fusion proteins, and/or gRNAs are operably linked to a promoter sequence that is recognized by a phage RNA polymerase for example, for in vitro mRNA synthesis. In such embodiments, the in vitro-transcribed RNA can be purified for use in the methods described herein. For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. In such embodiments, the expressed protein and/or RNAs can be purified for use in the methods of genome modification described herein.

[0093] In certain embodiments, the polynucleotide encoding the deaminase, fusion protein, and/or gRNA is linked to a polyadenylation signal (e.g., SV40 polyA signal and other signals functional in plants) and/or at least one transcriptional termination sequence. In some embodiments, the sequence encoding the deaminase or fusion protein is linked to sequence(s) encoding at least one nuclear localization signal, at least one cell-penetrating domain, and/or at least one signal peptide capable of trafficking proteins to particular subcellular locations, as described elsewhere herein.

[0094] In some embodiments, the polynucleotide encoding the deaminase, fusion protein, and/or gRNA is present in a vector or multiple vectors. A “vector” refers to a polynucleotide composition for transferring, delivering, or introducing a nucleic acid into a host cell. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors (e.g., lentiviral vectors, adeno-associated viral vectors, baculoviral vector). In some embodiments, the vector comprises additional expression control sequences (e.g., enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences), selectable marker sequences (e.g., antibiotic resistance genes), origins of replication, and the like. Additional information can be found in “Current Protocols in Molecular Biology” Ausubel et al., John Wiley & Sons, New York, 2003 or “Molecular Cloning: A Laboratory Manual” Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, N.Y., 3rd edition, 2001.

[0095] In some embodiments, the vector comprises a selectable marker gene for the selection of transformed cells. Selectable marker genes are utilized for the selection of transformed cells or tissues. Marker genes include genes encoding antibiotic resistance, such as those encoding neomycin phosphotransferase II (NEO) and hygromycin phosphotransferase (HPT), as well as genes conferring resistance to herbicidal compounds, such as glufosinate ammonium, bromoxynil, imidazolinones, and 2,4-dichlorophenoxyacetate (2,4-D).

[0096] In some embodiments, the expression cassette or vector comprising the sequence encoding a fusion protein comprising an RNA-guided DNA-binding polypeptide, such as an RGN, further comprises a sequence encoding a gRNA. In some embodiments, the sequence(s) encoding the gRNA are operably linked to at least one transcriptional control sequence for expression of the gRNA in the organism or host cell of interest. For example, the polynucleotide encoding the gRNA can be operably linked to a promoter sequence

that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters and rice U6 and U3 promoters.

[0097] As indicated, expression constructs comprising nucleotide sequences encoding the deaminases, fusion proteins, and/or gRNAs can be used to transform organisms of interest. Methods for transformation involve introducing a nucleotide construct into an organism of interest. By “introducing” is intended to introduce the nucleotide construct to the host cell in such a manner that the construct gains access to the interior of the host cell. The methods of the invention do not require a particular method for introducing a nucleotide construct to a host organism, only that the nucleotide construct gains access to the interior of at least one cell of the host organism. The host cell can be a eukaryotic or prokaryotic cell. In particular embodiments, the eukaryotic host cell is a plant cell, a mammalian cell, or an insect cell. Methods for introducing nucleotide constructs into plants and other host cells are known in the art including, but not limited to, stable transformation methods, transient transformation methods, and virus-mediated methods.

[0098] The methods result in a transformed organism, such as a plant, including whole plants, as well as plant organs (e.g., leaves, stems, roots, etc.), seeds, plant cells, propagules, embryos and progeny of the same. Plant cells can be differentiated or undifferentiated (e.g. callus, suspension culture cells, protoplasts, leaf cells, root cells, phloem cells, pollen).

[0099] “Transgenic organisms” or “transformed organisms” or “stably transformed” organisms or cells or tissues refers to organisms that have incorporated or integrated a polynucleotide encoding a deaminase of the invention. It is recognized that other exogenous or endogenous nucleic acid sequences or DNA fragments may also be incorporated into the host cell. *Agrobacterium*-and biostatic-mediated transformation remain the two predominantly employed approaches for transformation of plant cells. However, transformation of a host cell may be performed by infection, transfection, microinjection, electroporation, microporation, biolistics or particle bombardment, electroporation, silica/carbon fibers, ultrasound mediated, PEG mediated, calcium phosphate co-precipitation, polycation DMSO technique, DEAE dextran procedure, and viral mediated, liposome mediated and the like. Viral-mediated introduction of a polynucleotide encoding a deaminase, fusion protein, and/or gRNA includes retroviral, lentiviral, adenoviral, and adeno-associated viral mediated introduction and expression, as well as the use of Caulimoviruses (e.g., cauliflower mosaic virus), Geminiviruses (e.g., bean golden yellow mosaic virus or maize streak virus), and RNA plant viruses (e.g., tobacco mosaic virus).

[0100] Transformation protocols as well as protocols for introducing polypeptides or polynucleotide sequences into plants may vary depending on the type of host cell (e.g., monocot or dicot plant cell) targeted for transformation. Methods for transformation are known in the art and include those set forth in U.S. Pat. Nos. 8,575,425; 7,692,068; 8,802,934; 7,541,517; each of which is herein incorporated by reference. See, also, Rakoczy-Trojanowska, M. (2002) *Cell Mol Biol Lett.* 7:849-858; Jones et al. (2005) *Plant Methods* 1:5; Rivera et al. (2012) *Physics of Life Reviews* 9:308-345; Bartlett et al. (2008) *Plant Methods* 4:1-12; Bates, G.W. (1999) *Methods in Molecular Biology* 111:359-

366; Binns and Thomashow (1988) *Annual Reviews in Microbiology* 42:575-606; Christou, P. (1992) *The Plant Journal* 2:275-281; Christou, P. (1995) *Euphytica* 85:13-27; Tzfira et al. (2004) *TRENDS in Genetics* 20:375-383; Yao et al. (2006) *Journal of Experimental Botany* 57:3737-3746; Zupan and Zambryski (1995) *Plant Physiology* 107:1041-1047; Jones et al. (2005) *Plant Methods* 1:5;

[0101] Transformation may result in stable or transient incorporation of the nucleic acid into the cell. “Stable transformation” is intended to mean that the nucleotide construct introduced into a host cell integrates into the genome of the host cell and is capable of being inherited by the progeny thereof. “Transient transformation” is intended to mean that a polynucleotide is introduced into the host cell and does not integrate into the genome of the host cell.

[0102] Methods for transformation of chloroplasts are known in the art. See, for example, Svab et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530; Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917; Svab and Maliga (1993) *EMBO J.* 12:601-606. The method relies on particle gun delivery of DNA containing a selectable marker and targeting of the DNA to the plastid genome through homologous recombination. Additionally, plastid transformation can be accomplished by transactivation of a silent plastid-borne transgene by tissue-preferred expression of a nuclear-encoded and plastid-directed RNA polymerase. Such a system has been reported in McBride et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:7301-7305.

[0103] The cells that have been transformed may be grown into a transgenic organism, such as a plant, in accordance with conventional ways. See, for example, McCormick et al. (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same transformed strain or different strains, and the resulting hybrid having the deaminase or fusion protein polynucleotide identified. Two or more generations may be grown to ensure that the deaminase or fusion protein polynucleotide is stably maintained and inherited and the seeds harvested to ensure the presence of the deaminase or fusion protein polynucleotide. In this manner, the present invention provides transformed seed (also referred to as “transgenic seed”) having a nucleotide construct of the invention, for example, an expression cassette of the invention, stably incorporated into their genome.

[0104] In some embodiments, cells that have been transformed are introduced into an organism. These cells could have originated from the organism, wherein the cells are transformed in an ex vivo approach.

[0105] The sequences provided herein may be used for transformation of any plant species, including, but not limited to, monocots and dicots. Examples of plants of interest include, but are not limited to, corn (maize), sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, and oilseed rape, *Brassica* sp., alfalfa, rye, millet, safflower, peanuts, sweet potato, cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, papaya, cashew, macadamia, almond, oats, vegetables, ornamentals, and conifers.

[0106] Vegetables include, but are not limited to, tomatoes, lettuce, green beans, lima beans, peas, and members of the genus *Cucurbita* such as cucumber, cantaloupe, and musk melon. Ornamentals include, but are not limited to, azalea, hydrangea, hibiscus, roses, tulips, daffodils, petunias,

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

[0107] As used herein, the term plant includes plant cells, plant protoplasts, plant cell tissue cultures from which plants can be regenerated, plant calli, plant clumps, and plant cells that are intact in plants or parts of plants such as embryos, pollen, ovules, seeds, leaves, flowers, branches, fruit, kernels, ears, cobs, husks, stalks, roots, root tips, anthers, and the like. Grain is intended to mean the mature seed produced by commercial growers for purposes other than growing or reproducing the species. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced polynucleotides. Further provided is a processed plant product or byproduct that retains the sequences disclosed herein, including for example, soymeal.

[0108] In some embodiments, the polynucleotides encoding the deaminases, fusion proteins, and/or gRNAs are used to transform any eukaryotic species, including but not limited to animals (e.g., mammals, insects, fish, birds, and reptiles), fungi, amoeba, algae, and yeast. In some embodiments, the polynucleotides encoding the deaminases, fusion proteins, and/or gRNAs are used to transform any prokaryotic species, including but not limited to, archaea and bacteria (e.g., *Bacillus* spp., *Klebsiella* spp., *Streptomyces* spp., *Rhizobium* spp., *Escherichia* spp., *Pseudomonas* spp., *Salmonella* spp., *Shigella* spp., *Vibrio* spp., *Yersinia* spp., *Mycoplasma* spp., *Agrobacterium* spp., and *Lactobacillus* spp.).

[0109] In some embodiments, conventional viral and non-viral based gene transfer methods are used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding a deaminase or fusion protein of the invention and optionally a gRNA to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g., a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. Non-limiting examples include vectors utilizing Caulimoviruses (e.g., cauliflower mosaic virus), Geminiviruses (e.g., bean golden yellow mosaic virus or maize steak virus), and RNA plant viruses (e.g., tobacco mosaic virus). For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Feigner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6 (10): 1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51 (1): 31-44 (1995); Haddada et al., in *Current Topics in Microbiology and Immunology*, Doerfler and Bohm (eds) (1995); and Yu et al., *Gene Therapy* 1:13-26 (1994).

[0110] Methods of non-viral delivery of nucleic acids include lipofection, *Agrobacterium*-mediated transformation, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid: nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g.,

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

[0111] The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro*, and the modified cells may optionally be administered to patients (*ex vivo*). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0112] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of *cis*-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum *cis*-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Sommnerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700).

[0113] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy

procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Katin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989). Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψJ2 cells or PA317 cells, which package retrovirus.

[0114] Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences.

[0115] The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0116] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject.

[0117] In some embodiments, a cell that is transfected is a eukaryotic cell. In some embodiments, the eukaryotic cell is an animal cell (e.g., mammals, insects, fish, birds, and reptiles). In some embodiments, a cell that is transfected is a human cell. In some embodiments, a cell that is transfected is a cell of hematopoietic origin, such as an immune cell (i.e., a cell of the innate or adaptive immune system) including but not limited to a B cell, a T cell, a natural killer (NK) cell, a pluripotent stem cell, an induced pluripotent stem cell, a chimeric antigen receptor T (CAR-T) cell, a monocyte, a macrophage, and a dendritic cell.

[0118] In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. In some embodiments, the cell or cell line is prokaryotic. In some embodiments, the cell or cell line is eukaryotic. In further embodiments, the cell or cell line is derived from insect, avian, plant, or fungal species. In some embodiments, the cell or cell line may be mammalian, such as for example human, monkey, mouse, cow, swine, goat, hamster, rat, cat, or dog. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLaS3,

Huhl, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TFI, CTLL-2, CIR, Rat6, CVI, RPTE, AIO, T24, 182, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388DI, SEM-K2, WEHI-231, HB56, TIB55, lurkat, 145.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-I cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO—IR, CHO-KI, CHO-K2, CHO-T, CHO Dhfr-/-, COR-L23, COR-L23/CPR, COR-L235010, CORL23/R23, COS-7, COV-434, CML TI, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, lurkat, IY cells, K562 cells, Ku812, KCL22, KGI, KYOI, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCKII, MDCKII, MOR/0.2R, MONO-MAC 6, MTD-IA, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN/OPCT cell lines, Peer, PNT-1A/PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THPI cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)).

[0119] In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with a fusion protein of the invention and optionally a gRNA, or with a ribonucleoprotein complex of the invention, and modified through the activity of a fusion protein or ribonucleoprotein complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0120] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is an insect. In further embodiments, the insect is an insect pest, such as a mosquito or tick. In some embodiments, the insect is a plant pest, such as a corn rootworm or a fall armyworm. In some embodiments, the transgenic animal is a bird, such as a chicken, turkey, goose, or duck. In some embodiments, the transgenic animal is a mammal, such as a human, mouse, rat, hamster, monkey, ape, rabbit, swine, cow, horse, goat, sheep, cat, or dog.

## VI. Variants and Fragments of Polypeptides and Polynucleotides

[0121] The present disclosure provides novel adenine deaminases which are active on DNA molecules, the amino acid sequence of which are set forth as SEQ ID NO: 1-10 and 399-441, active variants or fragments thereof, and polynucleotides encoding the same.

[0122] While the activity of a variant or fragment may be altered compared to the polynucleotide or polypeptide of interest, the variant and fragment should retain the functionality of the polynucleotide or polypeptide of interest. For example, a variant or fragment may have increased activity, decreased activity, different spectrum of activity or any other alteration in activity when compared to the polynucleotide or polypeptide of interest.

[0123] Fragments and variants of deaminases of the invention which have adenine deaminase activity will retain said activity if they are part of a fusion protein further comprising a DNA-binding polypeptide or a fragment thereof.

[0124] The term "fragment" refers to a portion of a polynucleotide or polypeptide sequence of the invention. "Fragments" or "biologically active portions" include polynucleotides comprising a sufficient number of contiguous nucleotides to retain the biological activity (i.e., deaminase activity on nucleic acids). "Fragments" or "biologically active portions" include polypeptides comprising a sufficient number of contiguous amino acid residues to retain the biological activity. Fragments of the deaminases disclosed herein include those that are shorter than the full-length sequences due to the use of an alternate downstream start site. In some embodiments, a biologically active portion of a deaminase is a polypeptide that comprises, for example, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, or more contiguous amino acid residues of any of SEQ ID NOs: 1-10 and 399-441, or a variant thereof. Such biologically active portions can be prepared by recombinant techniques and evaluated for activity.

[0125] In general, "variants" is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a deletion and/or addition of one or more nucleotides at one or more internal sites within the native polynucleotide and/or a substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" or "wild type" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. For polynucleotides, conservative variants include those sequences that, because of the degeneracy of the genetic code, encode the native amino acid sequence of the gene of interest. Naturally occurring allelic variants such as these can be identified with the use of well-known molecular biology techniques, as, for example, with polymerase chain reaction (PCR) and hybridization techniques as outlined below. Variant polynucleotides also include synthetically derived polynucleotides, such as those generated, for example, by using site-directed mutagenesis but which still encode the polypeptide or the polynucleotide of interest. Generally, variants of a particular polynucleotide disclosed herein will have at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity to that particular polynucleotide as determined by sequence alignment programs and parameters described elsewhere herein.

[0126] Variants of a particular polynucleotide disclosed herein (i.e., the reference polynucleotide) can also be evaluated by comparison of the percent sequence identity between the polypeptide encoded by a variant polynucleotide and the polypeptide encoded by the reference polynucleotide. Percent sequence identity between any two polypeptides can be

calculated using sequence alignment programs and parameters described elsewhere herein. Where any given pair of polynucleotides disclosed herein is evaluated by comparison of the percent sequence identity shared by the two polypeptides they encode, the percent sequence identity between the two encoded polypeptides is at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or more sequence identity.

[0127] In particular embodiments, the presently disclosed polynucleotides encode an adenine deaminase comprising an amino acid sequence having at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 81%, at least 82%, at least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or greater identity to an amino acid sequence of any of SEQ ID NOs: 1-10 and 399-441.

[0128] A biologically active variant of an adenine deaminase of the invention may differ by as few as 1-15 amino acid residues, as few as 1-10, such as 6-10, as few as 5, as few as 4, as few as 3, as few as 2, or as few as 1 amino acid residue. In specific embodiments, the polypeptides comprise an N-terminal or a C-terminal truncation, which can comprise at least a deletion of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60 amino acids or more from either the N or C terminus of the polypeptide. In some embodiments, the polypeptides comprise an internal deletion which can comprise at least a deletion of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60 amino acids or more.

[0129] It is recognized that modifications may be made to the deaminases provided herein creating variant proteins and polynucleotides. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques. In some embodiments, native, as yet-unknown or as yet unidentified polynucleotides and/or polypeptides structurally and/or functionally-related to the sequences disclosed herein may also be identified that fall within the scope of the present invention. Conservative amino acid substitutions may be made in nonconserved regions that do not alter the function of the polypeptide as an adenine deaminase. In some embodiments, modifications are made that improve the adenine deaminase activity of the deaminase.

[0130] Variant polynucleotides and proteins also encompass sequences and proteins derived from a mutagenic and recombinogenic procedure such as DNA shuffling. With such a procedure, one or more different deaminases disclosed herein (e.g., SEQ ID NO: 1-10 and 399-441) is manipulated to create a new adenine deaminase possessing the desired properties. In this manner, libraries of recombinant polynucleotides are generated from a population of related sequence polynucleotides comprising sequence regions that have substantial sequence identity and can be homologously recombined in vitro or in vivo. For example, using this approach, sequence motifs encoding a domain of interest may be shuffled between the deaminase sequences provided herein and other subsequently identified deaminase genes to obtain a new gene coding for a protein with an

improved property of interest, such as an increased  $K_m$  in the case of an enzyme. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) *Proc. Natl. Acad. Sci. USA* 91:10747-10751; Stemmer (1994) *Nature* 370:389-391; Crameri et al. (1997) *Nature Biotech.* 15:436-438; Moore et al. (1997) *J. Mol. Biol.* 272:336-347; Zhang et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:4504-4509; Crameri et al. (1998) *Nature* 391:288-291; and U.S. Pat. Nos. 5,605,793 and 5,837,458. A "shuffled" nucleic acid is a nucleic acid produced by a shuffling procedure such as any shuffling procedure set forth herein. Shuffled nucleic acids are produced by recombining (physically or virtually) two or more nucleic acids (or character strings), for example in an artificial, and optionally recursive, fashion. Generally, one or more screening steps are used in shuffling processes to identify nucleic acids of interest; this screening step can be performed before or after any recombination step. In some (but not all) shuffling embodiments, it is desirable to perform multiple rounds of recombination prior to selection to increase the diversity of the pool to be screened. The overall process of recombination and selection are optionally repeated recursively. Depending on context, shuffling can refer to an overall process of recombination and selection, or, alternately, can simply refer to the recombinational portions of the overall process.

[0131] As used herein, "sequence identity" or "identity" in the context of two polynucleotides or polypeptide sequences makes reference to the residues in the two sequences that are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g., charge or hydrophobicity) and therefore do not change the functional properties of the molecule. When sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences that differ by such conservative substitutions are said to have "sequence similarity" or "similarity". Means for making this adjustment are well known to those of skill in the art. Typically, this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, e.g., as implemented in the program PC/GENE (Intelligenetics, Mountain View, California).

[0132] As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the

window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity.

[0133] Unless otherwise stated, sequence identity/similarity values provided herein refer to the value obtained using GAP Version 10 using the following parameters: % identity and % similarity for a nucleotide sequence using GAP Weight of 50 and Length Weight of 3, and the nwsgapdna.cmp scoring matrix; % identity and % similarity for an amino acid sequence using GAP Weight of 8 and Length Weight of 2, and the BLOSUM62 scoring matrix; or any equivalent program thereof. By "equivalent program" is intended any sequence comparison program that, for any two sequences in question, generates an alignment having identical nucleotide or amino acid residue matches and an identical percent sequence identity when compared to the corresponding alignment generated by GAP Version 10.

[0134] Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (e.g., BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, e.g., in Dayhoff et al. (1978) "A model of evolutionary change in proteins." In "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10915-10919. The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences, so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, e.g., gapped BLAST 2.0, described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402, and made available to the public at the National Center for Biotechnology Information Website ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Optimal alignments, including multiple alignments, can be prepared using, e.g., PSI-BLAST, available through [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) and described by Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402.

[0135] With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue "corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence as determined by simply counting from the N-terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the

reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

## VII. Antibodies

[0136] Antibodies to the deaminases, fusion proteins, or ribonucleoproteins comprising the deaminases of the present invention, including those having the amino acid sequence set forth as any one of SEQ ID NOs: 1-10 and 399-441 or active variants or fragments thereof, are also encompassed. Methods for producing antibodies are well known in the art (see, for example, Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; and U.S. Pat. No. 4,196,265). These antibodies can be used in kits for the detection and isolation of deaminases or fusion proteins or ribonucleoproteins comprising deaminases described herein. Thus, this disclosure provides kits comprising antibodies that specifically bind to the polypeptides or ribonucleoproteins described herein, including, for example, polypeptides comprising a sequence of at least 85% identity to any of SEQ ID NOs: 1-10 and 399-441.

## VIII. Systems and Ribonucleoprotein Complexes for Binding and/or Modifying a Target Sequence of Interest and Methods of Making the Same

[0137] The present disclosure provides a system which targets to a nucleic acid sequence and modifies a target nucleic acid sequence. In some embodiments, an RNA-guided, DNA-binding polypeptide, such as an RGN, and the gRNA are responsible for targeting the ribonucleoprotein complex to a nucleic acid sequence of interest; the deaminase polypeptide fused to the RGDBP is responsible for modifying the targeted nucleic acid sequence from A>N. In some embodiments, the deaminase converts A>G. The guide RNA hybridizes to the target sequence of interest and also forms a complex with the RNA-guided, DNA-binding polypeptide, thereby directing the RNA-guided, DNA-binding polypeptide to bind to the target sequence. The RNA-guided, DNA-binding polypeptide is one domain of a fusion protein; the second domain is a deaminase described herein. In some embodiments, the RNA-guided, DNA-binding polypeptide is an RGN, such as a Cas9. Other examples of RNA-guided, DNA-binding polypeptides include RGNs such as those described in International Patent Application Publication Nos. WO 2019/236566 and WO 2020/139783. In some embodiments, the RNA-guided, DNA-binding polypeptide is a Type II CRISPR-Cas polypeptide, or an active variant or fragment thereof. In some embodiments, the RNA-guided, DNA-binding polypeptide is a Type V CRISPR-Cas polypeptide, or an active variant or fragment thereof. In some embodiments, the RNA-guided, DNA-binding polypeptide is a Type VI CRISPR-Cas polypeptide. In some embodiments, the DNA-binding domain of the fusion protein does not require an RNA guide, such as a zinc finger nuclease, TALEN, or meganuclease polypeptide. In some embodiments, the nuclease activity of a DNA-binding domain has been partially or completely inactivated. In further embodiments, the RNA-guided, DNA-binding polypeptide comprises an amino acid sequence of an RGN, such as for example APG07433.1 (SEQ ID NO: 41), or an active variant

or fragment thereof such as nickase nAPG07433.1 (SEQ ID NO: 42) or other nickase RGN variants described in the Examples (SEQ ID NOs: 52-59, 61, 397, and 398).

[0138] In some embodiments, the system for binding and modifying a target sequence of interest provided herein is a ribonucleoprotein complex, which is at least one molecule of an RNA bound to at least one protein. The ribonucleoprotein complexes provided herein comprise at least one guide RNA as the RNA component and a fusion protein comprising a deaminase of the invention and an RNA-guided, DNA-binding polypeptide as the protein component. In some embodiments, the ribonucleoprotein complex is purified from a cell or organism that has been transformed with polynucleotides that encode the fusion protein and a guide RNA and cultured under conditions to allow for the expression of the fusion protein and guide RNA.

[0139] In various embodiments, ribonucleoprotein complexes comprising any of the fusion proteins described herein and a guide RNA bound to the DNA-binding polypeptide of the fusion protein, are provided. For example, provided herein is a ribonucleoprotein complex comprising a fusion protein with a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. In another instance, a ribonucleoprotein complex comprising a fusion protein with a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 399, is provided. In yet another example, a ribonucleoprotein complex comprising a fusion protein with a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405, is provided. In some of those embodiments of the ribonucleoprotein complexes described above, the fusion protein comprises an RGN selected from a CasX, a CasY, a C2c1, a C2c2, a C2c3, a GeoCas9, aSpCas9, aSaCas9, aNme2Cas9, aCjCas9, aCas12a (formerly known as Cpf1), aCas12b, aCas12g, aCas12h, aCas12i, aLbCas12a, aAsCas12a, aCasMINI, aCas13b, aCas13c, aCas13d, aCas14, aCsn2, anxCas9, anSpCas9-NG, anLbCas12a, anAsCas12a, aCas9-KKH, a circularly permuted Cas9, anArgonaute (Ago), aSmacCas9, aSpy-macCas9 domain, or anRGN with an amino acid sequence set forth in any one of SEQ ID NOs: 41, 60, 366, or 368. In some embodiments, the ribonucleoprotein complex comprises a nickase having an amino acid sequence with at least 95% sequence identity to any one of SEQ ID NOs: 42, 52-59, 61, 397, and 398, fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. In some embodiments, the ribonucleoprotein complex comprises a nickase having an amino acid sequence with at least 95% sequence identity to any one of SEQ ID NOs: 42, 52-59, 61, 397, and 398, fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 399. In some embodiments, the ribonucleoprotein complex comprises a nickase having an amino acid sequence with at least 95% sequence identity to any one of SEQ ID NOs: 42, 52-59, 61, 397, and 398, fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405. In some embodiments, the ribonucleoprotein complex comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 407. In some embodiments, the ribonucleoprotein complex comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405.

sequence identity to SEQ ID NO: 399. In some embodiments, the ribonucleoprotein complex comprises a Cas9 nickase fused to a deaminase comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO: 405. The Cas9 nickase, can be any Cas9 nickase disclosed in PCT Patent Publication No. WO2020181195, the entire contents of which is incorporated herein by reference herein. In various embodiments described herein, the ribonucleoprotein complex may also contain the gRNAs described herein.

[0140] Methods are provided for making a deaminase, a fusion protein, or a fusion protein ribonucleoprotein complex. Such methods comprise culturing a cell comprising a nucleotide sequence encoding a deaminase, a fusion protein, and in some embodiments a nucleotide sequence encoding a guide RNA, under conditions in which the deaminase or fusion protein (and in some embodiments, the guide RNA) is expressed. The deaminase, fusion protein, or fusion ribonucleoprotein can then be purified from a lysate of the cultured cells.

[0141] Methods for purifying a deaminase, fusion protein, or fusion ribonucleoprotein complex from a lysate of a biological sample are known in the art (e.g., size exclusion and/or affinity chromatography, 2D-PAGE, HPLC, reversed-phase chromatography, immunoprecipitation). In particular methods, the deaminase or fusion protein is recombinantly produced and comprises a purification tag to aid in its purification, including but not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly (NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, biotin carboxyl carrier protein (BCCP), and calmodulin. Generally, the tagged deaminase, fusion protein, or fusion ribonucleoprotein complex is purified using immunoprecipitation or other similar methods known in the art.

[0142] An “isolated” or “purified” polypeptide, or biologically active portion thereof, is substantially or essentially free from components that normally accompany or interact with the polypeptide as found in its naturally occurring environment. Thus, an isolated or purified polypeptide is substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized. A protein that is substantially free of cellular material includes preparations of protein having less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% (by dry weight) of contaminating protein. When the protein of the invention or biologically active portion thereof is recombinantly produced, optimally culture medium represents less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% (by dry weight) of chemical precursors or non-protein-of-interest chemicals.

[0143] Particular methods provided herein for binding and/or cleaving a target sequence of interest involve the use of a ribonucleoprotein complex. In some embodiments, the ribonucleoprotein complex is assembled *in vitro*. *In vitro* assembly of a ribonucleoprotein complex can be performed using any method known in the art in which an RGDBP polypeptide or a fusion protein comprising the same is contacted with a guide RNA under conditions to allow for binding of the RGDBP polypeptide or fusion protein comprising the same to the guide RNA. As used herein, “con-

tact”, “contacting”, “contacted,” refer to placing the components of a desired reaction together under conditions suitable for carrying out the desired reaction. In some embodiments of the described methods for modifying a target DNA molecule, the step of contacting is performed *in vitro*. In some embodiments, the step of contacting is performed *in vivo*. In some embodiments, the step of contacting is performed in a subject (e.g., a human subject or a non-human animal subject). In some embodiments, the step of contacting is performed in a cell, such as a human or non-human animal cell. The RGDBP polypeptide or fusion protein comprising the same can be purified from a biological sample, cell lysate, or culture medium, produced via *in vitro* translation, or chemically synthesized. The guide RNA can be purified from a biological sample, cell lysate, or culture medium, transcribed *in vitro*, or chemically synthesized. The RGDBP polypeptide or fusion protein comprising the same and guide RNA can be brought into contact in solution (e.g., buffered saline solution) to allow for *in vitro* assembly of the ribonucleoprotein complex.

## IX. Methods of Modifying a Target Sequence

[0144] The present disclosure provides methods for modifying a target nucleic acid molecule (e.g., target DNA molecule) of interest. The methods include delivering a fusion protein comprising a DNA-binding polypeptide and at least one deaminase of the invention or a polynucleotide encoding the same to a target sequence or a cell, organelle, or embryo comprising a target sequence. In certain embodiments, the methods include delivering a system comprising at least one guide RNA or a polynucleotide encoding the same, and at least one fusion protein comprising at least one deaminase of the invention and an RNA-guided, DNA-binding polypeptide or a polynucleotide encoding the same to the target sequence or a cell, organelle, or embryo comprising the target sequence. In some embodiments, the fusion protein comprises any one of the amino acid sequences of SEQ ID NOS: 1-10 and 399-441, or an active variant or fragment thereof.

[0145] In some embodiments, the methods comprise contacting a DNA molecule with (a) a fusion protein comprising a deaminase and an RNA-guided, DNA-binding polypeptide, such as for example a nuclease-inactive or a nickase Cas9 domain; and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA molecule; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleobase. In some embodiments, the target DNA molecule comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleobase results in a sequence that is not associated with a disease or disorder. In some embodiments, the disease or disorder affects animals. In further embodiments, the disease or disorder affects mammals, such as humans, cows, horses, dogs, cats, goats, sheep, swine, monkeys, rats, mice, or hamsters. In some embodiments, the target DNA sequence resides in an allele of a crop plant, wherein the particular allele of the trait of interest results in a plant of lesser agronomic value. The deamination of the nucleobase results in an allele that improves the trait and increases the agronomic value of the plant.

[0146] In those embodiments wherein the method comprises delivering a polynucleotide encoding a guide RNA and/or a fusion protein, the cell or embryo can then be

cultured under conditions in which the guide RNA and/or fusion protein are expressed. In various embodiments, the method comprises contacting a target sequence with a ribonucleoprotein complex comprising a gRNA and a fusion protein (which comprises a deaminase of the invention and an RNA-guided DNA-binding polypeptide). In certain embodiments, the method comprises introducing into a cell, organelle, or embryo comprising a target sequence a ribonucleoprotein complex of the invention. The ribonucleoprotein complex of the invention can be one that has been purified from a biological sample, recombinantly produced and subsequently purified, or in vitro-assembled as described herein. In those embodiments wherein the ribonucleoprotein complex that is contacted with the target sequence or a cell organelle, or embryo has been assembled in vitro, the method can further comprise the in vitro assembly of the complex prior to contact with the target sequence, cell, organelle, or embryo.

[0147] A purified or in vitro assembled ribonucleoprotein complex of the invention can be introduced into a cell, organelle, or embryo using any method known in the art, including, but not limited to electroporation. In some embodiments, a fusion protein comprising a deaminase of the invention and an RNA-guided, DNA-binding polypeptide, and a polynucleotide encoding or comprising the guide RNA is introduced into a cell, organelle, or embryo using any method known in the art (e.g., electroporation).

[0148] Upon delivery to or contact with the target sequence or cell, organelle, or embryo comprising the target sequence, the guide RNA directs the fusion protein to bind to the target sequence in a sequence-specific manner. The target sequence can subsequently be modified via the deaminase domain of the fusion protein. In some embodiments, the binding of this fusion protein to a target sequence results in modification of a nucleotide adjacent to the target sequence. The nucleobase adjacent to the target sequence that is modified by the deaminase may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 base pairs from the 5' or 3' end of the target sequence. A fusion protein comprising a deaminase of the invention and an RNA-guided, DNA-binding polypeptide can introduce targeted A>N, and preferably targeted A>G, mutations in the targeted DNA molecule.

[0149] In some embodiments of the described methods for modifying a target DNA molecule, the step of contacting is performed in vitro. In particular embodiments, the step of contacting is performed in vivo. In some embodiments, the step of contacting is performed in a subject (e.g., a human subject or a non-human animal subject). In some embodiments, the step of contacting is performed in a cell, such as a human or non-human animal cell.

[0150] Methods to measure binding of the fusion protein to a target sequence are known in the art and include chromatin immunoprecipitation assays, gel mobility shift assays, DNA pull-down assays, reporter assays, microplate capture and detection assays. Likewise, methods to measure cleavage or modification of a target sequence are known in the art and include in vitro or in vivo cleavage assays wherein cleavage is confirmed using PCR, sequencing, or gel electrophoresis, with or without the attachment of an appropriate label (e.g., radioisotope, fluorescent substance) to the target sequence to facilitate detection of degradation products. In some embodiments, the nicking triggered expo-

nential amplification reaction (NTEXPAR) assay is used (see, e.g., Zhang et al. (2016) *Chem. Sci.* 7:4951-4957). In vivo cleavage can be evaluated using the Surveyor assay (Guschin et al. (2010) *Methods Mol Biol* 649:247-256).

[0151] In some embodiments, the methods involve the use of an RNA-binding, DNA-guided domain, as part of the fusion protein, complexed with more than one guide RNA. The more than one guide RNA can target different regions of a single gene or can target multiple genes. This multiple targeting enables the deaminase domain of the fusion protein to modify nucleic acids, thereby introducing multiple mutations in the target nucleic acid molecule (e.g., genome) of interest.

[0152] In those embodiments wherein the method involves the use of an RNA-guided nuclease (RGN), such as a nickase RGN (i.e., is only able to cleave a single strand of a double-stranded polynucleotide, for example nAPG07433.1 (SEQ ID NO: 42 or SEQ ID NOs: 50-57), the method can comprise introducing two different RGNs or RGN variants that target identical or overlapping target sequences and cleave different strands of the polynucleotide. For example, an RGN nickase that only cleaves the positive (+) strand of a double-stranded polynucleotide can be introduced along with a second RGN nickase that only cleaves the negative (-) strand of a double-stranded polynucleotide. In some embodiments, two different fusion proteins are provided, where each fusion protein comprises a different RGN with a different PAM recognition sequence, so that a greater diversity of nucleotide sequences may be targeted for mutation.

[0153] One of ordinary skill in the art will appreciate that any of the presently disclosed methods can be used to target a single target sequence or multiple target sequences. Thus, methods comprise the use of a fusion protein comprising a single RNA-guided, DNA-binding polypeptide in combination with multiple, distinct guide RNAs, which can target multiple, distinct sequences within a single gene and/or multiple genes. The deaminase domain of the fusion protein would then introduce mutations at each of the targeted sequences. Also encompassed herein are methods wherein multiple, distinct guide RNAs are introduced in combination with multiple, distinct RNA-guided, DNA binding polypeptides. Such RNA-guided, DNA-binding polypeptides may be multiple RGN or RGN variants. These guide RNAs and guide RNA/fusion protein systems can target multiple, distinct sequences within a single gene and/or multiple genes.

[0154] In some embodiments, a fusion protein comprising an RNA-guided, DNA-binding polypeptide and a deaminase polypeptide of the invention may be used for generating mutations in a targeted gene or targeted region of a gene of interest. In some embodiments, a fusion protein of the invention may be used for saturation mutagenesis of a targeted gene or region of a targeted gene of interest followed by high-throughput forward genetic screening to identify novel mutations and/or phenotypes. In some embodiments, a fusion protein described herein may be used for generating mutations in a targeted genomic location, which may or may not comprise coding DNA sequence. Libraries of cell lines generated by the targeted mutagenesis described above may also be useful for study of gene function or gene expression.

## X. Target Polynucleotides

[0155] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal or plant (including microalgae) and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the human, non-human animal or plant (including micro-algae).

[0156] Using natural variability, plant breeders combine most useful genes for desirable qualities, such as yield, quality, uniformity, hardiness, and resistance against pests. These desirable qualities also include growth, day length preferences, temperature requirements, initiation date of floral or reproductive development, fatty acid content, insect resistance, disease resistance, nematode resistance, fungal resistance, herbicide resistance, tolerance to various environmental factors including drought, heat, wet, cold, wind, and adverse soil conditions including high salinity. The sources of these useful genes include native or foreign varieties, heirloom varieties, wild plant relatives, and induced mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can employ the present invention to induce the rise of useful genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[0157] The target polynucleotide of a deaminase or a fusion protein of the invention can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. In some embodiments, the target polynucleotide is a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). In some embodiments, the target sequence for a fusion protein of the invention is associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the RNA-guided DNA-binding polypeptide. The precise sequence and length requirements for the PAM differ depending on the RNA-guided DNA-binding polypeptide used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence).

[0158] The target polynucleotide of a fusion protein of the invention may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides. Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is

directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease (e.g., a causal mutation). The transcribed or translated products may be known or unknown, and further may be at a normal or abnormal level.

[0159] Non-limiting examples of disease-associated genes that can be targeted using the presently disclosed methods and compositions are provided in Table 34. In some embodiments, the disease-associated gene that is targeted are those disclosed in Table 34 having a G>A mutation. Additional examples of disease-associated genes and polynucleotides are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

[0160] In some embodiments, the target polynucleotide comprises a cystic fibrosis transmembrane conductance regulator (5) gene.

[0161] As used herein, the term "cystic fibrosis transmembrane conductance regulator" or "CFTR" refers to a cAMP regulated chloride channel located in the apical membrane of epithelial cells that catalyze the passage of small ions through the membrane. A non-limiting example of a CFTR gene is set forth as SEQ ID NO: 51.

[0162] As used herein, the term "target" or "targets," in relation to a spacer sequence and a target sequence, refers to the localization of an RNA-guided nuclease to a target sequence based on the ability of a spacer sequence within an associated guide RNA to hybridize sufficiently with a target sequence.

[0163] CRISPR RNAs (crRNAs) or nucleic acid molecules encoding the same, wherein the crRNA comprises a spacer sequence that targets a CFTR target sequence are provided. Guide RNAs comprising such crRNAs, one or more nucleic acid molecules encoding a guide RNA comprising such crRNAs, vectors comprising one or more nucleic acid molecules encoding a guide RNA comprising such crRNAs, and systems comprising such crRNAs are also provided. Methods of using such crRNAs or nucleic molecules encoding the same, guide RNAs comprising such crRNAs, one or more nucleic acid molecules encoding a guide RNA comprising such crRNAs, vectors comprising one or more nucleic acid molecules encoding a guide RNA comprising such crRNAs, and systems comprising such crRNAs to bind to, cleave, and/or modulate the target sequence are also provided.

[0164] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, 562, and 563, or the complement thereof. In some embodiments, a single guide RNA (sgRNA) comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0165] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 62-68, 80-85, 116-119, 128-131, 163, 164, 180, 181, 203-209, 219-225, 256-258, 274-276, 310-313, and 330-333, or the complement thereof, and the associated RGN polypeptide has an amino acid sequence

having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 53. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 98-104, 140-143, 197, 198, 235-241, 292-294, and 350-353, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 56.

[0166] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 68-71, 86-89, 120-122, 132-134, 152-156, 169-173, 213-215, 229-231, 251-255, 269-273, 305-309, and 325-329, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 55. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 104-107, 144-146, 186-190, 245-247, 287-291, and 345-349, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 55.

[0167] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 72, 73, 90, 91, 161, 162, 178, 179, 265, 266, 283, and 284, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 52. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 108, 109, 195, 196, 301, and 302, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 52.

[0168] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 74, 75, 92, 93, 123, 124, 135, 136, 167, 184, 216-218, 232-234, 259-261, 277-279, 314-317, and 334-337, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 56. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 56.

93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 110, 111, 147, 148, 201, 248-250, 295-297, and 354-357, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 56.

[0169] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 76, 94, 210-212, 226-228, 322, 342, 562, and 563, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 42. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 112, 242-244, 362, and 564, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 42.

[0170] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 77, 95, 125, 137, 157-160, 174-177, 323, and 343, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 54. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 113, 149, 191-194, and 363, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 54.

[0171] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 78, 96, 126, 138, 168, 185, 267, 285, 318, 319, 338, and 339, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 57. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOS: 114, 150, 202, 303, 358, and 359, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 57.

[0172] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOS: 79, 97, 127, 139, 262-264, 280-282,

324, and 344, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 58. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 115, 151, 298-300, and 364, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 58.

[0173] In some embodiments, the CFTR target sequence of a crRNA or a guide RNA has the sequence set forth in any one of SEQ ID NOs: 165, 166, 182, 183, 268, 286, 320, 321, 340, and 341, or the complement thereof and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 59. In some embodiments, a sgRNA comprising a crRNA having a spacer sequence that targets a CFTR target sequence has at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to any one of SEQ ID NOs: 199, 200, 304, 360, and 361, and the associated RGN polypeptide has an amino acid sequence having at least 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO: 59.

[0174] In some embodiments, the methods comprise contacting a DNA molecule comprising a target DNA sequence with a DNA-binding polypeptide-deaminase fusion protein of the invention, wherein the DNA molecule is contacted with the fusion protein in an amount effective and under conditions suitable for the deamination of a nucleobase. In certain embodiments, the methods comprise contacting a DNA molecule comprising a target DNA sequence with (a) an RGN-deaminase fusion protein of the invention; and (b) a gRNA targeting the fusion protein of (a) to a target nucleotide sequence of the DNA strand; wherein the DNA molecule is contacted with the fusion protein and the gRNA in an amount effective and under conditions suitable for the deamination of a nucleobase. In some embodiments, the target DNA sequence comprises a sequence associated with a disease or disorder, and wherein the deamination of the nucleobase results in a sequence that is not associated with a disease or disorder. In some embodiments, the target DNA sequence resides in an allele of a crop plant, wherein the particular allele of the trait of interest results in a plant of lesser agronomic value. The deamination of the nucleobase results in an allele that improves the trait and increases the agronomic value of the plant.

[0175] In some embodiments, the target DNA sequence comprises a G>A point mutation associated with a disease or disorder, and wherein the deamination of the mutant A base results in a sequence that is not associated with a disease or disorder. In some embodiments, the deamination corrects a point mutation in the sequence associated with the disease or disorder. In some embodiments, the sequence associated with the disease or disorder encodes a protein, and the

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

#### XI. Pharmaceutical Compositions and Methods of Treatment

[0176] Methods of treating a disease in a subject in need thereof are provided herein. The methods comprise administering to a subject in need thereof an effective amount of a presently disclosed fusion protein or a polynucleotide encoding the same, a presently disclosed gRNA or a polynucleotide encoding the same, a presently disclosed fusion protein system, a presently disclosed ribonucleoprotein complex, or a cell modified by or comprising any one of these compositions.

[0177] In some embodiments, the treatment comprises in vivo gene editing by administering to a subject in need thereof a presently disclosed fusion protein, gRNA, or a presently disclosed fusion protein system or polynucleotide(s) encoding the same. In some embodiments, the treatment comprises ex vivo gene editing wherein cells are genetically modified ex vivo with a presently disclosed fusion protein, gRNA, or a presently disclosed fusion protein system or polynucleotide(s) encoding the same and then the modified cells are administered to a subject. In some embodiments, the genetically modified cells originate from the subject that is then administered the modified cells, and the transplanted cells are referred to herein as autologous. In some embodiments, the genetically modified cells originate from a different subject (i.e., donor) within the same species as the subject that is administered the modified cells (i.e., recipient), and the transplanted cells are referred to herein as allogeneic. In some examples described herein, the cells can be expanded in culture prior to administration to a subject in need thereof.

[0178] For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., a genetic defect associated with the CFTR gene, an effective amount of ribonucleoprotein complex comprising a fusion protein with a deaminase having an amino acid sequence that is at least 80% identical to sequence set forth in any one of the SEQ ID NOs: 399, and 405-407. In the embodiments described herein, the administration of the ribonucleoprotein complex corrects the point mutation or introduces a deactivating mutation into a disease-associated CFTR gene. Other diseases that can be treated by correcting a point mutation or introducing a deactivating mutation into a disease-associated gene will be known to those of skill in the art, and the disclosure is not limited in this respect.

[0179] In some embodiments, the disease to be treated with the presently disclosed compositions is one that can be treated with immunotherapy, such as with a chimeric antigen receptor (CAR) T cell. Such diseases include but are not limited to cancer.

[0180] In some embodiments, the deamination of the target nucleobase results in the correction of a genetic

defect, e.g., to correct the CFTR gene, or in the correction of a point mutation that leads to a loss of function in a gene product. In some embodiments, the genetic defect is associated with a disease or disorder, e.g., a lysosomal storage disorder or a metabolic disease, such as, for example, type I diabetes. Thus, in some embodiments, the disease to be treated with the presently disclosed compositions is associated with a sequence (i.e., the sequence is causal for the disease or disorder or causal for symptoms associated with the disease or disorder) that is mutated in order to treat the disease or disorder or the reduction of symptoms associated with the disease or disorder.

[0181] In some embodiments, the disease to be treated with the presently disclosed compositions is associated with a causal mutation. As used herein, a “causal mutation” refers to a particular nucleotide, nucleotides, or nucleotide sequence in the genome that contributes to the severity or presence of a disease or disorder in a subject. The correction of the causal mutation leads to the improvement of at least one symptom resulting from a disease or disorder. In some embodiments, the correction of the causal mutation leads to the improvement of at least one symptom resulting from a disease or disorder. In some embodiments, the causal mutation is adjacent to a PAM site recognized by the RGDBP (e.g., RGN) fused to a deaminase disclosed herein. The causal mutation can be corrected with a fusion polypeptide comprising a RGDBP (e.g., RGN) and a presently disclosed deaminase. Non-limiting examples of diseases associated with a causal mutation include cystic fibrosis, Hurler syndrome, Friedreich's Ataxia, Huntington's Disease, and sickle cell disease. Additional non-limiting examples of disease-associated genes and mutations are available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology

[0182] Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web. In some embodiments, the methods provided herein are used to introduce a deactivating point mutation into a gene or allele that encodes a gene product that is associated with a disease or disorder. For example, in some embodiments, methods are provided herein that employ a fusion protein to introduce a deactivating point mutation into an oncogene (e.g., in the treatment of a proliferative disease). A deactivating mutation may, in some embodiments, generate a premature stop codon in a coding sequence, which results in the expression of a truncated gene product, e.g., a truncated protein lacking the function of the full-length protein. In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. The fusion proteins provided herein can be validated for gene editing-based human therapeutics *in vitro*, e.g., by correcting a disease associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins provided herein, e.g., the fusion proteins comprising an RNA-guided, DNA-binding polypeptide and deaminase polypeptide can be used to correct any single point G>A mutation. Deamination of the mutant A to G leads to a correction of the mutation.

[0183] As used herein, “treatment” or “treating,” or “palliating” or “ameliorating” are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is

meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested. In some embodiments, treatment may be administered after one or more symptoms have developed and/or after a disease has been diagnosed. In particular embodiments, treatment may be administered in the absence of symptoms, e.g., to prevent or delay onset of a symptom or inhibit onset or progression of a disease. For example, treatment may be administered to a susceptible individual prior to the onset of symptoms (e.g., in light of a history of symptoms and/or in light of genetic or other susceptibility factors). Treatment may also be continued after symptoms have resolved, for example, to prevent or delay their prevention or recurrence.

[0184] The term “effective amount” or “therapeutically effective amount” refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, and the delivery system in which it is carried.

[0185] The term “administering” refers to the placement of an active ingredient into a subject, by a method or route that results in at least partial localization of the introduced active ingredient at a desired site, such as a site of injury or repair, such that a desired effect(s) is produced. In some embodiments, the disclosure provides methods comprising delivering any of the isolated polypeptides, nucleic acid molecules fusion proteins, ribonucleoprotein complexes, vectors, pharmaceutical compositions and/or gRNAs described herein. In some embodiments, the disclosure further provides cells produced by such methods, and organisms (such as animals or plants) comprising or produced from such cells. In some embodiments, a deaminase, fusion protein and/or nucleic acid molecules as described herein in combination with (and optionally complexed with) a guide sequence is delivered to a cell.

[0186] In some embodiments, the administering comprises administering by viral delivery. Viral vectors comprising a nucleic acid encoding the fusion proteins, ribonucleoprotein complexes, or vectors disclosed herein may be administered directly to patients (i.e., *in vivo*) or they may be used to treat cells *in vitro*, and the modified cells may optionally be administered to patients (i.e., *ex vivo*). Conventional viral based systems may include, without limitation, retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. In applications where transient expression

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

[0187] In some embodiments, the administering comprises administering by electroporation. In some embodiments, the administering comprises administering by nanoparticle delivery. In some embodiments, the administering comprises administering by liposome delivery. Any effective route of administration can be used to administer an effective amount of a pharmaceutical composition described herein.

[0188] In some embodiments, the administering comprises administering by other non-viral delivery of nucleic acids. Exemplary non-viral delivery methods, without limitation, include RNP complexes, lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipidmucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO1991/17424; WO 1991/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[0189] As used herein, the term "subject" refers to any individual for whom diagnosis, treatment or therapy is desired. In some embodiments, the subject is an animal. In some embodiments, the subject is a mammal. In some embodiments, the subject is a human being.

[0190] The efficacy of a treatment can be determined by the skilled clinician. However, a treatment is considered an "effective treatment," if any one or all of the signs or symptoms of a disease or disorder are altered in a beneficial manner (e.g., decreased by at least 10%), or other clinically accepted symptoms or markers of disease are improved or ameliorated. Efficacy can also be measured by failure of an individual to worsen as assessed by hospitalization or need for medical interventions (e.g., progression of the disease is halted or at least slowed). Methods of measuring these indicators are known to those of skill in the art. Treatment includes: (1) inhibiting the disease, e.g., arresting, or slowing the progression of symptoms; or (2) relieving the disease, e.g., causing regression of symptoms; and (3) preventing or reducing the likelihood of the development of symptoms.

[0191] Pharmaceutical compositions comprising the presently disclosed RGN polypeptides or polynucleotides encoding the same, the presently disclosed gRNAs or polynucleotides encoding the same, the presently disclosed deaminases or polynucleotides encoding the same, the presently disclosed fusion proteins, the presently disclosed systems (such as those comprising a fusion protein), the presently disclosed ribonucleoprotein complex or cells comprising any of the RGN polypeptides or RGN-encoding polynucleotides, gRNA or gRNA-encoding polynucleotides, fusion protein-encoding polynucleotides, or the systems, and a pharmaceutically acceptable carrier are provided.

[0192] As used herein, a "pharmaceutically acceptable carrier" refers to a material that does not cause significant irritation to an organism and does not abrogate the activity and properties of the active ingredient (e.g., a deaminase or fusion protein or nucleic acid molecule encoding the same). Carriers must be of sufficiently high purity and of suffi-

ciently low toxicity to render them suitable for administration to a subject being treated. The carrier can be inert, or it can possess pharmaceutical benefits. In some embodiments, a pharmaceutically acceptable carrier comprises one or more compatible solid or liquid filler, diluents or encapsulating substances which are suitable for administration to a human or other vertebrate animal. In some embodiments, the pharmaceutical composition comprises a pharmaceutically acceptable carrier that is non-naturally occurring. In some embodiments, the pharmaceutically acceptable carrier and the active ingredient are not found together in nature and are thus, heterologous.

[0193] Pharmaceutical compositions used in the presently disclosed methods can be formulated with suitable carriers, excipients, and other agents that provide suitable transfer, delivery, tolerance, and the like. A multitude of appropriate formulations are known to those skilled in the art. See, e.g., Remington, The Science and Practice of Pharmacy (21st ed. 2005). Non-limiting examples include a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. Administered intravenously, particular carriers are physiological saline or phosphate buffered saline (PBS). Pharmaceutical compositions for oral or parenteral use may be prepared into dosage forms in a unit dose suited to fit a dose of the active ingredients. Such dosage forms in a unit dose include, for example, tablets, pills, capsules, injections (ampoules), suppositories, etc. These compositions also may contain adjuvants including preservative agents, wetting agents, emulsifying agents, and dispersing agents. Prevention of the action of microorganisms may be ensured by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, sorbic acid, and the like. It also may be desirable to include isotonic agents, for example, sugars, sodium chloride and the like. Prolonged absorption of the injectable pharmaceutical form may be brought about by the use of agents delaying absorption, for example, aluminum monostearate and gelatin.

[0194] In some embodiments wherein cells comprising or modified with the presently disclosed RGNs, gRNAs, deaminases, fusion proteins, systems (including those comprising fusion proteins) or polynucleotides encoding the same are administered to a subject, the cells are administered as a suspension with a pharmaceutically acceptable carrier. One of skill in the art will recognize that a pharmaceutically acceptable carrier to be used in a cell composition will not include buffers, compounds, cryopreservation agents, preservatives, or other agents in amounts that substantially interfere with the viability of the cells to be delivered to the subject. A formulation comprising cells can include e.g., osmotic buffers that permit cell membrane integrity to be maintained, and optionally, nutrients to maintain cell viability or enhance engraftment upon administration. Such formulations and suspensions are known to those of skill in the art and/or can be adapted for use with the cells described herein using routine experimentation.

[0195] A cell composition can also be emulsified or presented as a liposome composition, provided that the emulsification procedure does not adversely affect cell viability.

The cells and any other active ingredient can be mixed with excipients that are pharmaceutically acceptable and compatible with the active ingredient, and in amounts suitable for use in the therapeutic methods described herein.

[0196] Additional agents included in a cell composition can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids, such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases, such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

[0197] Suitable routes of administrating the pharmaceutical composition described herein include, without limitation: topical, subcutaneous, transdermal, intradermal, intralesional, intraarticular, intraperitoneal, intravesical, transmucosal, gingival, intradental, intracochlear, transtympanic, intraorgan, epidural, intrathecal, intramuscular, intravenous, intravascular, intraosseus, periocular, intratumoral, intracerebral, and intracerebroventricular administration.

[0198] In some embodiments, the pharmaceutical composition described herein is administered locally to a diseased site (e.g., the lung). In some embodiments, the pharmaceutical composition described herein is administered to a subject by injection, inhalation (e.g., of an aerosol), by means of a catheter, by means of a suppository, or by means of an implant, the implant being of a porous, non-porous, or gelatinous material, including a membrane, such as a sialastic membrane, or a fiber. In some embodiments, the pharmaceutical composition is formulated for delivery to a subject, e.g., for gene editing.

[0199] In some embodiments, the pharmaceutical composition is formulated in accordance with routine procedures as a composition adapted for intravenous or subcutaneous administration to a subject, e.g., a human. In some embodiments, pharmaceutical composition for administration by injection are solutions in sterile isotonic aqueous buffer. Where necessary, the pharmaceutical can also include a solubilizing agent and a local anesthetic such as lignocaine to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the quantity of active agent. Where the pharmaceutical is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the pharmaceutical composition is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients can be mixed prior to administration.

[0200] In some embodiments, the pharmaceutical composition can be contained within a lipid particle or vesicle, such as a liposome or microcrystal, which is also suitable for parenteral administration.

[0201] Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for administration to humans, it will be understood by the skilled artisan

that such compositions are generally suitable for administration to animals or organisms of all sorts.

#### Modifying Causal Mutations Using Base-Editing

[0202] An example of a genetically inherited disease which could be corrected using an approach that relies on an RGN-deaminase fusion protein of the invention is Cystic Fibrosis. Cystic fibrosis (CF) is an autosomal recessive disease caused by mutations in the cystic fibrosis transmembrane regulator (CFTR) gene (set forth as SEQ ID NO: 51). CFTR encodes for a cAMP regulated chloride channel located in the apical membrane of epithelial cells that catalyze the passage of small ions through the membrane. Dysregulation of this mechanism causes an impairment of salt and fluid homeostasis that results in multiorgan dysfunctions and ultimately mortality from respiratory failure.

[0203] Almost 2,000 mutations in the CFTR gene have been found to cause CF. CFTR mutations are divided into six classes based on the functional defect in either CFTR protein synthesis, trafficking, function, or stability, although it is acknowledged that many CFTR mutants present multiple defects. Class I mutations lead to severely defective protein production. They are primarily nonsense or frameshift mutations which introduce a premature termination codon (PTC), leading to unstable messenger RNA (mRNA) degraded by the mRNA decay pathway (NMD). Nonsense mutations due to single nucleotide changes comprise a major subset of Class I mutations (Marangi, M. and Pistrutto, G. 2018, *Front Pharmacol* 9, 396, doi: 10.3389/fphar.2018.00396; Pranke, I., et al., 2019, *Front Pharmacol* 10, 121, doi: 10.3389/fphar.2019.00121, both of which are incorporated by reference herein). Treatment for patients with Class I cystic fibrosis can be difficult, as no functional CFTR protein is produced. Notably, a significant fraction of these nonsense mutations are potentially addressable with A to G base editors (Geurts, M. H. et al. 2020, *Cell Stem Cell* 26, 503-510 e507, doi: 10.1016/j.stem.2020.01.019 incorporated by reference herein).

[0204] Geurts et al. were the first group to perform precise base editing in cultured lung epithelial cells with Class I mutations from cystic fibrosis patients, using a fusion protein comprising an adenine deaminase operably linked an RGN, namely either SpyCas9 or the xSpyCas9 variant. SpyCas9 recognizes a 5'-nGG-3'PAM, while the xSpyCas9 variant recognized the reduced 5'-nG-3'. The authors state that a major limitation of the base editing technology is the PAM requirement of the Cas protein being used. They find that the majority of nonsense mutations identified in the CFTR gene are not in the required targeting window for a fusion protein comprising the RGN SpyCas9. The PAM is a short motif, generally one to four nucleotides, on the target DNA sequence that is recognized by the RGN. The PAM sequence is intrinsic to each RGN protein, such that an RGN can only access the genomic space around a suitable PAM. Additionally, the base editing window for base editors is limited, frequently to just a portion of the nucleotides in the target sequence. If the nucleotide of interest is too close to the PAM, the RGN blocks access to the nucleotide. If the nucleotide is too far away from the PAM, the deaminase tethered to the RGN is unable to reach the nucleotide. Also, the amount of ssDNA exposed by the R-loop limits the accessibility of the deaminases. The present invention includes RGN-deaminase fusion proteins where the RGN recognizes a PAM which is proximal to a Class I mutation

of the CFTR gene and the deaminase is able to successfully modify the targeted causal mutation.

[0205] Another limitation to RGN-deaminase fusion proteins known in the art is that the vector construct encoding for the fusion protein is too large for methods of in vivo delivery. AAV delivery of these fusion proteins is not an option for SpyCas9-based fusion proteins because their size exceeds the limit for efficient AAV packaging. The RGN component of the fusion proteins described herein are smaller in size and are therefore viable candidates for AAV vector delivery strategies. The present invention also discloses guide RNAs which are specific for the RGNs described herein and which guide the fusion proteins of the invention to target sites of nonsense mutations in the CFTR gene which were previously unreachable. The present invention also teaches methods of using said fusion proteins for targeted base editing through in vivo AAV vector delivery.

[0206] Ideally, the coding sequence of an RGN-deaminase fusion protein of the invention and a corresponding guide RNA for targeting the fusion protein to the CFTR gene may all be packaged into a single AAV vector. The generally accepted size limit for AAV vectors is 4.7 kb, although larger sizes may be contemplated at the expense of reduced packing efficiency. The RGN nickases in Table 28 have a coding sequence length of about 3.15-3.45 kB. To ensure that the expression cassettes for both the fusion protein and its corresponding guide RNA could fit into an AAV vector, novel, active deletion variants of RGNs are described herein. In addition to shortening the amino acid sequence and therefore the coding sequence of the RGN of the fusion protein, the peptide linker which links the RGN and the deaminase may also be shortened. Finally, the genetic elements, such as the promoters, enhancers, and/or terminators, may also be engineered via deletion analysis to determine the minimal size required for each to be functional.

[0207] Some embodiments of the disclosure provide methods for editing a nucleic acid using the deaminases or the RGN complexes described herein to achieve the nucleobase change, e.g., an A:T base pair to G:C base pair. In some embodiments, the method is a method for editing a nucleobase of a nucleic acid (e.g., a base pair of a double-stranded DNA sequence). In some embodiments, the deaminases or the RGN complexes described herein are used to introduce a point mutation into a nucleic acid by deaminating and excising a target "A" nucleobase. In some embodiments, the deamination-and-excision of the target nucleobase results in the correction of a genetic defect, e.g., in the correction of a point mutation in a CFTR gene. In some embodiments, the genetic defect is associated with a disease, disorder, or condition, e.g., Cystic Fibrosis. For example, in some embodiments, methods are provided herein employ a base editing RGN complexes comprising a fusion protein with a deaminase having an amino acid sequence that is at least 80% identical to sequence set forth in any one of the SEQ ID NOS: 399, and 405-407, to correct a gene associated with a genetic defect, e.g., to correct a point mutation in a CFTR gene (e.g., in the treatment of a proliferative disease). In specific embodiments, the target sequence in the CFTR gene is 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, or 563.

[0208] In some embodiments, the purpose of the methods provided herein is to restore the function of a dysfunctional gene via genome editing. The base editor proteins provided herein may be validated for gene editing-based human

therapeutics in vitro, e.g., by correcting a disease-associated mutation in human cell culture. It will be understood by the skilled artisan that the fusion proteins and/or the RGN complexes provided herein comprising a nucleic acid binding protein (e.g., nCas9) and a nucleobase modification domain (e.g., deaminase with an amino acid sequence set forth in SEQ ID NO.: 407, 399, or 405 may be used to correct any single point of T to G or change a pairing of T: A to G:C.

[0209] In some embodiments, provided herein are the methods for the treatment of a subject diagnosed with a disease associated with or caused by a point mutation (e.g., mutation in CFTR gene) that can be corrected by a fusion protein or the RGN complexes described herein. For example, in some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., cystic fibrosis, an effective amount of a fusion protein or RGN complex disclosed herein that corrects the point mutation or introduces a deactivating mutation into a disease-associated gene. In some embodiments, a method is provided that comprises administering to a subject having such a disease, e.g., a cancer associated with a point mutation as described above, an effective amount of a fusion protein, RGN complex, or pharmaceutical composition disclosed herein that corrects the point mutation or introduces a deactivating mutation into a disease-associated gene. In specific embodiments, methods of treating cystic fibrosis are provided along with methods of reducing at least one symptom of cystic fibrosis by administering an effective amount of a pharmaceutical composition disclosed herein. An effective amount of a pharmaceutical composition for treating or reducing a symptom of cystic fibrosis can reduce a symptom (i.e., treat) of cystic fibrosis by about 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or more; or about 10-20%, 15-25%, 20-40%, 30-50%, 40-60%, 50-70%, 60-80%, 70-90%, 80-95%, or 90-95% when compared to a control patient. In specific embodiments, the control patient can be the same patient before administration of the effective amount of the pharmaceutical composition disclosed herein. Symptoms of cystic fibrosis can include, but are not limited to: sneezing, a persistent cough that produces mucus or phlegm, shortness of breath, especially when exercising, recurrent lung infections, a stuffy nose, stuffy sinuses, greasy foul-smelling stools, constipation, nausea, swollen abdomen, loss of appetite, among others. Methods of identifying and measuring symptoms of cystic fibrosis are known in the art.

[0210] In some embodiments of the described methods for modifying a target DNA molecule, the step of contacting is performed in vitro. In particular embodiments, the step of contacting is performed in vivo. In some embodiments, the step of contacting is performed in a subject (e.g., a human subject or a non-human animal subject). In some embodiments, the step of contacting is performed in a cell, such as a human or non-human animal cell.

## XII. Cells Comprising a Polynucleotide Genetic Modification

[0211] Provided herein are cells and organisms comprising a target nucleic acid molecule of interest that has been modified using a process mediated by a fusion protein, optionally with a gRNA, as described herein. In some embodiments, the fusion protein comprises a deaminase polypeptide comprising an amino acid sequence of any of

SEQ ID NOS: 1-10 and 399-441, or an active variant or fragment thereof. In some embodiments, the fusion protein comprises an adenine deaminase comprising an amino acid sequence having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to any of SEQ ID NOS: 1-10 and 399-441. In some embodiments, the fusion protein comprises a deaminase and a DNA-binding polypeptide (e.g., an RNA-guided, DNA-binding polypeptide). In further embodiments, the fusion protein comprises a deaminase and an RGN or a variant thereof, such as for example APG07433.1 (SEQ ID NO: 41) or its nickase variant nAPG07433.1 (SEQ ID NO: 42). In some embodiments, the fusion protein comprises a deaminase and a Cas9 or a variant thereof, such as for example dCas9 or nickase Cas9. In some embodiments, the fusion protein comprises a nuclease-inactive or nickase variant of a Type II CRISPR-Cas polypeptide. In some embodiments, the fusion protein comprises a nuclease-inactive or nickase variant of a Type V CRISPR-Cas polypeptide. In some embodiments, the fusion protein comprises a nuclease-inactive or nickase variant of a Type VI CRISPR-Cas polypeptide.

[0212] The modified cells can be eukaryotic (e.g., mammalian, plant, insect, avian cell) or prokaryotic. Also provided are organelles and embryos comprising at least one nucleotide sequence that has been modified by a process utilizing a fusion protein as described herein. The genetically modified cells, organisms, organelles, and embryos can be heterozygous or homozygous for the modified nucleotide sequence. The mutation(s) introduced by the deaminase domain of the fusion protein can result in altered expression (up-regulation or down-regulation), inactivation, or the expression of an altered protein product or an integrated sequence. In those instances wherein the mutation(s) results in either the inactivation of a gene or the expression of a non-functional protein product, the genetically modified cell, organism, organelle, or embryo is referred to as a "knock out". The knock out phenotype can be the result of a deletion mutation (i.e., deletion of at least one nucleotide), an insertion mutation (i.e., insertion of at least one nucleotide), or a nonsense mutation (i.e., substitution of at least one nucleotide such that a stop codon is introduced).

[0213] In some embodiments, the mutation(s) introduced by the deaminase domain of the fusion protein results in the production of a variant protein product. The expressed variant protein product can have at least one amino acid substitution and/or the addition or deletion of at least one amino acid. The variant protein product can exhibit modified characteristics or activities when compared to the wild-type protein, including but not limited to altered enzymatic activity or substrate specificity.

[0214] In some embodiments, the mutation(s) introduced by the deaminase domain of the fusion protein result in an altered expression pattern of a protein. As a non-limiting example, mutation(s) in the regulatory regions controlling the expression of a protein product can result in the over-expression or downregulation of the protein product or an altered tissue or temporal expression pattern.

[0215] The cells that have been modified can be grown into an organism, such as a plant, in accordance with conventional ways. See, for example, McCormick et al. (1986) *Plant Cell Reports* 5:81-84. These plants may then be grown, and either pollinated with the same modified strain

or different strains, and the resulting hybrid having the genetic modification. The present invention provides genetically modified seed. Progeny, variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the genetic modification. Further provided is a processed plant product or byproduct that retains the genetic modification, including for example, soymeal.

[0216] The methods provided herein may be used for modification of any plant species, including, but not limited to, monocots and dicots. Examples of plants of interest include, but are not limited to, corn (maize), sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, and oilseed rape, *Brassica* sp., alfalfa, rye, millet, safflower, peanuts, sweet potato, cassava, coffee, coconut, pineapple, citrus trees, cocoa, tea, banana, avocado, fig, guava, mango, olive, *papaya*, cashew, macadamia, almond, oats, vegetables, ornamentals, and conifers.

[0217] Vegetables include, but are not limited to, tomatoes, lettuce, green beans, lima beans, peas, and members of the genus *Cucumis* such as cucumber, cantaloupe, and musk melon. Ornamentals include, but are not limited to, azalea, *hydrangea*, hibiscus, roses, tulips, daffodils, petunias, carnation, poinsettia, and *chrysanthemum*. Preferably, plants of the present invention are crop plants (for example, maize, sorghum, wheat, sunflower, tomato, crucifers, peppers, potato, cotton, rice, soybean, sugarbeet, sugarcane, tobacco, barley, oilseed rape, etc.).

[0218] The methods provided herein can also be used to genetically modify any prokaryotic species, including but not limited to, archaea and bacteria (e.g., *Bacillus* sp., *Klebsiella* sp., *Streptomyces* sp., *Rhizobium* sp., *Escherichia* sp., *Pseudomonas* sp., *Salmonella* sp., *Shigella* sp., *Vibrio* sp., *Yersinia* sp., *Mycoplasma* sp., *Agrobacterium*, *Lactobacillus* sp.).

[0219] The methods provided herein can be used to genetically modify any eukaryotic species or cells therefrom, including but not limited to animals (e.g., mammals, insects, fish, birds, and reptiles), fungi, amoeba, algae, and yeast. In some embodiments, the cell that is modified by the presently disclosed methods include cells of hematopoietic origin, such as immune cells (i.e., a cell of the innate or adaptive immune system) including but not limited to B cells, T cells, natural killer (NK) cells, pluripotent stem cells, induced pluripotent stem cells, chimeric antigen receptor T (CAR-T) cells, monocytes, macrophages, and dendritic cells.

[0220] Cells that have been modified may be introduced into an organism. These cells could have originated from the same organism (e.g., person) in the case of autologous cellular transplants, wherein the cells are modified in an ex vivo approach. In some embodiments, the cells originated from another organism within the same species (e.g., another person) in the case of allogeneic cellular transplants.

### XIII. Kits

[0221] Some aspects of this disclosure provide kits comprising a deaminase of the invention. In certain embodiments, the disclosure provides kits comprising a fusion protein comprising a deaminase of the invention and a DNA-binding polypeptide (e.g., an RNA-guided, DNA-binding polypeptide, such as an RGN polypeptide, for example a nuclease-inactive Cas9 domain), and, optionally, a linker positioned between the DNA-binding polypeptide

domain and the deaminase. In addition, in some embodiments, the kit comprises suitable reagents, buffers, and/or instructions for using the fusion protein, e.g., for in vitro or in vivo DNA or RNA editing. In some embodiments, the kit comprises instructions regarding the design and use of suitable gRNAs for targeted editing of a nucleic acid sequence.

[0222] In some embodiments, the pharmaceutical composition may be provided as a pharmaceutical kit comprising (a) a container containing a composition of the disclosure in lyophilized form and (b) a second container containing a pharmaceutically acceptable diluent (e.g., sterile water) for injection. The pharmaceutically acceptable diluent can be used for reconstitution or dilution of the lyophilized compound of the disclosure. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration.

[0223] The article "a" and "an" are used herein to refer to one or more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "a polypeptide" means one or more polypeptides.

[0224] All publications and patent applications mentioned in the specification are indicative of the level of those skilled in the art to which this disclosure pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated herein by reference.

[0225] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

#### Non-Limiting Embodiments Include:

[0226] 1. An isolated polypeptide comprising an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441, wherein said polypeptide has deaminase activity.

[0227] 2. The isolated polypeptide of embodiment 1, comprising an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0228] 3. The isolated polypeptide of embodiment 1, comprising an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0229] 4. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein the deaminase is encoded by a nucleotide sequence that:

[0230] a) has at least 80% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485, or

[0231] b) encodes an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0232] 5. The nucleic acid molecule of embodiment 4, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0233] 6. The nucleic acid molecule of embodiment 4, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0234] 7. The nucleic acid molecule of embodiment 4, wherein the deaminase is encoded by a nucleotide sequence that has 100% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0235] 8. The nucleic acid molecule of embodiment 4, wherein the deaminase polypeptide has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0236] 9. The nucleic acid molecule of embodiment 4, wherein the deaminase polypeptide has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0237] 10. The nucleic acid molecule of any one of embodiments 4-9, wherein said nucleic acid molecule further comprises a heterologous promoter operably linked to said polynucleotide.

[0238] 11. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the polypeptide of any one of embodiments 1-3 or the nucleic acid molecule of any one of embodiments 4-10.

[0239] 12. The pharmaceutical composition of embodiment 11, wherein the pharmaceutically acceptable carrier is heterologous to said polypeptide or said nucleic acid molecule.

[0240] 13. The pharmaceutical composition of embodiment 11 or 12, wherein the pharmaceutically acceptable carrier is not naturally-occurring.

[0241] 14. A fusion protein comprising a DNA-binding polypeptide and a deaminase having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0242] 15. The fusion protein of embodiment 14, wherein said deaminase has at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0243] 16. The fusion protein of embodiment 14, wherein said deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0244] 17. The fusion protein of any one of embodiments 14-16, wherein the deaminase is an adenine deaminase.

[0245] 18. The fusion protein of any one of embodiments 14-17, wherein the DNA-binding polypeptide is a meganuclease, zinc finger fusion protein, or a TALEN.

[0246] 19. The fusion protein of any one of embodiments 14-17, wherein the DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide.

[0247] 20. The fusion protein of embodiment 19, wherein the RNA-guided, DNA-binding polypeptide is an RNA-guided nuclease (RGN) polypeptide.

[0248] 21. The fusion protein of embodiment 20, wherein the RGN is a Type II CRISPR-Cas polypeptide.

[0249] 22. The fusion protein of embodiment 20, wherein the RGN is a Type V CRISPR-Cas polypeptide.

[0250] 23. The fusion protein of any one of embodiments 20-22, wherein the RGN is an RGN nickase.

[0251] 24. The fusion protein of embodiment 20, wherein the RGN has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 41, 60, 366, and 368.

[0252] 25. The fusion protein of embodiment 20, wherein the RGN has an amino acid sequence of any one of SEQ ID NOS: 41, 60, 366, and 368.

[0253] 26. The fusion protein of embodiment 23, wherein the RGN nickase is any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0254] 27. The fusion protein of any of embodiments 14-26, wherein the fusion protein further comprises at least one nuclear localization signal (NLS).

[0255] 28. A nucleic acid molecule comprising a polynucleotide encoding a fusion protein comprising a DNA-binding polypeptide and a deaminase, wherein the deaminase is encoded by a nucleotide sequence that:

[0256] a) has at least 80% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485, or

[0257] b) encodes an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0258] 29. The nucleic acid molecule of embodiment 28, wherein said nucleotide sequence has at least 90% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0259] 30. The nucleic acid molecule of embodiment 28, wherein said nucleotide sequence has at least 95% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0260] 31. The nucleic acid molecule of embodiment 28, wherein said nucleotide sequence has 100% sequence identity to any one of SEQ ID NOS: 451, 449, 443, 11-20, 444-448, 450, and 452-485.

[0261] 32. The nucleic acid molecule of embodiment 28, wherein said nucleotide sequence encodes an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0262] 33. The nucleic acid molecule of embodiment 28, wherein said nucleotide sequence encodes an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0263] 34. The nucleic acid molecule of any one of embodiments 28-33, wherein the deaminase is an adenine deaminase.

[0264] 35. The nucleic acid molecule of any one of embodiments 28-34, wherein the DNA-binding polypeptide is a meganuclease, zinc finger fusion protein, or a TALEN.

[0265] 36. The nucleic acid molecule of any one of embodiments 28-34-, wherein the DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide.

[0266] 37. The nucleic acid molecule of embodiment 36, wherein the RNA-guided, DNA-binding polypeptide is an RNA-guided nuclease (RGN) polypeptide.

[0267] 38. The nucleic acid molecule of embodiment 37, wherein the RGN is a Type II CRISPR-Cas polypeptide.

[0268] 39. The nucleic acid molecule of embodiment 37, wherein the RGN is a Type V CRISPR-Cas polypeptide.

[0269] 40. The nucleic acid molecule of any one of embodiments 37-39, wherein the RGN is an RGN nickase.

[0270] 41. The nucleic acid molecule of embodiment 37, wherein the RGN has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 41, 60, 366, and 368.

[0271] 42. The nucleic acid molecule of embodiment 37, wherein the RGN is SEQ ID NO: 41, 60, 366, or 368.

[0272] 43. The nucleic acid molecule of embodiment 40, wherein the RGN nickase is any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0273] 44. The nucleic acid molecule of any of embodiments 28-43, wherein the polynucleotide encoding the fusion protein is operably linked at its 5' end to a heterologous promoter.

[0274] 45. The nucleic acid molecule of any of embodiments 28-44, wherein the polynucleotide encoding the fusion protein is operably linked at its 3' end to a heterologous terminator.

[0275] 46. The nucleic acid molecule of any of embodiments 28-45, wherein the fusion protein comprises one or more nuclear localization signals.

[0276] 47. The nucleic acid molecule of any of embodiments 28-46, wherein the fusion protein is codon optimized for expression in a eukaryotic cell.

[0277] 48. The nucleic acid molecule of any of embodiments 28-46, wherein the fusion protein is codon optimized for expression in a prokaryotic cell.

[0278] 49. A vector comprising the nucleic acid molecule of any one of embodiments 28-48.

[0279] 50. A vector comprising the nucleic acid molecule of any one of embodiments 28-48, further comprising at least one nucleotide sequence encoding a guide RNA (gRNA) capable of hybridizing to a target sequence.

[0280] 51. The vector of embodiment 50, wherein the gRNA is a single guide RNA.

[0281] 52. The vector of embodiment 50, wherein the gRNA is a dual guide RNA.

[0282] 53. A cell comprising the fusion protein of any of embodiments 14-27.

[0283] 54. A cell comprising the fusion protein of any one of embodiments 14-27, wherein the cell further comprises a guide RNA.

[0284] 55. A cell comprising the nucleic acid molecule of any one of embodiments 28-48.

[0285] 56. A cell comprising the vector of embodiments of any one of embodiments 49-52.

[0286] 57. The cell of any one of embodiments 53-56, wherein the cell is a prokaryotic cell.

[0287] 58. The cell of any one of embodiments 53-56, wherein the cell is a eukaryotic cell.

[0288] 59. The cell of embodiment 58, wherein the eukaryotic cell is a mammalian cell.

[0289] 60. The cell of embodiment 59, wherein the mammalian cell is a human cell.

[0290] 61. The cell of embodiment 60, wherein the human cell is an immune cell.

[0291] 62. The cell of embodiment 61, wherein the immune cell is a stem cell.

[0292] 63. The cell of embodiment 62, wherein the stem cell is an induced pluripotent stem cell.

[0293] 64. The cell of embodiment 58, wherein the eukaryotic cell is an insect or avian cell.

[0294] 65. The cell of embodiment 58, wherein the eukaryotic cell is a fungal cell.

- [0295] 66. The cell of embodiment 58, wherein the eukaryotic cell is a plant cell.
- [0296] 67. A plant comprising the cell of embodiment 66.
- [0297] 68. A seed comprising the cell of embodiment 66.
- [0298] 69. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the fusion protein of any one of embodiments 14-27, the nucleic acid molecule of any one of embodiments 28-48, the vector of any one of embodiments 49-52, or the cell of any one of embodiments 59-63.
- [0299] 70. A method for making a fusion protein comprising culturing the cell of any one of embodiments 53-66 under conditions in which the fusion protein is expressed.
- [0300] 71. A method for making a fusion protein comprising introducing into a cell the nucleic acid molecule of any of embodiments 28-48 or a vector of any one of embodiments 49-52 and culturing the cell under conditions in which the fusion protein is expressed.
- [0301] 72. The method of embodiment 70 or 71, further comprising purifying said fusion protein.
- [0302] 73. A method for making an RGN fusion ribonucleoprotein complex, comprising introducing into a cell the nucleic acid molecule of any one of embodiments 37-43 and a nucleic acid molecule comprising an expression cassette encoding for a guide RNA, or the vector of any of embodiments 50-52, and culturing the cell under conditions in which the fusion protein and the gRNA are expressed and form an RGN fusion ribonucleoprotein complex.
- [0303] 74. The method of embodiment 73, further comprising purifying said RGN fusion ribonucleoprotein complex.
- [0304] 75. A system for modifying a target DNA molecule comprising a target DNA sequence, said system comprising:
- [0305] a) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN) and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441, or a nucleotide sequence encoding said fusion protein; and
- [0306] b) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and
- [0307] wherein the one or more guide RNAs are capable of forming a complex with the fusion protein in order to direct said fusion protein to bind to said target DNA sequence and modify the target DNA molecule.
- [0308] 76. The system of embodiment 75, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0309] 77. The system of embodiment 75, wherein said deaminase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0310] 78. The system of any one of embodiments 75-77, wherein at least one of said nucleotide sequence encoding the one or more guide RNAs and said nucleotide sequence encoding the fusion protein is operably linked to a promoter heterologous to said nucleotide sequence.
- [0311] 79. The system of any one of embodiments 75-78, wherein the target DNA sequence is a eukaryotic target DNA sequence.
- [0312] 80. The system of any one of embodiments 75-79, wherein the target DNA sequence is located adjacent to a protospacer adjacent motif (PAM) that is recognized by the RGN.
- [0313] 81. The system of any one of embodiments 75-80, wherein the target DNA molecule is within a cell.
- [0314] 82. The system of embodiment 81, wherein the cell is a eukaryotic cell.
- [0315] 83. The system of embodiment 82, wherein the eukaryotic cell is a plant cell.
- [0316] 84. The system of embodiment 82, wherein the eukaryotic cell is a mammalian cell.
- [0317] 85. The system of embodiment 84, wherein the mammalian cell is a human cell.
- [0318] 86. The system of embodiment 85, wherein the human cell is an immune cell.
- [0319] 87. The system of embodiment 86, wherein the immune cell is a stem cell.
- [0320] 88. The system of embodiment 87, wherein the stem cell is an induced pluripotent stem cell.
- [0321] 89. The system of embodiment 82, wherein the eukaryotic cell is an insect cell.
- [0322] 90. The system of embodiment 81, wherein the cell is a prokaryotic cell.
- [0323] 91. The system of any one of embodiments 75-90, wherein the RGN of the fusion protein is a Type II CRISPR-Cas polypeptide.
- [0324] 92. The system of any one of embodiments 75-90, wherein the RGN of the fusion protein is a Type V CRISPR-Cas polypeptide.
- [0325] 93. The system of any one of embodiments 75-90, wherein the RGN of the fusion protein has an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 41, 60, 366, or 368.
- [0326] 94. The system of any one of embodiments 75-90, wherein the RGN of the fusion protein has an amino acid sequence of any one of SEQ ID NOS: 41, 60, 366, and 368.
- [0327] 95. The system of any one of embodiments 75-90, wherein the RGN of the fusion protein is an RGN nickase.
- [0328] 96. The system of embodiment 95, wherein the RGN nickase is any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.
- [0329] 97. The system of any of embodiments 75-96, wherein the fusion protein comprises one or more nuclear localization signals.
- [0330] 98. The system of any of embodiments 75-97, wherein the fusion protein is codon optimized for expression in a eukaryotic cell.
- [0331] 99. The system of any of embodiments 75-98, wherein nucleotide sequences encoding the one or more guide RNAs and the nucleotide sequence encoding a fusion protein are located on one vector.
- [0332] 100. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the system of any one of embodiments 75-99.
- [0333] 101. A method for modifying a target DNA molecule comprising a target DNA sequence, said method comprising delivering a system according to any one of embodiments 75-99 to said target DNA molecule or a cell comprising the target DNA molecule.
- [0334] 102. The method of embodiment 101, wherein said modified target DNA molecule comprises an A>N mutation of at least one nucleotide within the target DNA molecule, wherein N is C, G, or T.

[0335] 103. The method of embodiment 102, wherein said modified target DNA molecule comprises an A>G mutation of at least one nucleotide within the target DNA molecule.

[0336] 104. A method for modifying a target DNA molecule comprising a target sequence comprising:

[0337] a) assembling an RGN-deaminase ribonucleotide complex in vitro by combining:

[0338] i) one or more guide RNAs capable of hybridizing to the target DNA sequence; and

[0339] ii) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN), and at least one deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 407, 405, 399, 1-10, 400-404, 406, and 408-441;

[0340] under conditions suitable for formation of the RGN-deaminase ribonucleotide complex; and

[0341] b) contacting said target DNA molecule or a cell comprising said target DNA molecule with the in vitro-assembled RGN-deaminase ribonucleotide complex;

[0342] wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said fusion protein to bind to said target DNA sequence and modification of the target DNA molecule occurs.

[0343] 105. The method of embodiment 104, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0344] 106. The method of embodiment 104, wherein said deaminase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOs: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0345] 107. The method of any one of embodiments 104-106, wherein said modified target DNA molecule comprises an A>N mutation of at least one nucleotide within the target DNA molecule, wherein N is C, G, or T.

[0346] 108. The method of embodiment 107, wherein said modified target DNA molecule comprises an A>G mutation of at least one nucleotide within the target DNA molecule.

[0347] 109. The method of any one of embodiments 104-108, wherein the RGN of the fusion protein is a Type II CRISPR-Cas polypeptide.

[0348] 110. The method of any of embodiments 104-108, wherein the RGN of the fusion protein is a Type V CRISPR-Cas polypeptide.

[0349] 111. The method of any of embodiments 104-108, wherein the RGN of the fusion protein has an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 41, 60, 366, or 368.

[0350] 112. The method of any one of embodiments 104-108, wherein the RGN of the fusion protein has an amino acid sequence of any one of SEQ ID NOs: 41, 60, 366, and 368.

[0351] 113. The method of any of embodiments 104-108, wherein the RGN of the fusion protein is an RGN nickase.

[0352] 114. The method of embodiment 113, wherein the RGN nickase is any one of SEQ ID NOs: 42, 52-59, 61, 397, and 398.

[0353] 115. The method of any of embodiments 104-114, wherein the fusion protein comprises one or more nuclear localization signals.

[0354] 116. The method of any of embodiments 104-115, wherein the fusion protein is codon optimized for expression in a eukaryotic cell.

[0355] 117. The method of any one of embodiments 104-116, wherein said target DNA sequence is a eukaryotic target DNA sequence.

[0356] 118. The method of any of embodiments 104-117, wherein said target DNA sequence is located adjacent to a protospacer adjacent motif (PAM).

[0357] 119. The method of any of embodiments 104-118, wherein the target DNA molecule is within a cell.

[0358] 120. The method of embodiment 119, wherein the cell is a eukaryotic cell.

[0359] 121. The method of embodiment 120, wherein the eukaryotic cell is a plant cell.

[0360] 122. The method of embodiment 120, wherein the eukaryotic cell is a mammalian cell.

[0361] 123. The method of embodiment 122, wherein the mammalian cell is a human cell.

[0362] 124. The method of embodiment 123, wherein the human cell is an immune cell.

[0363] 125. The method of embodiment 124, wherein the immune cell is a stem cell.

[0364] 126. The method of embodiment 125, wherein the stem cell is an induced pluripotent stem cell.

[0365] 127. The method of embodiment 120, wherein the eukaryotic cell is an insect cell.

[0366] 128. The method of embodiment 119, wherein the cell is a prokaryotic cell.

[0367] 129. The method of any one of embodiments 119-128, further comprising selecting a cell comprising said modified DNA molecule.

[0368] 130. A cell comprising a modified target DNA sequence according to the method of embodiment 129.

[0369] 131. The cell of embodiment 130, wherein the cell is a eukaryotic cell.

[0370] 132. The cell of embodiment 131, wherein the eukaryotic cell is a plant cell.

[0371] 133. A plant comprising the cell of embodiment 132.

[0372] 134. A seed comprising the cell of embodiment 132.

[0373] 135. The cell of embodiment 131, wherein the eukaryotic cell is a mammalian cell.

[0374] 136. The cell of embodiment 135, wherein the mammalian cell is a human cell.

[0375] 137. The cell of embodiment 136, wherein the human cell is an immune cell.

[0376] 138. The cell of embodiment 137, wherein the immune cell is a stem cell.

[0377] 139. The cell of embodiment 138, wherein the stem cell is an induced pluripotent stem cell.

[0378] 140. The cell of embodiment 131, wherein the eukaryotic cell is an insect cell.

[0379] 141. The cell of embodiment 130, wherein the cell is a prokaryotic cell.

[0380] 142. A pharmaceutical composition comprising the cell of any one of embodiments 135-139, and a pharmaceutically acceptable carrier.

[0381] 143. A method for producing a genetically modified cell with a correction in a causal mutation for a genetically inherited disease, the method comprising introducing into the cell:

- [0382] a) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN) and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441, or a polynucleotide encoding said fusion protein, wherein said polynucleotide encoding the fusion protein is operably linked to a promoter to enable expression of the fusion protein in the cell; and
- [0383] b) one or more guide RNAs (gRNA) capable of hybridizing to a target DNA sequence, or a polynucleotide encoding said gRNA, wherein said polynucleotide encoding the gRNA is operably linked to a promoter to enable expression of the gRNA in the cell; whereby the fusion protein and gRNA target to the genomic location of the causal mutation and modify the genomic sequence to remove the causal mutation.
- [0384] 144. The method of embodiment 143, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0385] 145. The method of embodiment 143, wherein said deaminase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0386] 146. The method of any one of embodiments 143-145, wherein said RGN of the fusion protein is an RGN nickase.
- [0387] 147. The method of embodiment 146, wherein the RGN nickase is any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.
- [0388] 148. The method of any one of embodiments 143-147, wherein the genome modification comprises introducing an A>G mutation of at least one nucleotide within the target DNA sequence.
- [0389] 149. The method of any of embodiments 143-148, wherein the cell is an animal cell.
- [0390] 150. The method of embodiment 149, wherein the animal cell is a mammalian cell.
- [0391] 151. The method of embodiment 150, wherein the cell is derived from a dog, cat, mouse, rat, rabbit, horse, sheep, goat, cow, pig, or human.
- [0392] 152. The method of any one of embodiments 143-151, wherein the correction of the causal mutation comprises correcting a nonsense mutation.
- [0393] 153. The method of embodiment 149, wherein the genetically inherited disease is a disease listed in Table 34.
- [0394] 154. The method of embodiment 149, wherein the genetically inherited disease is cystic fibrosis.
- [0395] 155. The method of embodiment 154, wherein the gRNA further comprises a spacer sequence that targets any one of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.
- [0396] 156. The method of embodiment 155, wherein the gRNA comprises any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.
- [0397] 157. A CRISPR RNA (crRNA) or a nucleic acid molecule encoding the same, wherein said CRISPR RNA comprises a spacer sequence that targets a target DNA sequence within a cystic fibrosis transmembrane conductance regulator (CFTR) gene, wherein said target sequence has the sequence set forth as any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, 562, and 563, or the complement thereof.
- [0398] 158. A guide RNA comprising the crRNA of embodiment 157.
- [0399] 159. The guide RNA of embodiment 158, wherein said guide RNA is a dual-guide RNA.
- [0400] 160. The guide RNA of embodiment 158, wherein said guide RNA is a single guide RNA (sgRNA).
- [0401] 161. The guide RNA of embodiment 160, wherein said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.
- [0402] 162. The guide RNA of embodiment 160, wherein said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.
- [0403] 163. The guide RNA of embodiment 160, wherein said sgRNA has the sequence set forth as any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.
- [0404] 164. A vector comprising one or more nucleic acid molecules encoding said guide RNA of any one of embodiments 158-163.
- [0405] 165. A system for binding a target DNA sequence of a DNA molecule, said system comprising:
- [0406] a) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and
- [0407] b) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN) and an adenine deaminase, or a polynucleotide comprising a nucleotide sequence encoding the fusion protein;
- [0408] wherein the one or more guide RNAs are capable of hybridizing to the target DNA sequence,
- [0409] wherein the one or more guide RNAs are capable of forming a complex with the RGN polypeptide in order to direct said RGN polypeptide to bind to said target DNA sequence of the DNA molecule, and
- [0410] wherein at least one guide RNA comprises a CRISPR RNA (crRNA) comprising a spacer sequence that targets a target DNA sequence within a cystic fibrosis transmembrane conductance regulator (CFTR) gene, wherein said target sequence has the sequence set forth as any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, 562, and 563, or the complement thereof.
- [0411] 166. The system of embodiment 165, wherein at least one of said nucleotide sequences encoding the one or more guide RNAs and said nucleotide sequence encoding the fusion protein is operably linked to a promoter heterologous to said nucleotide sequence.
- [0412] 167. A system for binding a target DNA sequence of a DNA molecule, said system comprising:
- [0413] a) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more polynucleotides comprising one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and
- [0414] b) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN) and an adenine deaminase;
- [0415] wherein the one or more guide RNAs are capable of hybridizing to the target DNA sequence,

[0416] wherein the one or more guide RNAs are capable of forming a complex with the RGN polypeptide in order to direct said RGN polypeptide to bind to said target DNA sequence of the DNA molecule, and

[0417] wherein at least one guide RNA comprises a CRISPR RNA (crRNA) comprising a spacer sequence that targets a target DNA sequence within a cystic fibrosis transmembrane conductance regulator (CFTR) gene, wherein said target sequence has the sequence set forth as any one of SEQ ID NOs: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, 562, and 563, or the complement thereof.

[0418] 168. The system of embodiment 167, wherein at least one of said nucleotide sequences encoding the one or more guide RNAs is operably linked to a promoter heterologous to said nucleotide sequence.

[0419] 169. The system of any one of embodiments 165-168, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOs: 1-10 and 399-441.

[0420] 170. The system of any one of embodiments 165-168, wherein the deaminase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOs: 1-10 and 399-441.

[0421] 171. The system of any one of embodiments 165-168, wherein the deaminase has an amino acid sequence having the sequence set forth in any one of SEQ ID NOs: 1-10 and 399-441.

[0422] 172. The system of any one of embodiments 165-171, wherein said RGN polypeptide and said one or more guide RNAs are not found complexed to one another in nature.

[0423] 173. The system of any one of embodiments 165-172, wherein:

[0424] a) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 62-68, 80-85, 116-119, 128-131, 163, 164, 180, 181, 203-209, 219-225, 256-258, 274-276, 310-313, and 330-333, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 53;

[0425] b) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 68-71, 86-89, 120-122, 132-134, 152-156, 169-173, 213-215, 229-231, 251-255, 269-273, 305-309, and 325-329, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 55;

[0426] c) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 72, 73, 90, 91, 161, 162, 178, 179, 265, 266, 283, and 284 or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 52;

[0427] d) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 74, 75, 92, 93, 123, 124, 135, 136, 167, 184, 216-218, 232-234, 259-261, 277-279, 314-317, and 334-337, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 56;

[0428] e) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 76, 94, 210-212, 226-228, 322, 342, 562, and 563, or the complement

thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 42;

[0429] f) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 77, 95, 125, 137, 157-160, 174-177, 323, and 343, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 54;

[0430] g) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 78, 96, 126, 138, 168, 185, 267, 285, 318, 319, 338, and 339, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 57;

[0431] h) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 79, 97, 127, 139, 262-264, 280-282, 324, and 344, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 58; and

[0432] i) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 165, 166, 182, 183, 268, 286, 320, 321, 340, and 341, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 59.

[0433] 174. The system of any one of embodiments 165-172, wherein:

[0434] a) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 62-68, 80-85, 116-119, 128-131, 163, 164, 180, 181, 203-209, 219-225, 256-258, 274-276, 310-313, and 330-333, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 53;

[0435] b) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 68-71, 86-89, 120-122, 132-134, 152-156, 169-173, 213-215, 229-231, 251-255, 269-273, 305-309, and 325-329, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 55;

[0436] c) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 72, 73, 90, 91, 161, 162, 178, 179, 265, 266, 283, and 284 or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 52;

[0437] d) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 74, 75, 92, 93, 123, 124, 135, 136, 167, 184, 216-218, 232-234, 259-261, 277-279, 314-317, and 334-337, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 56;

[0438] e) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 76, 94, 210-212, 226-228, 322, 342, 562, and 563, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 42;

[0439] f) said target DNA sequence has the sequence set forth as any one of SEQ ID NOs: 77, 95, 125, 137,

157-160, 174-177, 323, and 343, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 54;

[0440] g) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 78, 96, 126, 138, 168, 185, 267, 285, 318, 319, 338, and 339, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 57;

[0441] h) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 79, 97, 127, 139, 262-264, 280-282, 324, and 344, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 58; and

[0442] i) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 165, 166, 182, 183, 268, 286, 320, 321, 340, and 341, or the complement thereof, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 59.

[0443] 175. The system of any one of embodiments 165-172, wherein:

[0444] a) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 62-68, 80-85, 116-119, 128-131, 163, 164, 180, 181, 203-209, 219-225, 256-258, 274-276, 310-313, and 330-333, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 53;

[0445] b) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 68-71, 86-89, 120-122, 132-134, 152-156, 169-173, 213-215, 229-231, 251-255, 269-273, 305-309, and 325-329, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 55;

[0446] c) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 72, 73, 90, 91, 161, 162, 178, 179, 265, 266, 283, and 284 or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 52;

[0447] d) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 74, 75, 92, 93, 123, 124, 135, 136, 167, 184, 216-218, 232-234, 259-261, 277-279, 314-317, and 334-337, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 56;

[0448] e) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 76, 94, 210-212, 226-228, 322, 342, 562, and 563, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 42;

[0449] f) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 77, 95, 125, 137, 157-160, 174-177, 323, and 343, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 54;

[0450] g) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 78, 96, 126, 138, 168, 185, 267, 285, 318, 319, 338, and 339, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 57;

[0451] h) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 79, 97, 127, 139, 262-264, 280-282, 324, and 344, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 58; and

[0452] i) said target DNA sequence has the sequence set forth as any one of SEQ ID NOS: 165, 166, 182, 183, 268, 286, 320, 321, 340, and 341, or the complement thereof, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 59.

[0453] 176. The system of any one of embodiments 165-175, wherein at least one guide RNA is a dual-guide RNA.

[0454] 177. The system of any one of embodiments 165-175, wherein at least one guide RNA is a single guide RNA (sgRNA).

[0455] 178. The system of embodiment 177, wherein:

[0456] a) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 98-104, 140-143, 197, 198, 235-241, 292-294, and 350-353, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 53;

[0457] b) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 104-107, 144-146, 186-190, 245-247, 287-291, and 345-349, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 55;

[0458] c) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 108, 109, 195, 196, 301, and 302, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 52;

[0459] d) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 110, 111, 147, 148, 201, 248-250, 295-297, and 354-357, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 56;

[0460] e) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 112, 242-244, 362, and 564, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 42;

[0461] f) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 113, 149, 191-194, and 363, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 54;

[0462] g) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 114, 150, 202, 303, 358, and 359, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 57;

[0463] h) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 115, 151, 298-300, and 364, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 58; and

- [0464] i) said sgRNA has at least 90% sequence identity to any one of SEQ ID NOS: 199, 200, 304, 360, and 361, and wherein said RGN polypeptide has a sequence having at least 90% sequence identity to SEQ ID NO: 59.
- [0465] 179. The system of embodiment 177, wherein:
- [0466] a) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 98-104, 140-143, 197, 198, 235-241, 292-294, and 350-353, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 53;
- [0467] b) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 104-107, 144-146, 186-190, 245-247, 287-291, and 345-349, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 55;
- [0468] c) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 108, 109, 195, 196, 301, and 302, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 52;
- [0469] d) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 110, 111, 147, 148, 201, 248-250, 295-297, and 354-357, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 56;
- [0470] e) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 112, 242-244, 362, and 564, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 42;
- [0471] f) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 113, 149, 191-194, and 363, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 54;
- [0472] g) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 114, 150, 202, 303, 358, and 359, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 57;
- [0473] h) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 115, 151, 298-300, and 364, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 58; and
- [0474] i) said sgRNA has at least 95% sequence identity to any one of SEQ ID NOS: 199, 200, 304, 360, and 361, and wherein said RGN polypeptide has a sequence having at least 95% sequence identity to SEQ ID NO: 59.
- [0475] 180. The system of embodiment 177, wherein:
- [0476] a) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 98-104, 140-143, 197, 198, 235-241, 292-294, and 350-353, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 53;
- [0477] b) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 104-107, 144-146, 186-190, 245-247, 287-291, and 345-349, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 55;
- [0478] c) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 108, 109, 195, 196, 301, and 302, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 52;
- [0479] d) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 110, 111, 147, 148, 201, 248-250, 295-297, and 354-357, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 56;
- [0480] e) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 112, 242-244, 362, and 564, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 42;
- [0481] f) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 113, 149, 191-194, and 363, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 54;
- [0482] g) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 114, 150, 202, 303, 358, and 359, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 57;
- [0483] h) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 115, 151, 298-300, and 364, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 58; and
- [0484] i) said sgRNA has 100% sequence identity to any one of SEQ ID NOS: 199, 200, 304, 360, and 361, and wherein said RGN polypeptide has a sequence having 100% sequence identity to SEQ ID NO: 59.
- [0485] 181. A cell comprising the crRNA or nucleic acid molecule of embodiment 157, the guide RNA of any one of embodiments 158-163, the vector of embodiment 164 or the system of any one of embodiments 165-180.
- [0486] 182. A pharmaceutical composition comprising the crRNA or nucleic acid molecule of embodiment 157, the guide RNA of any one of embodiments 158-163, the vector of embodiment 164, the cell of embodiment 181, or the system of any one of embodiments 165-180, and a pharmaceutically acceptable carrier.
- [0487] 183. A composition comprising:
- [0488] a) a fusion protein comprising a DNA-binding polypeptide and an adenine deaminase, or a nucleic acid molecule encoding the fusion protein; and
- [0489] b) a second adenine deaminase having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441; or a nucleic acid molecule encoding the deaminase.
- [0490] 184. The composition of embodiment 183, wherein said second adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0491] 185. The composition of embodiment 183, wherein said second adenine deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0492] 186. The composition of any one of embodiments 183-185, wherein the first adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0493] 187. The composition of any one of embodiments 183-186, wherein the first adenine deaminase has at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.
- [0494] 188. The composition of any one of embodiments 183-186, wherein the first adenine deaminase has 100%

sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0495] 189. The composition of any one of embodiments 183-188, wherein the DNA-binding polypeptide is a meganuclease, zinc finger fusion protein, or a TALEN.

[0496] 190. The composition of any one of embodiments 183-189, wherein the DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide.

[0497] 191. The composition of embodiment 190, wherein the RNA-guided, DNA-binding polypeptide is an RNA-guided nuclease (RGN) polypeptide.

[0498] 192. The composition of embodiment 191, wherein the RGN is an RGN nickase.

[0499] 193. A vector comprising a nucleic acid molecule encoding a fusion protein and a nucleic acid molecule encoding a second deaminase, wherein said fusion protein comprises a DNA-binding polypeptide and a first adenine deaminase, and wherein said second adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0500] 194. The vector of embodiment 193, wherein said second adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0501] 195. The vector of embodiment 193, wherein said second adenine deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0502] 196. The vector of any one of embodiments 193-195, wherein the first adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0503] 197. The vector of any one of embodiments 193-195, wherein the first adenine deaminase has at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0504] 198. The vector of any one of embodiments 193-195, wherein the first adenine deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0505] 199. The vector of any one of embodiments 193-198, wherein the DNA-binding polypeptide is a meganuclease, zinc finger fusion protein, or a TALEN.

[0506] 200. The vector of any one of embodiments 193-198, wherein the DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide.

[0507] 201. The vector of embodiment 200, wherein the RNA-guided, DNA-binding polypeptide is an RNA-guided nuclease (RGN) polypeptide.

[0508] 202. The vector of embodiment 201, wherein the RGN is an RGN nickase.

[0509] 203. A cell comprising the vector of any one of embodiments 193-202.

[0510] 204. A cell comprising:

[0511] a) a fusion protein comprising a DNA-binding polypeptide and a first adenine deaminase; or a nucleic acid molecule encoding the fusion protein; and

[0512] b) a second adenine deaminase having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441; or a nucleic acid molecule encoding the second adenine deaminase.

[0513] 205. The cell of embodiment 204, wherein said second adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0514] 206. The cell of embodiment 204, wherein said second adenine deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0515] 207. The cell of any one of embodiments 204-206, wherein the first adenine deaminase has at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0516] 208. The cell of any one of embodiments 204-206, wherein the first adenine deaminase has at least 95% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0517] 209. The cell of any one of embodiments 204-206, wherein the first adenine deaminase has 100% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441.

[0518] 210. The cell of any one of embodiments 204-209, wherein the DNA-binding polypeptide is a meganuclease, zinc finger fusion protein, or a TALEN.

[0519] 211. The cell of any one of embodiments 204-209, wherein the DNA-binding polypeptide is an RNA-guided, DNA-binding polypeptide.

[0520] 212. The cell of embodiment 211, wherein the RNA-guided, DNA-binding polypeptide is an RNA-guided nuclease (RGN) polypeptide.

[0521] 213. The cell of embodiment 212, wherein the RGN is an RGN nickase.

[0522] 214. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the composition of any one of embodiments 183-192, the vector of any one of embodiments 193-202, or the cell of any one of embodiments 203-213.

[0523] 215. A method for treating a disease, said method comprising administering to a subject in need thereof an effective amount of a pharmaceutical composition of any one of embodiments 69, 100, 142, and 214.

[0524] 216. The method of embodiment 215, wherein said disease is associated with a causal mutation and said effective amount of said pharmaceutical composition corrects said causal mutation.

[0525] 217. Use of the fusion protein of any one of embodiments 14-27, the nucleic acid molecule of any one of embodiments 28-48, the vector of any one of embodiments 49-52 and 193-202, the cell of any one of embodiments 59-63, 135-139, and 203-213, the system of any one of embodiments 75-99, or the composition of any one of embodiments 183-192 for the treatment of a disease in a subject.

[0526] 218. The use of embodiment 217, wherein said disease is associated with a causal mutation and said treating comprises correcting said causal mutation.

[0527] 219. Use of the fusion protein of any one of embodiments 14-27, the nucleic acid molecule of any one of embodiments 28-48, the vector of any one of embodiments 49-52 and 193-202, the cell of any one of embodiments 59-63, 135-139, and 203-213, the system of any one of embodiments 75-99, or the composition of any one of embodiments 183-192 for the manufacture of a medicament useful for treating a disease.

[0528] 220. The use of embodiment 219, wherein said disease is associated with a causal mutation and an effective amount of said medicament corrects said causal mutation.

[0529] 221. A nucleic acid molecule comprising a polynucleotide encoding an RNA-guided nuclease (RGN) polypeptide, wherein said polynucleotide comprises a nucleotide sequence encoding an RGN polypeptide comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 41 or 60, but lacking amino acid residues 590 to 597 of SEQ ID NO: 41 or 60;

[0530] wherein said RGN polypeptide is capable of binding a target DNA sequence in an RNA-guided sequence specific manner when bound to a guide RNA (gRNA) capable of hybridizing to said target DNA sequence.

[0531] 222. The nucleic acid molecule of embodiment 221, wherein said polynucleotide encoding an RGN polypeptide is operably linked to a promoter heterologous to said polynucleotide.

[0532] 223. The nucleic acid molecule of embodiment 221 or 222, wherein said RGN polypeptide comprises an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 366 or 368.

[0533] 224. The nucleic acid molecule of embodiment 221 or 222, wherein said RGN polypeptide comprises an amino acid sequence of SEQ ID NO: 366 or 368.

[0534] 225. The nucleic acid molecule of any one of embodiments 221-223, wherein said RGN polypeptide is nuclease dead or functions as a nickase.

[0535] 226. The nucleic acid molecule of embodiment 225, wherein said nickase has the amino acid sequence set forth in SEQ ID NO: 397 or 398.

[0536] 227. The nucleic acid molecule of any one of embodiments 221-226, wherein the RGN polypeptide is operably fused to a base-editing polypeptide.

[0537] 228. A vector comprising the nucleic acid molecule of any one of claims 221-227.

[0538] 229. An isolated polypeptide comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 41 or 60, but lacking amino acid residues 590 to 597 of SEQ ID NO: 41 or 60, wherein said polypeptide is an RNA-guided nuclease.

[0539] 230. The isolated polypeptide of embodiment 229, wherein said RGN polypeptide comprises an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 366 or 368.

[0540] 231. The isolated polypeptide of embodiment 230, wherein said RGN polypeptide comprises an amino acid sequence of SEQ ID NO: 366 or 368.

[0541] 232. The isolated polypeptide of embodiment 229 or 230, wherein said RGN polypeptide is nuclease dead or functions as a nickase.

[0542] 233. The isolated polypeptide of embodiment 232, wherein said nickase has the amino acid sequence set forth in SEQ ID NO: 397 or 398.

[0543] 234. The isolated polypeptide of any one of embodiments 229-233, wherein the RGN polypeptide is operably fused to a base-editing polypeptide.

[0544] 235. A cell comprising the nucleic acid molecule of any one of embodiments 221-227, the vector of claim 228, or the polypeptide of any one of claims 229-234.

[0545] 236. An isolated polypeptide comprising an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 407, wherein said polypeptide has deaminase activity.

[0546] 237. The isolated polypeptide of embodiment 236 comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 407, wherein said polypeptide has deaminase activity.

[0547] 238. The isolated polypeptide of embodiment 236, wherein the polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 407.

[0548] 239. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein the deaminase is encoded by a nucleotide sequence that:

[0549] a) has at least 80% sequence identity to SEQ ID NO: 451, or

[0550] b) encodes an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NO: 407.

[0551] 240. The nucleic acid molecule of embodiment 239, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 451.

[0552] 241. The nucleic acid molecule of embodiment 239, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 451.

[0553] 242. The nucleic acid molecule of embodiment 239, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 451.

[0554] 243. The nucleic acid molecule of embodiments 239-242, wherein said nucleic acid molecule further comprises a heterologous promoter operably linked to said polynucleotide.

[0555] 244. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the polypeptide of any one of embodiments 236-238 or the nucleic acid molecule of any one of embodiments 239-242.

[0556] 245. A fusion protein comprising a DNA-binding polypeptide and a deaminase having at least 90% sequence identity to SEQ ID NO: 407.

[0557] 246. A fusion protein of embodiment 245 comprising a DNA-binding polypeptide and a deaminase having at least 95% sequence identity to SEQ ID NO: 407.

[0558] 247. A fusion protein of embodiment 245 comprising a DNA-binding polypeptide and a deaminase having 100% sequence identity to SEQ ID NO: 407.

[0559] 248. The fusion protein of any one of embodiments 245-247, wherein the DNA-binding polypeptide is a RNA-guided nuclease (RGN) polypeptide.

[0560] 249. The fusion protein of embodiment 248, wherein the RGN polypeptide is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0561] 250. The fusion protein of any one of embodiments 248-249, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0562] 251. The fusion protein of any one of embodiments 248-250, wherein the RGN polypeptide is a nickase.

[0563] 252. The fusion protein of embodiment 251, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0564] 253. The fusion protein of embodiment 251, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0565] 254. A nucleic acid molecule comprising a polynucleotide encoding a fusion protein comprising a DNA-binding polypeptide and a deaminase, wherein the deaminase is encoded by a nucleotide sequence that:

[0566] a) has at least 80% sequence identity to SEQ ID NO: 451, or

[0567] b) encodes an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 407.

[0568] 255. The nucleic acid molecule of embodiment 254, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 451.

[0569] 256. The nucleic acid molecule of embodiment 254, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 451.

[0570] 257. The nucleic acid molecule of embodiment 254, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 451.

[0571] 258. The nucleic acid molecule of any one of embodiments 254-257, wherein the DNA-binding polypeptide is a RGN polypeptide.

[0572] 259. The nucleic acid molecule of embodiment 258, wherein the RGN is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0573] 260. The nucleic acid molecule of any one of embodiments 258-259, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Casl2a, a Casl2b, a Casl2g, a Casl2h, a Casl2i, a Casl3b, a Casl3c, a Casl3d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0574] 261. The nucleic acid molecule of any one of embodiments 258-260, wherein the RGN polypeptide is a nickase.

[0575] 262. The nucleic acid molecule of embodiment 261, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0576] 263. The nucleic acid molecule of embodiment 262, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0577] 264. A vector comprising the nucleic acid molecule of any one of embodiments 254-263.

[0578] 265. The vector of embodiment 264, further comprising at least one nucleotide sequence encoding a guide RNA (gRNA) capable of hybridizing to a target sequence.

[0579] 266. A ribonucleoprotein (RNP) complex comprising the fusion protein of any one of embodiments 245-253 and a guide RNA bound to the DNA-binding polypeptide of the fusion protein.

[0580] 267. A cell comprising the fusion protein of any of embodiments 245-253, the nucleic acid molecule of any one of embodiments 254-263, the vector of any one of embodiments 264-265, or the RNP complex of embodiment 266.

[0581] 268. A system for modifying a target DNA molecule comprising a target DNA sequence, said system comprising:

[0582] a) a fusion protein comprising an RNA-guided nuclease (RGN) polypeptide and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 407, or a nucleotide sequence encoding said fusion protein; and

[0583] b) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and

[0584] wherein the one or more guide RNAs are capable of forming a complex with the fusion protein in order to direct said fusion protein to bind to said target DNA sequence and modify the target DNA molecule.

[0585] 269. The system of embodiment 268, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 407.

[0586] 270. The system of embodiment 268, wherein said deaminase has an amino acid sequence having 100% sequence identity to SEQ ID NO: 407.

[0587] 271. The system of any one of embodiments 268-270, wherein at least one of said nucleotide sequence encoding the one or more guide RNAs and said nucleotide sequence encoding the fusion protein is operably linked to a promoter heterologous to said nucleotide sequence.

[0588] 272. The system of any one of embodiments 268-271, wherein the target DNA sequence is located adjacent to a protospacer adjacent motif (PAM) that is recognized by the RGN polypeptide.

[0589] 273. The system of any one of embodiments 268-272, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof. 274. The system of any one of embodiments 268-273, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0590] 275. The system of any one of embodiments 268-274, wherein the RGN polypeptide of the fusion protein is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0591] 276. The system of any one of embodiments 272-275, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Casl2a, a Casl2b, a Casl2g, a Casl2h, a Casl2i, a Casl3b, a Casl3c, a Casl3d, a Casl4, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0592] 277. The system of embodiment 276, wherein the RGN polypeptide is a nickase.

[0593] 278. The system of embodiment 277, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0594] 279. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the fusion protein of any of embodiments 245-253, the nucleic acid molecule of any one of embodiments 254-263, the vector of any one of embodiments 264-265, the RNP complex of embodiment 266, the cell of embodiment 267, or the system of any one of embodiments 268-28.

[0595] 280. A method for modifying a target DNA molecule comprising a target sequence comprising:

[0596] a) assembling an RGN-deaminase ribonucleotide complex by combining:

[0597] i) one or more guide RNAs capable of hybridizing to the target DNA sequence; and

[0598] ii) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN), and at least one deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 407;

[0599] under conditions suitable for formation of the RGN-deaminase ribonucleotide complex; and

[0600] b) contacting said target DNA molecule or a cell comprising said target DNA molecule with the assembled RGN-deaminase ribonucleotide complex;

[0601] wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said fusion protein to bind to said target DNA sequence and modification of the target DNA molecule occurs.

[0602] 281. The method of embodiment 280, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0603] 282. The method of any one of embodiments 280-281, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0604] 283. The method of any one of embodiments 280-283, wherein the method is performed in vitro, in vivo, or ex vivo.

[0605] 284. A method of treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

[0606] administering to the subject the fusion protein of any of embodiments 245-253, the nucleic acid molecule of any one of embodiments 254-263, the vector of any one of embodiments 264-265, the RNP complex of embodiment 266, the cell of embodiment 267, the system of any one of embodiments 268-28, or the pharmaceutical composition of embodiment 279.

[0607] 285. The method of embodiment 284, further comprising administering any one of a gRNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0608] 286. An isolated polypeptide comprising an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405, wherein said polypeptide has deaminase activity.

[0609] 287. The isolated polypeptide of embodiment 286 comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 405, wherein said polypeptide has deaminase activity.

[0610] 288. The isolated polypeptide of embodiment 286, wherein the polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 407.

[0611] 289. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein the deaminase is encoded by a nucleotide sequence that:

[0612] a) has at least 80% sequence identity to SEQ ID NO: 449, or

[0613] b) encodes an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NO: 405.

[0614] 290. The nucleic acid molecule of embodiment 289, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 449.

[0615] 291. The nucleic acid molecule of embodiment 289, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 449.

[0616] 292. The nucleic acid molecule of embodiment 289, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 449.

[0617] 293. The nucleic acid molecule of embodiments 289-292, wherein said nucleic acid molecule further comprises a heterologous promoter operably linked to said polynucleotide.

[0618] 294. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the polypeptide of any one of embodiments 286-288 or the nucleic acid molecule of any one of embodiments 289-293.

[0619] 295. A fusion protein comprising a DNA-binding polypeptide and a deaminase having at least 90% sequence identity to SEQ ID NO: 405.

[0620] 296. A fusion protein of embodiment 295 comprising a DNA-binding polypeptide and a deaminase having at least 95% sequence identity to SEQ ID NO: 405.

[0621] 297. A fusion protein of embodiment 295 comprising a DNA-binding polypeptide and a deaminase having 100% sequence identity to SEQ ID NO: 405.

[0622] 298. The fusion protein of any one of embodiments 295-297, wherein the DNA-binding polypeptide is a RNA-guided nuclease (RGN) polypeptide.

[0623] 299. The fusion protein of embodiment 298, wherein the RGN polypeptide is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0624] 300. The fusion protein of any one of embodiments 298-299, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0625] 301. The fusion protein of any one of embodiments 298-300, wherein the RGN polypeptide is a nickase.

[0626] 302. The fusion protein of embodiment 301, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0627] 303. The fusion protein of embodiment 301, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0628] 304. A nucleic acid molecule comprising a polynucleotide encoding a fusion protein comprising a DNA-binding polypeptide and a deaminase, wherein the deaminase is encoded by a nucleotide sequence that:

[0629] a) has at least 80% sequence identity to SEQ ID NO: 449, or

[0630] b) encodes an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405.

[0631] 305. The nucleic acid molecule of embodiment 304, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 449.

[0632] 306. The nucleic acid molecule of embodiment 304, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 449.

[0633] 307. The nucleic acid molecule of embodiment 304, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 449.

[0634] 308. The nucleic acid molecule of any one of embodiments 304-307, wherein the DNA-binding polypeptide is a RGN polypeptide.

[0635] 309. The nucleic acid molecule of embodiment 308, wherein the RGN is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0636] 310. The nucleic acid molecule of any one of embodiments 308-309, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2c1, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0637] 311. The nucleic acid molecule of any one of embodiments 308-310, wherein the RGN polypeptide is a nickase.

[0638] 312. The nucleic acid molecule of embodiment 311, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0639] 313. The nucleic acid molecule of embodiment 312, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0640] 314. A vector comprising the nucleic acid molecule of any one of embodiments 304-313.

[0641] 315. The vector of embodiment 314, further comprising at least one nucleotide sequence encoding a guide RNA (gRNA) capable of hybridizing to a target sequence.

[0642] 316. A ribonucleoprotein (RNP) complex comprising the fusion protein of any one of embodiments 295-303 and a guide RNA bound to the DNA-binding polypeptide of the fusion protein.

[0643] 317. A cell comprising the fusion protein of any of embodiments 295-303, the nucleic acid molecule of any one of embodiments 304-313, the vector of any one of embodiments 314-315, or the RNP complex of embodiment 316.

[0644] 318. A system for modifying a target DNA molecule comprising a target DNA sequence, said system comprising:

[0645] a) a fusion protein comprising an RNA-guided nuclease (RGN) polypeptide and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405, or a nucleotide sequence encoding said fusion protein; and

[0646] b) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and

[0647] wherein the one or more guide RNAs are capable of forming a complex with the fusion protein in order to direct said fusion protein to bind to said target DNA sequence and modify the target DNA molecule.

[0648] 319. The system of embodiment 318, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 405.

[0649] 320. The system of embodiment 318, wherein said deaminase has an amino acid sequence having 100% sequence identity to SEQ ID NO: 405.

[0650] 321. The system of any one of embodiments 318-320, wherein at least one of said nucleotide sequence encoding the one or more guide RNAs and said nucleotide sequence encoding the fusion protein is operably linked to a promoter heterologous to said nucleotide sequence.

[0651] 322. The system of any one of embodiments 318-321, wherein the target DNA sequence is located adjacent to a protospacer adjacent motif (PAM) that is recognized by the RGN polypeptide.

[0652] 323. The system of any one of embodiments 318-322, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0653] 324. The system of any one of embodiments 318-323, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0654] 325. The system of any one of embodiments 318-324, wherein the RGN polypeptide of the fusion protein is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0655] 326. The system of any one of embodiments 322-325, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2c1, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0656] 327. The system of embodiment 326, wherein the RGN polypeptide is a nickase.

[0657] 328. The system of embodiment 327, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0658] 329. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the fusion protein of any of embodiments 295-303, the nucleic acid molecule of any one of embodiments 304-313, the vector of any one of embodiments 314-315, the RNP complex of embodiment 316, the cell of embodiment 317, or the system of any one of embodiments 318-328.

[0659] 330. A method for modifying a target DNA molecule comprising a target sequence comprising:

[0660] a) assembling an RGN-deaminase ribonucleotide complex by combining:

[0661] i) one or more guide RNAs capable of hybridizing to the target DNA sequence; and

[0662] ii) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN), and at least one deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405;

[0663] under conditions suitable for formation of the RGN-deaminase ribonucleotide complex; and

[0664] b) contacting said target DNA molecule or a cell comprising said target DNA molecule with the assembled RGN-deaminase ribonucleotide complex;

[0665] wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said fusion protein to bind to said target DNA sequence and modification of the target DNA molecule occurs.

[0666] 331. The method of embodiment 330, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0667] 332. The method of any one of embodiments 330-331, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0668] 333. The method of any one of embodiments 330-332, wherein the method is performed *in vitro*, *in vivo*, or *ex vivo*.

[0669] 334. A method of treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

[0670] administering to the subject the fusion protein of any of embodiments 295-303, the nucleic acid molecule of any one of embodiments 304-313, the vector of any one of embodiments 314-315, the RNP complex of embodiment 316, the cell of embodiment 317, the system of any one of embodiments 318-328, or the pharmaceutical composition of embodiment 329.

[0671] 335. The method of embodiment 334, further comprising administering any one of a gRNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0672] 336. An isolated polypeptide comprising an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 399, wherein said polypeptide has deaminase activity.

[0673] 337. The isolated polypeptide of embodiment 336 comprising an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 399, wherein said polypeptide has deaminase activity.

[0674] 338. The isolated polypeptide of embodiment 336, wherein the polypeptide comprises an amino acid sequence set forth in SEQ ID NO: 399.

[0675] 339. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein the deaminase is encoded by a nucleotide sequence that:

[0676] a) has at least 80% sequence identity to SEQ ID NO: 443, or

[0677] b) encodes an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NO: 399.

[0678] 340. The nucleic acid molecule of embodiment 339, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 443.

[0679] 341. The nucleic acid molecule of embodiment 339, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 443.

[0680] 342. The nucleic acid molecule of embodiment 339, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 443.

[0681] 343. The nucleic acid molecule of embodiments 339-342, wherein said nucleic acid molecule further comprises a heterologous promoter operably linked to said polynucleotide.

[0682] 344. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the polypeptide of any one of embodiments 336-338 or the nucleic acid molecule of any one of embodiments 339-342.

[0683] 345. A fusion protein comprising a DNA-binding polypeptide and a deaminase having at least 90% sequence identity to SEQ ID NO: 399.

[0684] 346. A fusion protein of embodiment 345 comprising a DNA-binding polypeptide and a deaminase having at least 95% sequence identity to SEQ ID NO: 399.

[0685] 347. A fusion protein of embodiment 345 comprising a DNA-binding polypeptide and a deaminase having 100% sequence identity to SEQ ID NO: 399.

[0686] 348. The fusion protein of any one of embodiments 345-347, wherein the DNA-binding polypeptide is a RNA-guided nuclease (RGN) polypeptide.

[0687] 349. The fusion protein of embodiment 348, wherein the RGN polypeptide is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0688] 350. The fusion protein of any one of embodiments 348-349, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2c1, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0689] 351. The fusion protein of any one of embodiments 348-350, wherein the RGN polypeptide is a nickase.

[0690] 352. The fusion protein of embodiment 351, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0691] 353. The fusion protein of embodiment 351, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0692] 354. A nucleic acid molecule comprising a polynucleotide encoding a fusion protein comprising a DNA-binding polypeptide and a deaminase, wherein the deaminase is encoded by a nucleotide sequence that:

[0693] a) has at least 80% sequence identity to SEQ ID NO: 443, or

[0694] b) encodes an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 399.

[0695] 355. The nucleic acid molecule of embodiment 354, wherein the deaminase is encoded by a nucleotide sequence that has at least 90% sequence identity to SEQ ID NO: 443.

[0696] 356. The nucleic acid molecule of embodiment 354, wherein the deaminase is encoded by a nucleotide sequence that has at least 95% sequence identity to SEQ ID NO: 443.

[0697] 357. The nucleic acid molecule of embodiment 354, wherein the deaminase is encoded by a nucleotide sequence that has at least 100% sequence identity to SEQ ID NO: 443.

[0698] 358. The nucleic acid molecule of any one of embodiments 354-357, wherein the DNA-binding polypeptide is a RGN polypeptide.

[0699] 359. The nucleic acid molecule of embodiment 358, wherein the RGN is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0700] 360. The nucleic acid molecule of any one of embodiments 358-359, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN polypeptide with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0701] 361. The nucleic acid molecule of any one of embodiments 358-360, wherein the RGN polypeptide is a nickase.

[0702] 362. The nucleic acid molecule of embodiment 361, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0703] 363. The nucleic acid molecule of embodiment 362, wherein the nickase has an amino acid sequence having 100% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0704] 364. A vector comprising the nucleic acid molecule of any one of embodiments 354-363.

[0705] 365. The vector of embodiment 364, further comprising at least one nucleotide sequence encoding a guide RNA (gRNA) capable of hybridizing to a target sequence.

[0706] 366. A ribonucleoprotein (RNP) complex comprising the fusion protein of any one of embodiments 345-353 and a guide RNA bound to the DNA-binding polypeptide of the fusion protein.

[0707] 367. A cell comprising the fusion protein of any of embodiments 345-353, the nucleic acid molecule of any one of embodiments 354-363, the vector of any one of embodiments 364-365, or the RNP complex of embodiment 366.

[0708] 368. A system for modifying a target DNA molecule comprising a target DNA sequence, said system comprising:

[0709] a) a fusion protein comprising an RNA-guided nuclease (RGN) polypeptide and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 399, or a nucleotide sequence encoding said fusion protein; and

[0710] b) one or more guide RNAs capable of hybridizing to said target DNA sequence or one or more nucleotide sequences encoding the one or more guide RNAs (gRNAs); and

[0711] wherein the one or more guide RNAs are capable of forming a complex with the fusion protein in order to direct said fusion protein to bind to said target DNA sequence and modify the target DNA molecule.

[0712] 369. The system of embodiment 368, wherein said deaminase has an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 399.

[0713] 370. The system of embodiment 368, wherein said deaminase has an amino acid sequence having 100% sequence identity to SEQ ID NO: 399.

[0714] 371. The system of any one of embodiments 368-370, wherein at least one of said nucleotide sequence encoding the one or more guide RNAs and said nucleotide sequence encoding the fusion protein is operably linked to a promoter heterologous to said nucleotide sequence.

[0715] 372. The system of any one of embodiments 368-371, wherein the target DNA sequence is located adjacent to a protospacer adjacent motif (PAM) that is recognized by the RGN polypeptide.

[0716] 373. The system of any one of embodiments 368-372, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0717] 374. The system of any one of embodiments 368-373, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0718] 375. The system of any one of embodiments 368-374, wherein the RGN polypeptide of the fusion protein is a Type II CRISPR-Cas polypeptide or a Type V CRISPR-Cas polypeptide.

[0719] 376. The system of any one of embodiments 372-375, wherein the RGN polypeptide is a Cas9, a CasX, a CasY, a Cpf1, a C2cl, a C2c2, a C2c3, a GeoCas9, a CjCas9, a Cas12a, a Cas12b, a Cas12g, a Cas12h, a Cas12i, a Cas13b, a Cas13c, a Cas13d, a Cas14, a Csn2, an xCas9, an SpCas9-NG, an LbCas12a, an AsCas12a, a Cas9-KKH, a circularly permuted Cas9, an Argonaute (Ago), a SmacCas9, a Spy-macCas9 domain, or a RGN with an amino acid sequence set forth in any one of SEQ ID NOS: 41, 60, 366, or 368.

[0720] 377. The system of embodiment 376, wherein the RGN polypeptide is a nickase.

[0721] 378. The system of embodiment 377, wherein the nickase has an amino acid sequence having at least 95% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0722] 379. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and the fusion protein of any of embodiments 345-353, the nucleic acid molecule of any one of embodiments 354-363, the vector of any one of

embodiments 364-365, the RNP complex of embodiment 366, the cell of embodiment 367, or the system of any one of embodiments 368-378.

[0723] 380. A method for modifying a target DNA molecule comprising a target sequence comprising:

[0724] a) assembling an RGN-deaminase ribonucleotide complex by combining:

[0725] i) one or more guide RNAs capable of hybridizing to the target DNA sequence; and

[0726] ii) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN), and at least one deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 399;

[0727] under conditions suitable for formation of the RGN-deaminase ribonucleotide complex; and

[0728] b) contacting said target DNA molecule or a cell comprising said target DNA molecule with the assembled RGN-deaminase ribonucleotide complex;

[0729] wherein the one or more guide RNAs hybridize to the target DNA sequence, thereby directing said fusion protein to bind to said target DNA sequence and modification of the target DNA molecule occurs.

[0730] 381. The method of embodiment 380, wherein the target DNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0731] 382. The method of any one of embodiments 380-381, wherein the gRNA sequence comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0732] 383. The method of any one of embodiments 380-382, wherein the method is performed in vitro, in vivo, or ex vivo.

[0733] 384. A method of treating a subject having or at risk of developing a disease, disorder, or condition, the method comprising:

[0734] administering to the subject the fusion protein of any of embodiments 345-353, the nucleic acid molecule of any one of embodiments 354-363, the vector of any one of embodiments 364-365, the RNP complex of embodiment 366, the cell of embodiment 367, the system of any one of embodiments 368-378, or the pharmaceutical composition of embodiment 379.

[0735] 385. The method of embodiment 384, further comprising administering any one of a gRNA comprising a nucleic acid sequence selected from the group consisting of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0736] 386. A method for producing a treating or reducing at least one symptom of cystic fibrosis, the method comprising administering to a subject in need thereof an effective amount of:

[0737] a) a fusion protein comprising an RNA-guided nuclease polypeptide (RGN) and a deaminase, wherein the deaminase has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 407, 405, 399, 1-10, 400-404, 406, and 408-441, or a polynucleotide encoding said fusion protein, wherein said polynucleotide encoding the fusion protein is operably linked to a promoter to enable expression of the fusion protein in the cell; and

[0738] b) one or more guide RNAs (gRNA) capable of hybridizing to a target DNA sequence, or a polynucleotide encoding said gRNA, wherein said polynucleotide encoding the gRNA is operably linked to a promoter to enable expression of the gRNA in the cell; whereby the fusion protein and gRNA target to the genomic location of the causal mutation and modify the genomic sequence to remove the causal mutation.

[0739] 387. The method of embodiment 386, wherein the gRNA comprises a spacer sequence that targets any one of SEQ ID NOS: 62-97, 116-139, 152-185, 203-234, 251-286, 305-344, 562, and 563, or the complement thereof.

[0740] 388. The method of embodiments 386 or 387, wherein the gRNA comprises any one of SEQ ID NOS: 98-115, 140-151, 186-202, 235-250, 287-304, 345-364, and 564.

[0741] 389. The method of any one of claims 386-388, wherein said the RGN has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 41, 60, 366, and 368.

[0742] 390. The method of any one of claims 386-389, wherein said the RGN has an amino acid sequence having at least 90% sequence identity to any one of SEQ ID NOS: 42, 52-59, 61, 397, and 398.

[0743] The following examples are offered by way of illustration and not by way of limitation.

## EXPERIMENTAL

### Example 1: Demonstration of Base Editing in Mammalian Cells

[0744] The deaminases shown in Table 1 below were produced based on naturally occurring deaminases which were then mutated and selected for adenine deaminase activity in prokaryotic cells.

TABLE 1

Deaminase sequences	
Deaminase	SEQ ID NO.
APG09982	1
APG03724	2
APG09949	3
APG08196	4
APG06333	5
APG06489	6
APG08449	7
APG05174	8
APG09102	9
APG05723	10

[0745] To determine if the deaminases of Table 1 are able to perform adenine base editing in mammalian cells, each deaminase was operably fused to an RGN nuclease to produce a fusion protein. Residues predicted to deactivate the RuvC domain of the RGN APG07433.1 (set forth as SEQ ID NO: 41; described in PCT publication WO 2019/236566, incorporated by reference herein) were identified and the RGN was modified to a nuclease variant (nAPG07433.1; SEQ ID NO: 42). A nuclease variant of an RGN is referred to herein as "nRGN". It should be understood that any nuclease variant of an RGN may be used to produce a fusion protein of the invention.

**[0746]** Deaminase and nRGN nucleotide sequences codon optimized for mammalian expression were synthesized as fusion proteins with an N-terminal nuclear localization tag and cloned into the pTwist CMV (Twist Biosciences) expression plasmid. Each fusion protein comprises, starting at the amino terminus, the SV40 NLS (SEQ ID NO: 43) operably linked at the C-terminal end to 3× FLAG Tag (SEQ ID NO: 44), operably linked at the C-terminal end to a deaminase, operably linked at the C-terminal end to a peptide linker (SEQ ID NO: 45), operably linked at the C-terminal end to an nRGN (for example, nAPG07433.1, which is SEQ ID NO: 42), finally operably linked at the C-terminal end to the nucleoplasmin NLS (SEQ ID NO: 45). All fusion proteins comprise at least one NLS and a 3× FLAG Tag, as described above.

**[0747]** Expression plasmids comprising an expression cassette encoding a sgRNA expressed by a human U6 promoter (SEQ ID NO: 50) were also produced. Human genomic target sequences and the sgRNA sequences for guiding the fusion proteins to the genomic targets are indicated in Table 2.

TABLE 2

Guide RNA sequences				
sgRNA ID	Target sequence	sgRNA sequence	Forward Primer for amplification	Reverse Primer for amplification
SGN000930	21	26	31	32
SGN000186	22	27	33	34
SGN000194	23	28	35	36
SGN000143	24	29	37	38
SGN000139	25	30	39	40

**[0748]** 500 ng of plasmid comprising an expression cassette comprising a coding sequence for a fusion protein for each deaminase described in Table 1 and 500 ng of plasmid comprising an expression cassette encoding an sgRNA shown in Table 2 were co-transfected into HEK293 FT cells at 75-90% confluence in 24-well plates using Lipofectamine 2000 reagent (Life Technologies). Cells were then incubated at 37°C for 72 h. Following incubation, genomic DNA was then extracted using NucleoSpin 96 Tissue (Macherey-Nagel) following the manufacturer's protocol. The genomic region flanking the targeted genomic site was PCR amplified using the primers in Table 2 and products were purified using ZR-96 DNA Clean and Concentrator (Zymo Research) following the manufacturer's protocol. The purified PCR products underwent Next Generation Sequencing on Illumina MiSeq. Typically, 100,000 of 250 bp paired-end reads (2×100,000 reads) are generated per amplicon. The reads were analyzed using CRISPResso (Pinello, et al. 2016 *Nature Biotech*, 34:695-697) to calculate the rates of editing. Output alignments were analyzed for INDEL formation or introduction of specific adenine mutations. Tables 3 through 7 show adenine base editing for each fusion protein comprising nAPG07433.1 and a deaminase from Table 1 and a guide RNA from Table 2. The deaminase component of each fusion protein is indicated. The editing rate for adenines within or proximal to the target sequence is indicated. "A5" indicates, for example, an adenine at position 5 of the target sequence. The position of each nucleotide in the target sequence was determined by numbering the first nucleotide in the target sequence closest to the PAM as position 1, and

the position number increases in the 3' direction away from the PAM sequence. The tables also show which nucleotide the adenine was changed to, and at what rate. For example, Table 3 shows that for the APG09982-nAPG07433.1 fusion protein, the adenine at position 13 was mutated to a guanine at a rate of 1.2%.

TABLE 3

A > N Editing Rate using guide SGN000139					
Deaminase	A5	A12	A13	A20	A22
APG09982	C	0	0	0.3	0
	G	0	0.5	1.2	0
	T	0	0	0	0
APG03724	C	0	0	0	0.3
	G	0	0.7	0.7	0.1
	T	0	0	0	0
APG09949	C	0	0	0	0.3
	G	0.1	0.6	0.7	0
	T	0	0	0	0
APG08196	C	0	0	0	0.6
	G	0	0.6	0.6	0
	T	0	0	0	0
APG06333	C	0	0	0.2	0
	G	0	0.5	1	0
	T	0	0	0	0
APG06489	C	0	0	0	0.2
	G	0	0.6	0.4	0
	T	0	0	0	0
APG08449	C	0	0	0	0.3
	G	0	0.8	0.8	0
	T	0	0	0	0
APG05174	C	0	0	0	0.6
	G	0	0.6	0.7	0
	T	0	0	0	0
APG09102	C	0	0	0	0.1
	G	0	0.6	0.6	0
	T	0	0	0	0
APG05723	C	0	0	0	0.1
	G	0	0.4	0.5	0.1
	T	0	0	0	0

**[0749]** All fusion proteins showed detectable A>G conversion at positions A12 and A13. APG09982 and APG06333 showed at least 1% editing at position A13.

TABLE 4

A > N Editing Rate using guide SGN000143								
Deaminase	A1	A4	A6	A9	A11	A14	A19	A30
APG09982	C	0	0	0	0	0	0	0
	G	0	0	0.1	4.5	1.7	0	0
	T	0	0	0	0	0	0	0
APG03724	C	0	0	0	0	0	0	0
	G	0	0	0.1	0.1	1.3	1.1	0
	T	0	0	0	0	0	0	0
APG09949	C	0	0	0	0	0	0	0.1
	G	0	0	0.1	0.8	0.7	0	0
	T	0	0	0	0	0	0	0
APG08196	C	0	0	0	0	0	0	0
	G	0	0	0	0.4	0.7	0.5	0.1
	T	0	0	0	0	0	0	0
APG06333	C	0	0	0	0	0	0	0
	G	0	0	0	0	1.3	0.8	0.1
	T	0	0	0	0	0	0	0
APG06489	C	0	0	0	0	0	0	0
	G	0	0	0.1	0.6	1.8	0.8	0.1
	T	0	0	0	0	0	0	0
APG08449	C	0	0	0	0	0	0	0.1
	G	0	0	0	0	2.4	1.2	0
	T	0	0	0	0	0	0	0

TABLE 4-continued

**[0750]** All fusion proteins showed A>G conversion at positions A11 and A14. APG09982 showed 4.5% conversion of A11 to G and 1.7% conversion of A14 to G.

TABLE 5

A > N Editing Rate using guide SGN000186								
Deaminase		A9	A16	A18	A22	A25	A28	A30
APG09982	C	0	0	0	0	0	0	0
	G	1.7	4.5	2	0	0	0	0
	T	0	0	0	0	0	0	0
APG03724	C	0	0	0	0	0.1	0	0
	G	0.7	4.1	1.4	0	0	0	0
	T	0	0	0	0	0	0	0
APG09949	C	0	0	0.1	0	0.1	0	0
	G	0.6	3.4	1.1	0	0	0	0
	T	0	0	0	0	0.1	0	0
APG08196	C	0	0	0.1	0	0.1	0	0
	G	1	3.3	1.4	0	0	0.1	0
	T	0	0	0	0	0.1	0	0
APG06333	C	0	0	0	0	0	0	0
	G	1.4	4.2	1.9	0	0	0	0
	T	0	0	0	0	0	0	0
APG06489	C	0	0	0	0	0	0	0
	G	1.7	2.5	1.4	0	0	0	0.1
	T	0	0	0	0	0	0	0
APG08449	C	0	0	0.1	0	0.1	0	0
	G	1.5	5.3	1.6	0	0	0	0
	T	0	0	0	0	0.1	0	0
APG05174	C	0	0	0.1	0	0	0	0
	G	0.9	3.2	1	0	0	0.1	0
	T	0	0	0	0	0.1	0	0
APG09102	C	0	0	0	0	0	0	0
	G	2.3	6.2	2.1	0	0	0	0
	T	0	0	0	0	0	0	0

TABLE 5-continued

A > N Editing Rate using guide SGN000186							
Deaminase	A9	A16	A18	A22	A25	A28	A30
APG05723	C G T	0 1.1 0	0 1.9 0	0 1.2 0	0 0 0	0 0 0	0 0 0

**[0751]** All fusion proteins showed base editing of over 1% at multiple locations in target SGN000186. APG09102 showed 6.2% A>G conversion at position A16; it also showed over 2% base editing at positions A9 and A18. For all fusion proteins tested, position A16 was the most highly edited.

TABLE 6

A > N Editing Rate using guide SGN000194								
Deaminase	A6	A10	A13	A15	A21	A23	A26	A27
APG09982	C	0	0	0	0	0	0	0
	G	0	0.3	0.6	1.5	0	0	0
	T	0	0	0	0	0	0	0
APG03724	C	0	0	0	0	0	0	0
	G	0	0.1	0.3	1	0	0	0
	T	0	0	0	0	0	0	0
APG09949	C	0	0	0	0	0	0	0
	G	0	0.2	0.3	1.6	0	0	0
	T	0	0	0	0	0	0	0
APG08196	C	0	0	0	0	0	0	0
	G	0.1	0.4	0.1	0.9	0	0	0
	T	0	0	0	0	0	0	0
APG06333	C	0	0	0	0	0	0	0
	G	0	0.2	0.3	1	0	0	0
	T	0	0	0	0	0	0	0
APG06489	C	0	0	0	0	0	0	0
	G	0	0.4	0.2	1.1	0	0	0
	T	0	0	0	0	0	0	0
APG08449	C	0	0	0	0	0	0	0
	G	0.1	0.3	0.4	1.8	0	0	0
	T	0	0	0	0	0	0	0
APG05174	C	0	0	0	0	0	0	0
	G	0.1	0.1	0.3	0.9	0	0	0
	T	0	0	0	0	0	0	0
APG09102	C	0	0	0	0	0	0	0
	G	0	0.2	0.7	1.6	0	0	0
	T	0	0	0	0	0	0	0
APG05723	C	0	0	0	0	0	0	0
	G	0	0	0.1	0.9	0	0	0
	T	0	0	0	0	0	0	0

[0752] With SGN00194, all fusion proteins showed 0.9% -1.8% A>G editing at position A15. No detectable editing was seen in positions A21, A23, A26 and A27.

TABLE 7

TABLE 7-continued

Deaminase	A > N Editing Rate using guide SGN000930															
	A2	A4	A5	A8	A9	A10	A14	A15	A16	A20	A21	A23	A24	A26	A27	A29
APG08196	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0.1	0	0.2	0.7	0.3	0.2	0.4	0.1	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG06333	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0.1	0.1	0	0.3	0.4	0.3	0.9	0.2	0.1	0.1	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG06489	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0.3	0.1	0.2	0.8	0.3	0.4	0.6	0	0.1	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG08449	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0.1	0.1	0.3	0.6	0.4	0.2	0.4	0.1	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG05174	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0.1	0.2	0.8	0.3	0.4	0.2	0.2	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG09102	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	0	0.9	0.1	0.1	0.6	0.3	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
APG05723	C	0	0	0	0	0.1	0.1	1.2	0.6	0.2	0.5	0	0	0	0	0
	G	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[0753] A14 was the most highly edited position in SGN000930 with all fusion proteins tested. The editing rate ranged from 0.3% -1.2% for A>G conversions.

#### Example 2: Fluorescence Assay for Targeted Adenine Base Editing

[0754] A vector harboring Enhanced Green Fluorescent Protein (EGFP) containing a W58x mutation which causes a premature stop codon (GFP-STOP, SEQ ID NO: 47) was constructed such that the W58 codon can be reverted from a stop codon (TGA) to the wild-type tryptophan (TGG) residue using an adenine deaminase to alter the third position A to G. Successful A to G conversion results in the expression of EGFP which can be quantified. A second vector capable of expressing a guide RNA which targets the deaminase-RGN fusion protein to the region around the W58x mutation (SEQ ID NO: 48) was also produced.

[0755] This GFP-STOP reporter vector, along with the vectors capable of expressing a deaminase-nRGN fusion protein and the corresponding guide RNA, were transfected into HEK293T cells, using either lipofection or electroporation. For lipofection, cells were seeded at  $1 \times 10^5$  cells/well in 24-well plates the day prior to transfection in growth medium (DMEM+10% Fetal Bovine Serum+1% Penicillin/streptomycin). 500 ng each of the GFP-STOP reporter vector, deaminase-RGN expression vector, and guide RNA expression vector were transfected using Lipofectamine® 3000 reagent (Thermo Fisher Scientific) following manufacturer's instructions. For electroporation, cells were electroporated using the Neon® Transfection System (Thermo Fisher Scientific) following manufacturer's instructions.

[0756] In addition to transient transfection of the fluorescent GFP-STOP reporter, a stable cell line harboring a chromosomally integrated GFP-STOP cassette was generated. Once the stable line was established, for transfection, cells were seeded at  $1 \times 10^5$  cells/well in 24-well plates the day prior to transfection in growth medium (DMEM+10% Fetal Bovine Serum+1% Penicillin/streptomycin). 500 ng each of the deaminase-nRGN expression vector and guide RNA expression vector were transfected using Lipofectamine® 3000 reagent (Thermo Fisher Scientific) following manufacturer's instructions. For electroporation, cells

were electroporated using the Neon® Transfection System (Thermo Fisher Scientific) following manufacturer's instructions.

[0757] 24-48 hours after lipofection or electroporation, the expression of GFP was determined by microscopically surveying the cells for the presence of GFP+ cells. Following visual inspection, the proportion of GFP+ cells versus GFP- cells may be determined. Fluorescence was observed in mammalian cells expressing each of the deaminase-nRGN fusion proteins, indicating the fusion protein successfully targeted to the GFP-STOP mutation and edited the mutation to restore fluorescence of the GFP protein.

[0758] Following microscopic analysis, the cells were lysed in RIPA buffer and the resulting lysate was analyzed on a fluorescence plate reader to determine the fluorescence intensity of GFP (Table 8). A person of skill in the art will appreciate that the cells may be analyzed by flow cytometry or fluorescence activated cell sorting to determine the exact proportions of GFP+ and GFP-cells.

TABLE 8

GFP-STOP assay results	
Deaminase of fusion protein	GFP+ cells detected
APG09982	++
APG03724	++
APG09949	++
APG08196	++
APG06333	+++
APG06489	++
APG08449	++
APG05174	+++
APG09102	++
APG05723	++

N.D = None Detected;

+ = few GFP+ cells detected;

++ = several GFP+ cells detected;

+++ = many GFP+ cells detected

**Example 3: Demonstration of A Base Editing in Mammalian Cells**

**[0759]** The deaminases shown in Table 9 below were produced based on naturally occurring deaminases which were then mutated and selected for adenine deaminase activity in prokaryotic cells.

TABLE 9

Deaminase sequences	
Deaminase	SEQ ID NO.
LPG50140	399
LPG50141	400
LPG50142	401
LPG50143	402
LPG50144	403
LPG50145	404
LPG50146	405
LPG50147	406
LPG50148	407
LPG50149	408
LPG50150	409
LPG50151	410
LPG50152	411
LPG50153	412
LPG50154	413
LPG50155	414
LPG50156	415
LPG50157	416
LPG50158	417
LPG50159	418
LPG50160	419
LPG50161	420
LPG50162	421
LPG50163	422
LPG50164	423
LPG50165	424
LPG50166	425
LPG50167	426
LPG50168	427
LPG50169	428
LPG50170	429
LPG50171	430
LPG50172	431
LPG50173	432
LPG50174	433
LPG50175	434
LPG50176	435
LPG50177	436
LPG50178	437
LPG50179	438
LPG50180	439
LPG50181	440
LPG50182	441

**[0760]** To determine if the deaminases of Table 9 are able to perform adenine base editing in mammalian cells, each deaminase was operably fused to an RGN nickase to produce a fusion protein. Residues predicted to deactivate the RuvC domain of the RGN APG07433.1 (set forth as SEQ ID NO: 41; described in PCT publication WO 2019/236566, incorporated by reference herein) were identified and the RGN was modified to a nickase variant (nAPG07433.1; SEQ ID NO: 42). A nickase variant of an RGN is referred to herein as “nRGN”. It should be understood that any nickase variant of an RGN may be used to produce a fusion protein of the invention.

**[0761]** Deaminase and nRGN nucleotide sequences codon optimized for mammalian expression were synthesized as fusion proteins with an N-terminal nuclear localization tag and cloned into the pTwist CMV (Twist Biosciences)

expression plasmid. Each fusion protein comprises, starting at the amino terminus, the SV40 NLS (SEQ ID NO: 43) operably linked at the C-terminal end to 3× FLAG Tag (SEQ ID NO: 44), operably linked at the C-terminal end to a deaminase, operably linked at the C-terminal end to a peptide linker (SEQ ID NO: 442), operably linked at the C-terminal end to an nRGN (for example, nAPG07433.1, which is SEQ ID NO: 42), finally operably linked at the C-terminal end to the nucleoplasmin NLS (SEQ ID NO: 46). The nAPG07433.1 and peptide linker nucleotide sequences codon optimized for mammalian expression are set forth as SEQ ID NOs: 486 and 487, respectively. Table 10 shows the fusion proteins produced and tested for activity. All fusion proteins comprise at least one NLS and a 3× FLAG Tag, as described above.

TABLE 10

Fusion protein sequences with N-terminus SV40 NLS, 3X FLAG Tag and C-terminus Nucleoplasmin NLS	
Fusion Protein	SEQ ID
LPG50140-nAPG07433.1	488
LPG50141-nAPG07433.1	489
LPG50142-nAPG07433.1	490
LPG50143-nAPG07433.1	491
LPG50144-nAPG07433.1	492
LPG50145-nAPG07433.1	493
LPG50146-nAPG07433.1	494
LPG50147-nAPG07433.1	495
LPG50148-nAPG07433.1	496
LPG50149-nAPG07433.1	497
LPG50150-nAPG07433.1	498
LPG50151-nAPG07433.1	499
LPG50152-nAPG07433.1	500
LPG50153-nAPG07433.1	501
LPG50154-nAPG07433.1	502
LPG50155-nAPG07433.1	503
LPG50156-nAPG07433.1	504
LPG50157-nAPG07433.1	505
LPG50158-nAPG07433.1	506
LPG50159-nAPG07433.1	507
LPG50160-nAPG07433.1	508
LPG50161-nAPG07433.1	509
LPG50162-nAPG07433.1	510
LPG50163-nAPG07433.1	511
LPG50164-nAPG07433.1	512
LPG50165-nAPG07433.1	513
LPG50166-nAPG07433.1	514
LPG50167-nAPG07433.1	515
LPG50168-nAPG07433.1	516
LPG50169-nAPG07433.1	517
LPG50170-nAPG07433.1	518
LPG50171-nAPG07433.1	519
LPG50172-nAPG07433.1	520
LPG50173-nAPG07433.1	521
LPG50174-nAPG07433.1	522
LPG50175-nAPG07433.1	523
LPG50176-nAPG07433.1	524
LPG50177-nAPG07433.1	525
LPG50178-nAPG07433.1	526
LPG50179-nAPG07433.1	527
LPG50180-nAPG07433.1	528
LPG50181-nAPG07433.1	529
LPG50182-nAPG07433.1	530

**[0762]** Expression plasmids comprising an expression cassette encoding for a sgRNA were also produced. Human genomic target sequences and the sgRNA sequences for guiding the fusion proteins to the genomic targets are indicated in Table 11.

TABLE 11

Guide RNA sequences				
sgRNA ID	Target sequence	sgRNA sequence	Forward Primer for amplification	Reverse Primer for amplification
SGN000139	537	531	543	549
SGN000143	538	532	544	550
SGN000186	539	533	545	551
SGN000194	540	534	546	552
SGN000930	541	535	547	553
SGN001681	542	536	548	554

[0763] 500 ng of plasmid comprising an expression cassette comprising a coding sequence for a fusion protein shown in Table 10 and 500 ng of plasmid comprising an expression cassette encoding for an sgRNA shown in Table 11 were co-transfected into HEK293 FT cells at 75-90% confluence in 24-well plates using Lipofectamine 2000 reagent (Life Technologies). Cells were then incubated at 37° C. for 72 h. Following incubation, genomic DNA was then extracted using NucleoSpin 96 Tissue (Macherey-Nagel) following the manufacturer's protocol. The genomic region flanking the targeted genomic site was PCR amplified using the primers in Table 11 and products were purified using ZR-96 DNA Clean and Concentrator (Zymo Research) following the manufacturer's protocol. The purified PCR products underwent Next Generation Sequencing on Illumina MiSeq. Typically, 100,000 of 250 bp paired-end reads (2×100,000 reads) are generated per amplicon. The reads were analyzed using CRISPResso (Pinello, et al. 2016 *Nature Biotech.*, 34:695-697) to calculate the rates of editing. Output alignments were analyzed for INDEL formation or introduction of specific adenine mutations.

[0764] Table 12 shows all of the adenine base editing for each adenine deaminase fusion in Table 10 and a guide RNA from Table 12. Tables 13-27 show the specific nucleotide mutation profile for select exemplary samples. The editing rate for adenines within or proximal to the target sequence is indicated. "A5" indicates, for example, an adenine at position 5 of the target sequence. The position of each nucleotide in the target sequence was determined by numbering the first nucleotide in the target sequence closest to the PAM (which is 3' of the target for APG07433.1) as position 1, and the position number increases in the 5' direction away from the PAM sequence. The tables also show which nucleotide the adenine was changed to, and at what rate. For example, Table 13 shows that for the LPG50148-nAPG07433.1 fusion protein, the adenine at position 13 was mutated to a guanine at a rate of 9.7%.

TABLE 12

Estimate of base editing rates for each adenine deaminase		
Deaminase	SGN	% Mutated Reads
LPG50140	SGN001681	30.01%
LPG50140	SGN000139	6.91%
LPG50140	SGN000143	16.09%
LPG50140	SGN000186	18.76%
LPG50140	SGN000194	9.77%
LPG50140	SGN000930	3.51%
LPG50141	SGN001681	21.37%
LPG50141	SGN000139	2.43%
LPG50141	SGN000143	6.93%

TABLE 12-continued

Estimate of base editing rates for each adenine deaminase		
Deaminase	SGN	% Mutated Reads
LPG50141	SGN000186	9.79%
LPG50141	SGN000194	4.45%
LPG50141	SGN000930	5.29%
LPG50142	SGN001681	34.19%
LPG50142	SGN000139	3.10%
LPG50142	SGN000143	8.67%
LPG50142	SGN000186	14.12%
LPG50142	SGN000194	10.04%
LPG50142	SGN000930	6.78%
LPG50143	SGN001681	20.62%
LPG50143	SGN000139	1.99%
LPG50143	SGN000143	6.09%
LPG50143	SGN000186	10.58%
LPG50143	SGN000194	5.60%
LPG50143	SGN000930	3.98%
LPG50144	SGN001681	28.26%
LPG50144	SGN000139	3.55%
LPG50144	SGN000143	5.77%
LPG50144	SGN000186	12.22%
LPG50144	SGN000194	6.40%
LPG50144	SGN000930	5.81%
LPG50145	SGN001681	29.23%
LPG50145	SGN000139	2.53%
LPG50145	SGN000143	3.75%
LPG50145	SGN000186	9.93%
LPG50145	SGN000194	3.98%
LPG50145	SGN000930	3.84%
LPG50146	SGN001681	32.53%
LPG50146	SGN000139	5.95%
LPG50146	SGN000143	11.30%
LPG50146	SGN000186	17.78%
LPG50146	SGN000194	7.38%
LPG50146	SGN000930	7.13%
LPG50147	SGN001681	49.10%
LPG50147	SGN000139	3.26%
LPG50147	SGN000143	8.59%
LPG50147	SGN000186	12.61%
LPG50147	SGN000194	8.80%
LPG50147	SGN000930	4.96%
LPG50148	SGN001681	49.39%
LPG50148	SGN000139	10.80%
LPG50148	SGN000143	12.49%
LPG50148	SGN000186	32.65%
LPG50148	SGN000194	16.60%
LPG50148	SGN000930	7.61%
LPG50149	SGN001681	27.62%
LPG50149	SGN000139	2.83%
LPG50149	SGN000143	9.33%
LPG50149	SGN000186	22.12%
LPG50149	SGN000194	7.94%
LPG50149	SGN000930	7.06%
LPG50150	SGN001681	28.46%
LPG50150	SGN000139	3.06%
LPG50150	SGN000143	6.00%
LPG50150	SGN000186	23.67%
LPG50150	SGN000194	9.47%
LPG50150	SGN000930	5.41%
LPG50151	SGN001681	3.01%
LPG50151	SGN000139	0%
LPG50151	SGN000143	1.53%
LPG50151	SGN000186	7.76%
LPG50151	SGN000194	1.43%
LPG50151	SGN000930	0%
LPG50152	SGN001681	26.06%
LPG50152	SGN000139	2%
LPG50152	SGN000143	3%
LPG50152	SGN000186	18%
LPG50152	SGN000194	3%
LPG50152	SGN000930	6%
LPG50153	SGN001681	1.12%
LPG50153	SGN000139	0%
LPG50153	SGN000143	0%
LPG50153	SGN000186	0%

TABLE 12-continued

Estimate of base editing rates for each adenine deaminase		
Deaminase	SGN	% Mutated Reads
LPG50153	SGN000194	1%
LPG50153	SGN000930	0%
LPG50154	SGN001681	2.26%
LPG50154	SGN000139	0%
LPG50154	SGN000143	0%
LPG50154	SGN000186	0%
LPG50154	SGN000194	1%
LPG50154	SGN000930	0%
LPG50155	SGN001681	14.91%
LPG50155	SGN000139	2%
LPG50155	SGN000143	4%
LPG50155	SGN000186	17%
LPG50155	SGN000194	7%
LPG50155	SGN000930	5%
LPG50156	SGN001681	11.19%
LPG50156	SGN000139	3.79%
LPG50156	SGN000143	6.44%
LPG50156	SGN000186	12.69%
LPG50156	SGN000194	6.87%
LPG50156	SGN000930	4.10%
LPG50157	SGN001681	20.66%
LPG50157	SGN000139	3.37%
LPG50157	SGN000143	6.91%
LPG50157	SGN000186	12.15%
LPG50157	SGN000194	9.98%
LPG50157	SGN000930	5.55%
LPG50158	SGN001681	1.56%
LPG50158	SGN000139	0%
LPG50158	SGN000143	1.15%
LPG50158	SGN000186	4.91%
LPG50158	SGN000194	1.73%
LPG50158	SGN000930	0%
LPG50159	SGN001681	5.85%
LPG50159	SGN000139	0%
LPG50159	SGN000143	2.78%
LPG50159	SGN000186	6.99%
LPG50159	SGN000194	4.40%
LPG50159	SGN000930	2.60%
LPG50160	SGN001681	22.20%
LPG50160	SGN000139	4%
LPG50160	SGN000143	8%
LPG50160	SGN000186	16%
LPG50160	SGN000194	5%
LPG50160	SGN000930	6%
LPG50161	SGN001681	1.47%
LPG50161	SGN000139	0%
LPG50161	SGN000143	0%
LPG50161	SGN000186	0%
LPG50161	SGN000194	0%
LPG50161	SGN000930	0%
LPG50162	SGN001681	21.73%
LPG50162	SGN000139	2%
LPG50162	SGN000143	5%
LPG50162	SGN000186	14%
LPG50162	SGN000194	6%
LPG50162	SGN000930	5%
LPG50163	SGN001681	12.80%
LPG50163	SGN000139	0%
LPG50163	SGN000143	2%
LPG50163	SGN000186	10%
LPG50163	SGN000194	4%
LPG50163	SGN000930	3%
LPG50164	SGN001681	4.28%
LPG50164	SGN000139	0%
LPG50164	SGN000143	3.36%
LPG50164	SGN000186	7.38%
LPG50164	SGN000194	2.73%

TABLE 12-continued

Estimate of base editing rates for each adenine deaminase		
Deaminase	SGN	% Mutated Reads
LPG50164	SGN000930	1.47%
LPG50165	SGN001681	25.66%
LPG50165	SGN000139	2%
LPG50165	SGN000143	5.11%
LPG50165	SGN000186	9.88%
LPG50165	SGN000194	3.97%
LPG50165	SGN000930	3.18%
LPG50166	SGN000139	2%
LPG50166	SGN000143	4%
LPG50166	SGN000186	8%
LPG50166	SGN000194	2%
LPG50166	SGN000930	4%
LPG50167	SGN001681	20.56%
LPG50167	SGN000139	2%
LPG50167	SGN000143	4%
LPG50167	SGN000186	8%
LPG50167	SGN000194	5%
LPG50167	SGN000930	4%
LPG50168	SGN001681	13.81%
LPG50168	SGN000139	2%
LPG50168	SGN000143	3%
LPG50168	SGN000186	7%
LPG50168	SGN000194	2%
LPG50168	SGN000930	3%
LPG50169	SGN001681	25.73%
LPG50169	SGN000139	4%
LPG50169	SGN000143	25.73%
LPG50169	SGN000186	12.87%
LPG50169	SGN000194	8%
LPG50169	SGN000186	13%
LPG50169	SGN000194	9%
LPG50169	SGN000930	8%
LPG50170	SGN001681	12.65%
LPG50170	SGN000139	1.50%
LPG50170	SGN000143	3.14%
LPG50170	SGN000186	12.16%
LPG50170	SGN000194	2.76%
LPG50170	SGN000930	4.10%
LPG50171	SGN001681	27.16%
LPG50171	SGN000139	1.75%
LPG50171	SGN000143	6.14%
LPG50171	SGN000186	12.65%
LPG50171	SGN000194	5.60%
LPG50171	SGN000930	4.55%
LPG50172	SGN001681	1.78%
LPG50172	SGN000139	0%
LPG50172	SGN000143	0%
LPG50172	SGN000186	0%
LPG50172	SGN000194	0%
LPG50172	SGN000930	0%
LPG50173	SGN001681	12.64%
LPG50173	SGN000139	1.00%
LPG50173	SGN000143	3.23%
LPG50173	SGN000186	7.88%
LPG50173	SGN000194	2.66%
LPG50173	SGN000930	1.77%
LPG50174	SGN001681	14.11%
LPG50174	SGN000139	0%
LPG50174	SGN000143	3%
LPG50174	SGN000186	8%
LPG50174	SGN000194	2%
LPG50174	SGN000930	3%
LPG50175	SGN001681	22.29%
LPG50175	SGN000139	4%
LPG50175	SGN000143	9%
LPG50175	SGN000186	14%
LPG50175	SGN000194	13%
LPG50175	SGN000930	5%
LPG50176	SGN001681	9.52%
LPG50176	SGN000139	0%

TABLE 12-continued

Estimate of base editing rates for each adenine deaminase			
Deaminase	SGN	% Mutated	Reads
LPG50176	SGN000143	2%	
LPG50176	SGN000186	7%	
LPG50176	SGN000194	2%	
LPG50176	SGN000930	0%	
LPG50177	SGN001681	7.98%	
LPG50177	SGN000139	2%	
LPG50177	SGN000143	4%	
LPG50177	SGN000186	11%	
LPG50177	SGN000194	3%	
LPG50177	SGN000930	9%	
LPG50178	SGN000139	2.00%	
LPG50178	SGN000143	6.19%	
LPG50178	SGN000186	12.94%	
LPG50178	SGN000194	5.51%	
LPG50178	SGN000930	3.95%	
LPG50179	SGN001681	23.35%	
LPG50179	SGN000139	2.00%	
LPG50179	SGN000143	5.08%	
LPG50179	SGN000186	12.50%	
LPG50179	SGN000194	4.49%	
LPG50179	SGN000930	4.62%	
LPG50180	SGN001681	1.80%	
LPG50180	SGN000139	0%	
LPG50180	SGN000143	0%	
LPG50180	SGN000186	0%	
LPG50180	SGN000194	0%	
LPG50180	SGN000930	0%	
LPG50181	SGN001681	7.93%	
LPG50181	SGN000139	2.88%	
LPG50181	SGN000143	3.78%	
LPG50181	SGN000186	12.56%	
LPG50181	SGN000194	3.39%	
LPG50181	SGN000930	1.20%	
LPG50182	SGN001681	16.49%	
LPG50182	SGN000139	1.00%	
LPG50182	SGN000143	5%	
LPG50182	SGN000186	9%	
LPG50182	SGN000194	6%	
LPG50182	SGN000930	3%	

TABLE 13

TABLE 13

A > N Editing Rate using deaminase LPG50148 and guide SGN000139					
SGN000139					
	A5	A12	A13	A20	A22
LPG50148	C	0	0	0	0.1
	G	0	2.2	9.7	0.2
	T	0	0	0	0

[0765] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A12 and A13. LPG50148 showed over 9% editing at position A13.

TABLE 14

A > N Editing Rate using deaminase  
LPG50148 and guide SGN000143

[0766] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A11 and A14. LPG50148 showed over 11% editing at position A11.

TABLE 15

A > N Editing Rate using deaminase  
LPG50148 and guide SGN000186

[0767] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A16 and A18. LPG50148 showed over 23% editing at positions A9 and A16.

TABLE 16

[0768] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A13 and A15. LPG50148 showed over 12% editing at positions A13 and A15.

TABLE 17

[0769] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A10, A14, A15, A16, A20 and A21. LPG50148 showed over 2% editing at positions A10, A14, A16, A20 and A21.

TABLE 18

A > N Editing Rate using deaminase LPG50146 and guide SGN000139						
SGN000139						
	A5	A12	A13	A20	A22	
LPG50146	C	0	0	0	0.4	0.1
	G	0	2.1	4.1	0	0
	T	0	0	0	0	0

TABLE 21

A > N Editing Rate using deaminase LPG50146 and guide SGN000194						
SGN000194						
	A6	A10	A13	A15	A21	A23
LPG50146	C	0	0	0	0	0
	G	0	1.8	3.2	4.5	0
	T	0	0	0	0	0

[0773] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A13 and A15. LPG50146 showed over 3% editing at positions A13 and A15.

TABLE 22

A > N Editing Rate using deaminase LPG50146 and guide SGN000930																		
SGN000930																		
	A2	A4	A5	A8	A9	A10	A14	A15	A16	A20	A21	A23	A24	A26	A27	A29	A30	
LPG50146	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	G	0	0	0	0.1	0.1	0.7	2.9	2.6	2.4	1	0.8	0	0	0	0	0	
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

[0770] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A12 and A13. LPG50146 showed over 4% editing at position A13.

TABLE 19

A > N Editing Rate using deaminase LPG50146 and guide SGN000143								
SGN000143								
	A1	A4	A6	A9	A11	A14	A19	A30
LPG50146	C	0	0	0	0	0	0	0
	G	0	0	0	0.8	8.4	5	0
	T	0	0	0	0	0	0	0

[0771] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A11 and A14. LPG50146 showed over 8% editing at position A11.

TABLE 20

A > N Editing Rate using deaminase LPG50146 and guide SGN000186						
SGN000186						
	A9	A16	A18	A22	A25	A28
LPG50146	C	0	0	0	0.2	0
	G	7.4	13.4	3.1	0.1	0
	T	0	0	0	0	0

[0772] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A16 and A18. LPG50146 showed over 13% editing at position A16.

TABLE 23

A > N Editing Rate using deaminase LPG50140 and guide SGN000139				
SGN000139				
	A5	A12	A13	A20
LPG50140	C	0	0	0.4
	G	0	0.5	5.5
	T	0	0	0

[0775] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A12 and A13. LPG50140 showed over 5% editing at position A13.

TABLE 24

A > N Editing Rate using deaminase LPG50140 and guide SGN000143							
SGN000143							
	A1	A4	A6	A9	A11	A14	A19
LPG50140	C	0	0	0	0	0	0
	G	0	0	0	1.2	14	5.6
	T	0	0	0	0	0	0

[0776] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A11 and A14. LPG50140 showed 14% editing at position A11.

TABLE 25

		A > N Editing Rate using deaminase LPG50140 and guide SGN000186						
		SGN000186						
		A9	A16	A18	A22	A25	A28	A30
LPG50140	C	0	0	0	0	0.2	0	0
	G	9.4	15	1.7	0	0	0	0
	T	0	0	0	0	0	0	0

[0777] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A9, A16 and A18. LPG50140 showed over 9% editing at positions A9 and A16.

TABLE 26

		A > N Editing Rate using deaminase LPG50140 and guide SGN000194							
		SGN000194							
		A6	A10	A13	A15	A21	A23	A26	A27
LPG50140	C	0	0	0	0	0	0	0	0
	G	0	0	6.7	7.8	0	0	0	0
	T	0	0	0	0	0	0	0	0

[0778] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A13 and A15. LPG50140 showed over 6% editing at positions A13 and A15.

TABLE 27

		A > N Editing Rate using deaminase LPG50140 and guide SGN000930																
		SGN000930																
		A2	A4	A5	A8	A9	A10	A14	A15	A16	A20	A21	A23	A24	A26	A27	A29	A30
LPG50140	C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	0.4	1.4	0.6	1.1	0.4	0.5	0	0	0	0	0	0
	T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[0779] LPG50140, LPG50146, and LPG50148 showed detectable A>G conversion at positions A10, A14, A15, A16, A20 and A21. LPG50140 showed over 1% editing at positions A14 and A16.

8[0780] Table 28 below shows the average editing rates for LPG50148-nAPG07433.1 at several guides tested in HEK293T cells by lipofection of two plasmids. The base editor was encoded on one plasmid and the guide RNA was encoded on a second plasmid. Total substitution rate in the target is used to measure the base editing rate.

TABLE 28

		Average Editing Rate for LPG50148-nAPG07433.1		
Gene	SGN	Average % Substitution rate	N	
Gene A	SGN000139	10.8	1	
Gene A	SGN000143	29.65	2	
Gene B	SGN000487	34.68	2	
Gene B	SGN000488	39.94	1	
Gene B	SGN001061	9.18	2	
Gene B	SGN001062	32.77	1	
Gene B	SGN001270	8.34	3	
Gene B	SGN001946	5.1	1	

TABLE 28-continued

Average Editing Rate for LPG50148-nAPG07433.1			
Gene	SGN	Average % Substitution rate	N
Gene B	SGN001947	16.43	1
Gene B	SGN001948	0.46	1
Gene B	SGN001949	1.44	1
Gene B	SGN001950	10.96	1
Gene B	SGN001951	5.38	1
Gene B	SGN001952	6.29	1
Gene B	SGN001953	5.28	1
Gene B	SGN001954	7.95	1
Gene B	SGN001955	7.83	1
Gene B	SGN001956	4.78	1
Gene B	SGN001959	1.43	1
Gene B	SGN001960	17.4	1
Gene B	SGN001961	1.46	1
Gene B	SGN001962	1.62	1
Gene B	SGN001963	11.31	1
Gene B	SGN001964	2.03	1
Gene B	SGN001965	9.3	1
Gene B	SGN001966	1.51	1
CFTR	SGN001101	17.06	1
Gene D	SGN001196	14.58	1
Gene D	SGN001199	42.05	1
Gene E	SGN001681	48.85	1
Gene F	SGN000169	55.13	2
Gene F	SGN000173	47.13	1
Gene G	SGN000412	16.58	1
Gene G	SGN000414	14.5	2
Gene G	SGN001259	24.16	1
Gene G	SGN001274	10.45	2

TABLE 28-continued

Average Editing Rate for LPG50148-nAPG07433.1			
Gene	SGN	Average % Substitution rate	N
Gene G	SGN001275	5.25	1
Gene H	SGN000186	32.65	1
Gene I	SGN000754	30.76	1
Gene I	SGN000909	21.57	2
Gene I	SGN000927	3.8	1
Gene I	SGN000928	28.77	1
Gene I	SGN000929	17.58	2
Gene I	SGN000949	26.43	1
Gene I	SGN001268	16.64	2
Gene I	SGN001269	6.42	1
Gene I	SGN001967	1.45	1
Gene I	SGN001968	5.61	1
Gene I	SGN001973	5.14	1
Gene I	SGN001975	0.16	1
Gene I	SGN001976	0.62	1
Gene I	SGN001977	0.65	1
Gene I	SGN001978	3.09	1
Gene I	SGN001981	2.34	1

[0781] LPG50148-nAPG07433.1 shows editing at many different guides across the genome.

[0782] Table 29 shows the editing rates of adenine bases in each guide from LPG50148-nAPG07433.1. Only the adenine positions are shown below. The rate of adenine conversion is the average of multiple experiments when appropriate.

TABLE 29

Editing rate of A nucleotides in mammalian cells for top 10 guides												
SGN	Position											
	A1	A3	A4	A5	A6	A7	A8	A9	A10	A11	A13	A14
SGN001681				13				47				
SGN000169				0.2		1.3	17			22		
SGN001199					3.5						42	
SGN000186								24				
SGN000754				0		0		1.3				6.1
SGN000143	0		0		0.4			4.4		27		17
SGN000928	0.3			0.2	0.3		6.1					
SGN000487	0.2	0.2						12			25	
SGN001259			0					12				
SGN001062				0		0.7		0				
Position												
SGN	A15	A16	A17	A18	A19	A20	A21	A22	A24	A25		
	SGN001681											
SGN000169		43			11				1.7			
SGN001199												
SGN000186	29			4.1					0.2			0.4
SGN000754		29										
SGN000143					0.3							
SGN000928				26								
SGN000487	8.7			7.6			14					
SGN001259		16										
SGN001062	10			5.8		13		2.4	1	0.1	0	

[0783] LPG50148-nAPG07433.1 shows adenine base editing in positions 6 through 21 in the target region depending on the guide RNA used. Editing rates vary by guide RNAs used.

#### Example 4: Correction of Class I Cystic Fibrosis Nonsense Mutations

##### Example 4.1: Identification of RGNs and Guide RNAs

[0784] Cystic fibrosis is generally caused by deleterious mutations in the CFTR gene (SEQ ID NO: 51). Six of the most common nonsense mutations are G542X, W1282X,

R553X, R1162X, E60X, R785X, and Q493X. Each of these stop mutations could be edited to restore a coding codon by an RGN-deaminase fusion protein described herein. To target each mutation, the following must be determined: 1) an RGN which has a PAM recognition site proximal to the nonsense mutation; and 2) a guide RNA which optimally targets the RGN-deaminase fusion protein to the target DNA. Table 30 below shows nickase variants of RGNs which possess PAMs that are proximal to each of the six nonsense mutations and the number of guide RNAs which can be used for each RGN. Table 31 describes the genetic loci for each guide RNA. The PAM recognition site for each genetic locus is underlined. The target sequence for the guide RNA and the guide RNA sequence itself are also indicated.

TABLE 30

RGN nickases and number of guide RNAs for nonsense mutations in CFTR						
RGN nickase	SEQ ID NO. for RGN nickase	E60X	G542X	Q493X	R1162X	R553X
nAPG00969	52		2		2	
nAPG07433.1	42		1			3
nAPG06646	53	6	4	2	3	7
nAPG09748	54	1	1	4		1
nAPG09882	55	4	3	5	5	3
nAPG03850	56	2	2	1	3	3
nAPG07553	57	1	1	1	1	2
nAPG05586	58	1	1		3	1
nAPG01604	59			2	1	2

TABLE 31

guide RNAs for nonsense mutations in CFTR				
Guide ID	Genetic locus	Genetic locus (SEQ ID NO.)	Target (SEQ ID NO.)	gRNA (SEQ ID NO.)
E60X nAPG06646 Target 1	AATGAGTTAGGATTTCTTGAGGCCAGCTATCTATCCCATT <u>CTCTGCAAAAGAATAAAAAGT</u>	62	80	98
E60X nAPG06646 Target 2	ATTAATGAGTTAGGATTTCTTGAGGCCAGCTATCTATCCC <u>ATTCTCTGCAAAAGAATAAAA</u>	63	81	99
E60X nAPG06646 Target 3	GCATTAATGAGTTAGGATTTCTTGAGGCCAGCTATCTATC <u>CCATTCTGCAAAAGAATAAA</u>	64	82	100
E60X nAPG06646 Target 4	AAGGGCATTAAATGAGTTAGGATTTCTTGAGGCCAGCTATC <u>TATCCCATTCTGCAAAAGA</u>	65	83	101
E60X nAPG06646 Target 5	GAAGGGCATTAAATGAGTTAGGATTTCTTGAGGCCAGCTATC <u>CTATCCCATTCTGCAAAAG</u>	66	84	102
E60X nAPG06646 Target 6	CGAAGGGCATTAAATGAGTTAGGATTTCTTGAGGCCAGCTA <u>TCTATCCCATTCTGCAAAAGA</u>	67	85	103
E60X nAPG09882 Target 1	GAGTTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCTC <u>TGCAAAAGAATAAAAAGTGGG</u>	68	86	104
E60X nAPG09882 Target 2	TGAGTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAGTGG</u>	69	87	105
E60X nAPG09882 Target 3	ATGAGTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCTC <u>TCTGCAAAAGAATAAAAAGTGT</u>	70	88	106
E60X nAPG09882 Target 4	AGGGCATTAAATGAGTTAGGATTTCTTGAGGCCAGCTATCT <u>ATCCCATTCTGCAAAAGAAA</u>	71	89	107
E60X nAPG00969 Target 1	GTTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAGTGGGAC</u>	72	90	108
E60X nAPG00969 Target 2	AGTTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAGTGGGA</u>	73	91	109
E60X nAPG03850 Target 1	GGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAGTGGGAC</u>	74	92	110
E60X nAPG03850 Target 2	AGTTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAG</u>	75	93	111
E60X nAPG07433.1 Target 1	GAAGGGCATTAAATGAGTTAGGATTTCTTGAGGCCAGCTATC <u>CTATCCCATTCTGCAAAAG</u>	76	94	112
E60X nAPG09748 Target 1	GTCCCACCTTTATTCTTGAGAGAATGGGATAGATAGCTGG <u>CTTCAAAGAAAATCC</u>	77	95	113
E60X nAPG07553 Target 1	AGTTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAG</u>	78	96	114
E60X nAPG05586 Target 1	TTTAGGATTTCTTGAGGCCAGCTATCTATCCCATTCT <u>CTGCAAAAGAATAAAAAGTGT</u>	79	97	115
G542X nAPG06646 Target 1	CGTTGACCTCACTCAGTGTGATTCCACCTCTCAAAGAACTAT <u>ATTGTCTCTGCAAACTT</u>	116	128	140
G542X nAPG06646 Target 2	GACCTCACTCAGTGTGATTCCACCTCTCAAAGAACTATATTG <u>TCTTCTCTGCAAACTTGGGAG</u>	117	129	141
G542X nAPG06646 Target 3	CCTCCACTCAGTGTGATTCCACCTCTCAAAGAACTATATTG <u>TCTCTCTGCAAACTTGGGAG</u>	118	130	142
G542X nAPG06646 Target 4	CCACTCAGTGTGATTCCACCTCTCAAAGAACTATATTGTC <u>CTCTGCAAACTTGGGAGATGTC</u>	119	131	143
G542X nAPG09882 Target 1	TCTTGCTCGTTGACCTCACTCAGTGTGATTCCACCTCT <u>CAAAGAATATTGTCTGCAAA</u> <u>GAACATATATTGTCTCTG</u>	120	132	144

TABLE 31-continued

## guide RNAs for nonsense mutations in CFTR

Guide ID	Genetic locus	Genetic locus (SEQ ID NO.)	Target (SEQ ID NO.)	gRNA (SEQ ID NO.)
G542X nAPG09882 Target 2	TTGCTCGTTGACCTCCACTCAGTGTGATTCCACCTCTCAAAGA <u>ACTATATTGCTTCCTGCA</u>	121	133	145
G542X nAPG09882 Target 3	CACTCAGTGTGATTCCACCTCTCAAAGAACTATATTGTCTTC <u>TCTGCAAAC</u> ACTGGAGATGTCC	122	134	146
G542X nAPG03850 Target 1	TGACCTCCACTCAGTGTGATTCCACCTCTCAAAGAACTATATT <u>GTCTTCTCTGCAAC</u>	123	135	147
G542X nAPG03850 Target 2	TCAGTGTGATTCCACCTCTCAAAGAACTATATTGTCTTC <u>GCAAAC</u> TTGGAGATGT	124	136	148
G542X nAPG09748 Target 1	AGAGAAAAGACAATATA <u>AGTTCTTGAGAAGGTGGAATCACACTGA</u> GTGGAGGTCAACGAGC	125	137	149
G542X nAPG07553 Target 1	TCAGTGTGATTCCACCTCTCAAAGAACTATATTGTCTTC <u>GCAAAC</u> TTGGAGATGT	126	138	150
G542X nAPG05586 Target 1	CGTTGACCTCCACTCAGTGTGATTCCACCTCTCAAAGAACTAT <u>ATTGTCTTC</u> CTGCA	127	139	151
Q493X nAPG09882 Target 1	GATATTTCTTAATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGA</u> <u>ACAGAATGAAATTCTTCCAC</u>	152	169	186
Q493X nAPG09882 Target 2	ATATTTCTTAATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGA</u> <u>ACAGAATGAAATTCTTCCACT</u>	153	170	187
Q493X nAPG09882 Target 3	TTTCTTAATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGA</u> <u>GAATGAAATTCTTCCACTGTG</u>	154	171	188
Q493X nAPG09882 Target 4	TTTCTTAATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGA</u> <u>AATGAAATTCTTCCACTGTG</u>	155	172	189
Q493X nAPG09882 Target 5	TTCTTAA <u>ATGGTGC CAGGCATAATCCAGGAAA</u> ACTAAGAACAGA <u>ATGAAATTCTTCCACTGTG</u>	156	173	190
Q493X nAPG09748 Target 1	TAAGCACAGTGGAAA <u>ATTTCATTCTGTTCTTAGTTTC</u> CCTGGA TTATGCCTGGCACCAT	157	174	191
Q493X nAPG09748 Target 2	AAGCACAGTGGAAA <u>ATTTCATTCTGTTCTTAGTTTC</u> CCTGGAT TATGCCTGGCACCAT	158	175	192
Q493X nAPG09748 Target 3	ACAGTGGAAA <u>ATTTCATTCTGTTCTTAGTTTC</u> CCTGGATTATG CCTGGCACCATAAAG	159	176	193
Q02VAPC007/0+A	GGAAA <u>ATTCATTCTGTTCTTAGTTTC</u> CCTGGATTATGCCTGG CACCATAAAGAAAAT	160	177	194
Q493X nAPG00969 Target 1	GATATTTCTTAATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGA</u> <u>ACAGAATGAAATTCTTCC</u> AC	161	178	195
Q493X nAPG00969 Target 2	TTCTTAA <u>ATGGTGC CAGGCATAATCCAGGAAA</u> ACTAAGAACAGA <u>ATGAAATTCTTCCACTGTG</u> CT	162	179	196
Q493X nAPG06646 Target 1	TTAA <u>ATGGTGC CAGGCATAATCCAGGAAA</u> ACTAAGAACAGAATG <u>AAATTCTTCCACTGTG</u> CTT	163	180	197
Q493X nAPG06646 Target 2	AATGGTGC CAGGCATAATCCAGGAAA <u>ACTAAGAACAGAATGAAA</u> <u>TTCTTCCACTGTG</u> CTTAATT	164	181	198
Q493X nAPG01604 Target 1	TTCTTAA <u>ATGGTGC CAGGCATAATCCAGGAAA</u> ACTAAGAACAGA <u>ATGAAATTCTTCCACT</u>	165	182	199
Q493X nAPG01604 Target 2	TTAA <u>ATGGTGC CAGGCATAATCCAGGAAA</u> ACTAAGAACAGA <u>ATGA</u> <u>AATTCTTCCACTGTG</u> CT	166	183	200

TABLE 31-continued

## guide RNAs for nonsense mutations in CFTR

Guide ID	Genetic locus	Genetic locus (SEQ ID NO.)	Target (SEQ ID NO.)	gRNA (SEQ ID NO.)
Q493X nAPG03850 Target 1	CTTTAATGGTGCAGGCATAATCCAGGAAAACTAAGAACAGAAT <u>GAAATTCTTCACTGT</u>	167	184	201
Q493X nAPG07553 Target 1	CTTTAATGGTGCAGGCATAATCCAGGAAAACTAAGAACAGAAT <u>GAAATTCTTCACTGT</u>	168	185	202
R553X nAPG06646 Target 1	CCAATAATTAGTTATTCACCTTGCTAAAGAAATTCTTGCTCATT <u>GACCTCCACTCAGTGTGATT</u> C	203	219	235
R553X nAPG06646 Target 2	CAATAATTAGTTATTCACCTTGCTAAAGAAATTCTTGCTCATT <u>GACCTCCACTCAGTGTGATT</u> C	204	220	236
R553X nAPG06646 Target 3	ATAATTAGTTATTCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGATTCCA</u> C	205	221	237
R553X nAPG06646 Target 4	AATTAGTTATT <u>TCACCTTGCTAAAGAAATTCTTGCTCATTGACCTCCACTCAGTGT</u> CCACTGATTCCAACC	206	222	238
R553X nAPG06646 Target 5	TCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGATT</u> CGATTCCAAGAA	207	223	239
R553X nAPG06646 Target 6	CACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGATT</u> CGATTCCAAGAA	208	224	240
R553X nAPG06646 Target 7	CCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	209	225	241
R553X nAPG07433.1 Target 1	CCAATAATTAGTTATT <u>TCACCTTGCTAAAGAAATTCTTGCTCATT</u> GACCTCCACTCAGTGTGATT	210	226	242
R553X nAPG07433.1 Target 2	TCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGATT</u> CGATTCCAAGAA	211	227	243
R553X nAPG07433.1 Target 3	CCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	212	228	244
R553X nAPG09882 Target 1	AATAATTAGTTATT <u>TCACCTTGCTAAAGAAATTCTTGCTCATTGAC</u> CTCCACTCAGTGTGA	213	229	245
R553X nAPG09882 Target 2	ATTAGTTATT <u>TCACCTTGCTAAAGAAATTCTTGCTCATTGAC</u> CTCCACTCAGTGTGA	214	230	246
R553X nAPG09882 Target 3	TATTCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	215	231	247
R553X nAPG03850 Target 1	TATTCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	216	232	248
R553X nAPG03850 Target 2	TTCACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	217	233	249
R553X nAPG03850 Target 3	CACCTTGCTAAAGAAATTCTTGCTCATTGAC <u>CTCCACTCAGTGTGA</u> TTCCACCTTCAGAAACTA	218	234	250

TABLE 31-continued

guide RNAs for nonsense mutations in CFTR					
Guide ID	Genetic locus	Genetic locus (SEQ ID NO.)	Target (SEQ ID NO.)	gRNA (SEQ ID NO.)	
R1162X nAPG09882 Target 1	GGTTTACCTCTGGCATGTCAATGAACCTAAAGACTCAGCT <u>CACAGATCGCATCTGAAAT</u> AA	251	269	287	
R1162X nAPG09882 Target 2	ACCTTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAG <u>ATCGCATCTGAAATAAAAA</u> TA	252	270	288	
R1162X nAPG09882 Target 3	CTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGC <u>ATCTGAAATAAAAATAACA</u> AC	253	271	289	
R1162X nAPG09882 Target 4	TGTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCA <u>TCTGAAATAAAAATAACAA</u> CA	254	272	290	
R1162X nAPG09882 Target 5	GTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCAT <u>CTGAAATAAAAATAACAA</u> AT	255	273	291	
R1162X nAPG06646 Target 1	TTTACCTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCA <u>CAGATCGCATCTGAAATA</u> AA	256	274	292	
R1162X nAPG06646 Target 2	TACCTTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCAC <u>GATCGCATCTGAAATAAAA</u> AT	257	275	293	
R1162X nAPG06646 Target 3	TGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCATCT <u>GAATAAAAATAACAA</u>	258	276	294	
R1162X nAPG03850 Target 1	TACCTTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCAC <u>GATCGCATCTGAAATA</u>	259	277	295	
R1162X nAPG03850 Target 2	TTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATC <u>GCATCTGAAATAAAA</u>	260	278	296	
R1162X nAPG03850 Target 3	TGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCATCT <u>GAATAAAAATAACAA</u>	261	279	297	
R1162X nAPG05586 Target 1	TTACCTTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCAC <u>AGATCGCATCTGAAAT</u>	262	280	298	
R1162X nAPG05586 Target 2	CTGTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGC <u>ATCTGAAATAAAAATA</u>	263	281	299	
R1162X nAPG05586 Target 3	TGTCAATGAACCTAAAGACTCAGCTCACAGATCGCATCTGAAAT <u>AAAAATAACAAACATT</u>	264	282	300	
R1162X nAPG00969 Target 1	GGTTTACCTCTGTTGGCATGTCAATGAACCTAAAGACTCAGCT <u>CACAGATCGCATCTGAAAT</u> AA	265	283	301	
R1162X nAPG00969 Target 2	GTTGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCAT <u>CTGAAATAAAAATAACAA</u> AT	266	284	302	
R1162X nAPG07553 Target 1	TGGCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCATCT <u>GAATAAAAATAACAA</u>	267	285	303	
R1162X nAPG01604 Target 1	GCATGTCAATGAACCTAAAGACTCAGCTCACAGATCGCATCTGA <u>AATAAAAATAACAA</u>	268	286	304	
W1282X nAPG09882 Target 1	GTGTGTTGGATTCAAATAACTTGCAACAGTGAAGGAAAGCC <u>TTTGGAGTACCCACAGG</u> TG	305	325	345	

TABLE 31-continued

## guide RNAs for nonsense mutations in CFTR

Guide ID	Genetic locus	Genetic locus (SEQ ID NO.)	Target (SEQ ID NO.)	gRNA (SEQ ID NO.)
W1282X nAPG09882 Target 2	GTCTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>GAGTGATACCACAGGTGAG</u> CA	306	326	346
W1282X nAPG09882 Target 3	CTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>GAGTGATACCACAGGTGAGCA</u> AA	307	327	347
W1282X nAPG09882 Target 4	GGGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>ATACACAGGTGAGCAA</u> AGG	308	328	348
W1282X nAPG09882 Target 5	GATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>ACCCACAGGTGAGCAAAG</u> GAC	309	329	349
W1282X nAPG06646 Target 1	TCGATGGTGTCTTGGGATTCAATAACTTGCAACAGTGAAGG <u>AAAGCCTTGAGTGATAC</u> CA	310	330	350
W1282X nAPG06646 Target 2	TTGGGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>TGAGTGATACCACAGGTGAGCA</u> AAA	311	331	351
W1282X nAPG06646 Target 3	TGGGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>GATACCACAGGTGAGCAA</u> AAG	312	332	352
W1282X nAPG06646 Target 4	GGATTCAATAACTTGCAACAGTGAAGGAAAGCCTTG <u>TACCCACAGGTGAGCAAAG</u> GGA	313	333	353
W1282X nAPG03850 Target 1	TGTCTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GGAGTGATACCACAGG</u>	314	334	354
W1282X nAPG03850 Target 2	GTCTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GAGTGATACCACAGGT</u>	315	335	355
W1282X nAPG03850 Target 3	CTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GAGTGATACCACAGGTGA</u>	316	336	356
W1282X nAPG03850 Target 4	TGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GATACCACAGGTGAGC</u>	317	337	357
W1282X nAPG07553 Target 1	CTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GAGTGATACCACAGGTGA</u>	318	338	358
W1282X nAPG07553 Target 2	TGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GATACCACAGGTGAGC</u>	319	339	359
W1282X nAPG01604 Target 1	TCTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GAGTGATACCACAGGTG</u>	320	340	360
W1282X nAPG01604 Target 2	CTTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GAGTGATACCACAGGTGA</u>	321	341	361
W1282X nAPG07433.1 Target 1	TTGGGATTCAATAACTTGCAACAGTGAAGGAAAGC <u>TGAGTGATACCACAGGTGAGCA</u> AAA	322	342	362
W1282X nAPG09748 Target 1	GTATCACTCCAAGGCTTCCTCACTGTTGCAAAGTTATTGAA <u>TCCCCAACACACCAT</u>	323	343	363
W1282X nAPG05586 Target 1	GATTCAATAACTTGCAACAGTGAAGGAAAGC <u>GGAGTGAGCAAACACAGGTGAGCAA</u>	324	344	364
F508de nAPG07433.1 SGN001101 Target1	ACCAAAAGATGATTTCTTAATGGTGCCAGGCATAATCCAGG <u>AAAACTGAGAACAGAACATGAAA</u>	562	563	564

[0785] Table 28 in Example 3 provides editing data for the SGN001101 sgRNA targeting CFTR.

[0786] To assay for activity of the other guide RNAs, a guide RNA of Table 31 is provided with the corresponding nickase variant of each RGN described in Table 30, which is operably linked to a deaminase of the invention to produce a fusion protein. It is recognized that nuclease inactive variants of each RGN may be tested similarly as well. Each guide and fusion protein combination is assayed for the ability to edit at the target location in 16HBE140-immortalized bronchial epithelial cells. Currently, three HBE cell lines containing the CFTR nonsense mutations are available (Cystic Fibrosis Foundation, Lexington, MA). These cell lines are used to assay the G542X, W1282X, and R1162x nonsense mutation targets and compared to the 16HBE140-line. The fusion protein and guide RNA is delivered to the cells as ribonucleoproteins (RNPs), which are nucleofected into the 16HBE140-cell line following culturing and transformation methods provided in Valley et al (Valley et al, 2019. J Cyst Fibros 18, 476-483, incorporated by reference herein). The guide RNA is provided as a single guide RNA or as a 1:1 or 1:1.2 molar ratio of traerRNA:crRNA duplex with RGN proteins. Nucleofection of RNPs into cells is performed on a Lonza 4D-Nucleofector. Cells are then incubated at 37° C. for 72 h. In some embodiments, the fusion protein and gRNA are delivered to the cells as RNA molecules, with the fusion protein encoded in an mRNA.

[0787] Because there are no cell lines available for the E60X, R553X, and Q493X, these mutations are assayed in HEK293 cells using a modification of the GFP restoration assay described in Example 2, where the mutant locus containing the nonsense mutation is cloned into the GFP reading frame 2.

[0788] Following incubation, genomic DNA is then extracted using NucleoSpin 96 Tissue (Macherey-Nagel) following the manufacturer's protocol. The genomic region flanking the targeted genomic site is PCR amplified and products are purified using ZR-96 DNA Clean and Concentrator (Zymo Research) following the manufacturer's protocol. The purified PCR products are then sent for Next Generation Sequencing on Illumina MiSeq. Typically, 100,000 of 250 bp paired-end reads (2×100,000 reads) are generated per amplicon. The reads are analyzed using CRISPResso (Pinello, et al. 2016) to calculate the rates of editing. Output alignments are hand-curated to confirm introduction of the base-edited mutations of interest and also to screen for undesirable INDEL formation.

[0789] In addition to efficiency of base editing, the protein product of the base-edited CFTR gene is evaluated for function. For two of the nonsense mutations, Glu60x and Gly542X, the base edited change of adenine to guanine does not restore the wildtype sequence, as these mutations are caused by guanine to thymine transversions. The targeted activity of the fusion protein changes the Glu60x to Glu60GIn and Gly452x to Gly542Arg. While these mutations do allow for a full-length protein to be made, the stability and functionality of the CFTR protein is also confirmed.

#### Example 4.2: Engineering RGNs for Decreased Size

[0790] Ideally, the coding sequence of an RGN-deaminase fusion protein of the invention and a corresponding guide RNA for targeting the fusion protein to the CFTR gene is all

packaged into a single AAV vector. The generally accepted size limit for AAV vectors is 4.7 kb, although larger sizes may be contemplated at the expense of reduced packing efficiency. The RGN nickases in Table 30 have a coding sequence length of about 3.15-3.45 kB. To ensure that the expression cassettes for both the fusion protein and its corresponding guide RNA could fit into an AAV vector, shortening the length of RGN amino acid and its corresponding nucleic acid coding sequence is desirable.

[0791] Through alignment with closely related homologs, a unique 8 amino acid region at positions 590-597 was identified in APG07433.1 and its close homolog APG08290.1 (described in WO 2019/236566 and set forth herein as SEQ ID NO: 60). This region, set forth as SEQ ID NO: 365 for APG07433.1 and SEQ ID NO: 367 for APG08290.1, was removed from both proteins, resulting in variant RGNs APG07433.1-del (SEQ ID NO: 366) and APG08290.1-del (SEQ ID NO: 368). These deletion variants and their corresponding wild-type RGNs were assayed for editing activity in HEK293T cells using the guide RNAs indicated in Tables 32 and 33 following methods similar to those described in Example 1. Rates of editing of the target sequences are shown in Tables 32 and 33 below.

TABLE 32

Editing Rate for APG07433.1 Protein Deletion Variants				
guide RNA	Target (SEQ ID NO.)	sgRNA (SEQ ID NO.)	APG07433.1	APG07433.1-del
SGN000139	369	383	11.09%	1.00%
SGN000143	370	384	2.68%	0.71%
SGN000169	371	385	13.37%	15.48%
SGN000173	372	386	13.65%	15.37%
SGN000186	373	387	14.72%	15.16%
SGN000194	374	388	11.91%	7.66%
SGN000927	376	390	9.53%	11.47%
SGN000929	378	392	6.14%	13.10%
SGN000930	379	393	7.52%	9.51%
SGN000935	381	395	11.08%	15.99%
SGN001101	382	396	6.16%	6.75%

[0792] For targets SGN000169, SGN000173, SGN000186, SGN000927, SGN000930, and SGN001101, the editing rate of the wild type APG07433.1 protein and the engineered variant was similar. For targets SGN000139, SGN000143, and SGN000194, the editing rate is decreased when the engineered variant was used compared to the wild type protein. With SGN000929 and SGN000935, the editing rate increased with the engineered APG07433.1 variant compared to the wild type sequence.

TABLE 33

Editing Rate for APG08290.1 Protein Deletion Variants				
sgRNA ID	Target (SEQ ID NO.)	sgRNA (SEQ ID NO.)	APG08290.1	APG08290.1-del
SGN000926	375	389	N.D.	6.47%
SGN000929	378	392	1.83%	0.61%
SGN000930	379	393	9.93%	6.47%
SGN000928	377	391	N.D.	0.13%
SGN000931	380	394	0%	0%

N.D. = Not determined

[0793] The APG08290.1 deletion variant showed editing in all samples where the wild type APG08290.1 protein also showed editing. The lowest editing rate detected was 0.13% with the engineered protein. Target SGN000926 showed the highest editing rate: 9.17%.

[0794] Fusion proteins comprising APG07433.1-del or APG08290.1-del and a deaminase of the invention are produced and assayed for base editing activity using methods similar to Example 1.

[0795] A fusion protein comprises an RGN and a deaminase linked by a flexible peptide linker, such as that set forth as SEQ ID NO: 45. The linker of SEQ ID NO: 45 is 16 amino acids in length; this size may be reduced to reduce the size of the coding sequence of the fusion protein. Peptide linkers of less than 16 amino acids can be produced and operably link RGNs APG07433.1-del or APG08290.1-del and a deaminase of the invention and tested for base editing activity using methods similar to Example 1. Because the peptide linker between the RGN and the deaminase can determine the editing window of the fusion protein, testing of alternative linkers with different lengths and rigidity may also lead to improvements in editing efficiency while reducing off-target mutations. Therefore, fusion proteins with the highest editing rate are then assayed following methods similar to Example 4.1 to determine editing efficiency for each of the CFTR target sequences. Fusion protein-gRNA combinations with the highest editing efficiency are selected as the preferred guide for editing at that location and are used for AAV vector design.

#### Example 4.3: AAV delivery

[0796] The coding sequences for validated fusion protein/gRNA combinations with the highest editing rate are packaged into AAV vectors. AAV delivery has a number of benefits including a lack of pathogenicity, low immunogenicity, high transduction rates, and a defined path to manufacturing. Also, AAV dosing of the lungs has been shown to be safe and at least to some degree, efficacious with both single and repeat dosing (Guggino et al., 2017, *Expert Opin Biol Ther* 17, 1265-1273). After a fusion protein/gRNA combination has been cloned into an AAV vector, it may be packaged into several different serotypes to optimize tissue specific infectivity. For treatment of CF, the target for base editing is progenitor apical epithelium cells of the lungs, which will allow the correction to persist throughout cell turnover. To target respiratory epithelium, the capsid for serotypes AAV1, AAV5 or AAV6 are utilized, as these serotypes have been shown to have high infectivity in respiratory epithelium cells (Zabner et al., 2000, *J Virol* 74, 3852-3858).

[0797] Once the AAV vectors are produced, they are transduced into human airway epithelial cells in culture. The three HBE cell lines containing the CFTR G542X, R1162X, and W1282X nonsense mutation targets are used to validate the constructs for correction of those mutations. The 16HBE140-line is used to test the constructs correcting the other nonsense mutations. A range of multiplicities of infection (MOIs) are tested. In either case, reversion of the nonsense mutation to the wild type CFTR sequence is assessed. After 2-3 days in culture, genomic DNA is harvested, amplicons around the targeted sites are generated by

PCR, and NGS is performed to determine editing rates at each locus similar to the methods described in Example 1. Because airway epithelial cells are used, AAV introduction and editing rates are as similar to an in vivo treatment as possible while using a cultured cell system. AAVs with different serotypes are compared to determine which serotype is optimal for delivery of the fusion protein/gRNA into airway cells. The editing rates achieved by AAV introduction of these systems are compared with the RNP editing rates observed in Example 4.2.

[0798] Because cell lines for the nonsense mutations R553X, E60X, and Q493X are not available, fusion protein/gRNA systems targeting these mutations are evaluated in wild type 16HBE140-cells to assay for AAV introduction, base editor expression, and off-target editing rates at the location of interest. To determine the rate of stop codon correction, the mutant locus is cloned into GFP for a GFP restoration assay as described in Example 4.1.

[0799] In parallel with determining editing rates by NGS, total protein lysates from cells harboring CFTR mutations edited with fusion protein/gRNA systems are collected and the levels of full-length CFTR protein assessed by western blotting. To test whether functional CFTR protein is formed, forskolin activation assays are performed using methods similar to those described by Devor et al (2000, *Am J Physiol Cell Physiol* 279, C461-479, incorporated by reference herein) and/or Dousmais et al (2002, *J Gen Physiol* 119, 545-559, incorporated by reference herein). In these experiments, edited CFTR mutant cells are treated with forskolin, an activator of adenylate cyclase, to increase intracellular levels of cAMP. Elevated cAMP levels then activate CFTR, and the influx of Cl<sup>-</sup> is measured by either a genetically-encoded yellow fluorescent protein based Cl<sup>-</sup> sensor or a small molecule fluorescent indicator of chloride such as MQAE. The G542X, R1162X, and W1282X edited cell lines are tested in this assay.

[0800] To determine the rate of off-target mutations, a bioinformatic approach which is customized with information about the seed region and flexible off-target PAM recognition space of each specific nuclease is used. These pieces of information have been determined bioinformatically for each protein and are used to rank the likelihood of off-target activity for each protein.

[0801] To complement bioinformatic prediction of off targets, biochemical detection of off-targets via a modified SITE-seq protocol (Cameron et al., 2017, *Nat Methods* 14, 600-606, herein incorporated by reference) is also performed. Briefly, genomic DNA from human airway epithelial cells is obtained. This DNA is then treated with the RGN of interest at several different concentrations. Any DNA double stranded breaks are labelled, selectively isolated, and PCR amplified with adapter sequences that allow for NGS. Sequencing reads are then mapped to the genome and “pileups” of reads are identified at sites of double stranded breaks, marking putative off target locations. In a subsequent set of experiments, cells are edited with the RGN or RGN-deaminase fusion protein of interest and these putative sites are individually sequenced to confirm if they are bona fide off-targets. Since chromatin context, DNA accessibility, and other factors can impact the efficiency of genome editors in living cells, biochemical methods typically overestimate the number of off-targets. Therefore, both bioinformatic and biochemical methods together provide complementary methods to identify putative off-target sites, but these sites

must be verified by amplicon sequencing to get an accurate assessment of off-target editing.

**[0802]** Once putative off-target sites are identified, amplicon sequencing on 16HBE airway epithelial cells edited with the same optimized fusion protein and guide(s) ensures that the off-target profile established for these systems matches the expected profile in patient lungs as closely as possible.

**[0803]** To determine if the fusion proteins described herein induce changes in cellular RNA, careful analysis of the cellular transcriptome following editing is necessary. Fortunately, RNA-seq techniques to assess adenine base-editing off-target effects have become routine (Grunewald et al, 2017, *Nature* 569, 433-437; Zhou et al, *Nature* 571, 275-278, both incorporated by reference herein). Briefly, after editing cells with the fusion protein/gRNA systems determined in Example 4.2, total cellular mRNA is collected and subjected to RNA-seq. Transcriptomes from edited cells are compared to cells transfected with the ABE alone, and significant differences in RNA sequence are identified.

#### Example 5: Targeted Base-Editing for Correction of Causal Disease Mutations

**[0804]** A database of clinical variants was obtained from NCBI Clin Var database, which is available through the world wide web at the NCBI Clin Var website. Pathogenic Single Nucleotide Polymorphisms (SNPs) were identified from this list. Using the genomic locus information, CRISPR targets in the region overlapping and surrounding each SNP were identified. A selection of SNPs that can be corrected using base editing in combination with an RGN, such as for example an RGN listed in Table 30 or a variant thereof, to target the causal mutation ("Cas1 Mut.") is listed in Table 34. In Table 34 below, only one alias of each disease is listed. The "RS #" corresponds to the RS accession number through the SNP database at the NCBI website. The "AlleleID" corresponds to a causal allele accession number. The "Name" column contains the genetic locus identifier, the gene name, the location of the mutation in the gene, and the change resulting from the mutation.

TABLE 34

Disease Targets for Base Editing			
RS#	AlleleID	Name	GeneSymbol
36053993	20333	NM_001128425.1(MUTYH): c.1187G > A (p.Gly396Asp)	MUTYH
41293455	32714	NM_007294.3(BRCA1): c.4327C > T (p.Arg1443Ter)	BRCA1
62625308	32710	NM_007294.3(BRCA1): c.3607C > T (p.Arg1203Ter)	BRCA1
41293465	70268	NM_007294.3(BRCA1): c.5503C > T (p.Arg1835Ter)	BRCA1
80357123	70147	NM_007294.3(BRCA1): c.5251C > T (p.Arg1751Ter)	BRCA1
137929307	171217	NM_000527.4(LDLR): c.1775G > A (p.Gly592Glu)	LDLR
80356898	45982	NM_007294.3(BRCA1): c.1687C > T (p.Gln563Ter)	BRCA1
28936415	22745	NM_000303.2(PMM2): c.422G > A (p.Arg41His)	PMM2
11555217	34125	NM_001360.2(DHCR7): c.452G > A (p.Trp151Ter)	DHCR7
55770810	70063	NM_007294.3(BRCA1): c.5095C > T (p.Arg1699Trp)	BRCA1
28934906	26850	NM_004992.3(MECP2): c.473C > T (p.Thr158Met)	MECP2
28929474	33006	NM_001127701.1(SERPINA1): c.1096G > A (p.Glu366Lys)	SERPINA1
371898076	52045	NM_000257.4(MYH7): c.198G > A (p.Arg663His)	MYH7
5030858	15616	NM_000277.3(PAH): c.1222C > T (p.Arg408Trp)	PAH
80356945	69207	NM_007294.3(BRCA1): c.2338C > T (p.Gln780Ter)	BRCA1
1800553	22927	NM_000350.2(ABCA4): c.5882G > A (p.Gly1961Glu)	ABCA4
80356962	70247	NM_007294.3(BRCA1): c.5444G > A (p.Trp1815Ter)	BRCA1
104894396	32041	NM_004004.6(GJB2): c.71G > A (p.Trp24Ter)	GJB2
113994095	28535	NM_002693.2(POLG): c.1399G > A (p.Ala467Thr)	POLG
61749721	26868	NM_004992.3(MECP2): c.763C > T (p.Arg255Ter)	MECP2
137852700	23943	NM_000310.3(PPT1): c.451C > T (p.Arg151Ter)	PPT1
75527207	22159	NM_000492.3(CFTR): c.1652G > A (p.Gly551Asp)	CFTR
78655421	22148	NM_000492.3(CFTR): c.350G > A (p.Arg117His)	CFTR
80356885	69888	NM_007294.3(BRCA1): c.4524G > A (p.Trp1508Ter)	BRCA1
113994098	28541	NM_002693.2(POLG): c.2542G > A (p.Gly848Ser)	POLG
61750240	26854	NM_004992.3(MECP2): c.808C > T (p.Arg270Ter)	MECP2
61751362	26858	NM_001110792.1(MECP2): c.916C > T (p.Arg306Ter)	MECP2
80357260	69792	NM_007294.3(BRCA1): c.4183C > T (p.Gln1395Ter)	BRCA1
80359071	67203	NM_000059.3(BRCA2): c.8243G > A (p.Gly274Asp)	BRCA2
62625307	69596	NM_007294.3(BRCA1): c.3598C > T (p.Gln1200Ter)	BRCA1
76992529	28465	NM_000371.3(TTR): c.424G > A (p.Val142Ile)	TTR
77010898	22168	NM_000492.3(CFTR): c.3846G > A (p.Trp1282Ter)	CFTR
80359003	67069	NM_000059.3(BRCA2): c.7757G > A (p.Trp2586Ter)	BRCA2
61750420	22555	NM_000466.2(PEX1): c.2528G > A (p.Gly843Asp)	PEX1
80357284	46214	NM_007294.3(BRCA1): c.5346G > A (p.Trp1782Ter)	BRCA1
200411226	174776	NM_000256.3(MYBPC3): c.1484G > A (p.Arg495Gln)	MYBPC3
5030857	98638	NM_000277.3(PAH): c.1208C > T (p.Ala403Val)	PAH
28935468	26863	NM_004992.3(MECP2): c.916C > T (p.Arg306Cys)	MECP2
62642937	15667	NM_000277.3(PAH): c.1139C > T (p.Thr380Met)	PAH
80356989	69812	NM_007294.3(BRCA1): c.4222C > T (p.Gln1408Ter)	BRCA1
28942080	18735	NM_000527.4(LDLR): c.1567G > A (p.Val523Met)	LDLR
121908039	18778	NM_000527.4(LDLR): c.551G > A (p.Cys184Tyr)	LDLR
267607213	18780	NM_000527.4(LDLR): c.131G > A (p.Trp44Ter)	LDLR
3218716	52071	NM_000257.3(MYH7): c.2389G > A (p.Ala797Thr)	MYH7
104895097	17588	NM_000243.2(MEFV): c.2282G > A (p.Arg761His)	MEFV
397516074	51962	NM_000256.3(MYBPC3): c.772G > A (p.Glu258Lys)	MYBPC3
119455955	17682	NM_000391.3(TPP1): c.622C > T (p.Arg208Ter)	TPP1

TABLE 34-continued

Disease Targets for Base Editing			
RS#	AlleleID Name		GeneSymbol
75184679	16301 NM_024570.3(RNASEH2B): c.529G > A (p.Ala177Thr)	RNASEH2B	
80338901	26909 NM_000137.2(FAH): c.1062 + 5G > A	FAH	
119450941	17501 NM_000026.3(ADSL): c.1277G > A (p.Arg426His)	ADSL	
121965019	26947 NM_000203.4(IDUA): c.1205G > A (p.Trp402Ter)	IDUA	
141659620	21858 NM_003119.3(SPG7): c.1045G > A (p.Gly349Ser)	SPG7	
41276738	15335 NM_000552.4(VWF): c.2561G > A (p.Arg854Gln)	VWF	
80338940	32068 NM_004004.5(GJB2): c.-23 + 1G > A	GJB2	
80357292	46268 NM_007294.3(BRCA1): c.962G > A (p.Trp321Ter)	BRCA1	
121913627	29130 NM_000257.3(MYH7): c.1816G > A (p.Val606Met)	MYH7	
137854601	24416 NM_198056.2(SCN5A): c.5350G > A (p.Glu1784Lys)	SCN5A	
80338933	17521 NM_024577.3(SH3TC2): c.2860C > T (p.Arg954Ter)	SH3TC2	
80338948	32048 NM_004004.5(GJB2): c.427C > T (p.Arg143Trp)	GJB2	
80356903	69645 NM_007294.3(BRCA1): c.3718C > T (p.Gln1240Ter)	BRCA1	
80356969	70213 NM_007294.3(BRCA1): c.5353C > T (p.Gln1785Ter)	BRCA1	
80357010	45971 NM_007294.3(BRCA1): c.1480C > T (p.Gln494Ter)	BRCA1	
116987552	17337 NM_005609.3(PYGM): c.148C > T (p.Arg50Ter)	PYGM	
121913625	29128 NM_000257.4(MYH7): c.1357C > T (p.Arg453Cys)	MYH7	
387907267	45725 NM_000256.3(MYBPC3): c.2827C > T (p.Arg943Ter)	MYBPC3	
28934897	26968 NM_000431.3(MVK): c.1129G > A (p.Val377Ile)	MVK	
76713772	22151 NM_000492.3(CFTR): c.1585 - 1G > A	CFTR	
137852959	19587 NM_153638.3(PANK2): c.1561G > A (p.Gly521Arg)	PANK2	
199682486	101428 NM_013339.4(ALG6): c.257 + 5G > A	ALG6	
397507389	46666 NM_000059.3(BRCA2): c.7618 - 1G > A	BRCA2	
769370816	228176 NM_000527.4(LDLR): c.1618G > A (p.Ala540Thr)	LDLR	
36211715	29159 NM_000257.4(MYH7): c.2609G > A (p.Arg870His)	MYH7	
76434661	53916 NM_004004.5(GJB2): c.416G > A (p.Ser139Asn)	GJB2	
104894368	29104 NM_000432.3(MYL2): c.64G > A (p.Glu22Lys)	MYL2	
104894635	20146 NM_000199.3(SGSH): c.734G > A (p.Arg245His)	SGSH	
121913628	29131 NM_000257.3(MYH7): c.2770G > A (p.Glu924Lys)	MYH7	
193922390	45304 NM_000257.4(MYH7): c.5135G > A (p.Arg1712Gln)	MYH7	
397515757	51454 NM_000138.4(FBN1): c.1468 + 5G > A	FBN1	
11549407	30441 NM_000518.5(HBB): c.118C > T (p.Gln40Ter)	HBB	
61751374	22933 NM_000350.2(ABCA4): c.3113C > T (p.Ala1038Val)	ABCA4	
121434420	21793 NM_004572.3(PKP2): c.235C > T (p.Arg79Ter)	PKP2	
137853007	20631 NM_007194.4(CHEK2): c.433C > T (p.Arg145Trp)	CHEK2	
1137887	18083 NM_000051.3(ATM): c.2250G > A (p.Lys750=)	ATM	
28934872	27436 NM_000548.3(TSC2): c.1832G > A (p.Arg611Gln)	TSC2	
80224560	47062 NM_000492.3(CFTR): c.2657 + 5G > A	CFTR	
80359004	46672 NM_000059.3(BRCA2): c.7758G > A (p.Trp2586Ter)	BRCA2	
121434274	18627 NM_000016.5(ACADM): c.799G > A (p.Gly267Arg)	ACADM	
121908529	38436 NM_000030.2(AGXT): c.508G > A (p.Gly170Arg)	AGXT	
121918007	28709 NM_000478.4(ALPL): c.571G > A (p.Glu1911Lys)	ALPL	
121918243	16464 NM_015506.2(MMACHC): c.482G > A (p.Arg161Gln)	MMACHC	
397518423	94255 NM_005026.4(PIK3CD): c.3061G > A (p.Glu1021Lys)	PIK3CD	
587781629	150997 NM_000059.3(BRCA2): c.1909 + 1G > A	BRCA2	
765696008	228162 NM_000527.4(LDLR): c.1187 - 10G > A	LDLR	
3218713	29127 NM_000257.3(MYH7): c.746G > A (p.Arg249Gln)	MYH7	
5030855	15646 NM_000277.3(PAH): c.1066 - 11G > A	PAH	
55851803	69067 NM_007294.3(BRCA1): c.191G > A (p.Cys64Tyr)	BRCA1	
62508698	15619 NM_000277.1(PAH): c.838G > A (p.Glu280Lys)	PAH	
62516152	108520 NM_000277.3(PAH): c.688G > A (p.Val230Ile)	PAH	
62644499	15656 NM_000277.3(PAH): c.1243G > A (p.Asp415Asn)	PAH	
80338815	18090 NM_000487.5(ARSA): c.465 + 1G > A	ARSA	
121908987	21885 NM_016203.3(PRKAG2): c.905G > A (p.Arg302Gln)	PRKAG2	
121964962	15156 NM_000071.2(CBS): c.919G > A (p.Gly307Ser)	CBS	
5030851	15628 NM_000277.3(PAH): c.842C > T (p.Pro281Leu)	PAH	
63750871	24273 NM_000535.6(PMS2): c.400C > T (p.Arg134Ter)	PMS2	
80338853	21822 NM_001360.2(DHCR7): c.278C > T (p.Thr93Met)	DHCR7	
80356893	68976 NM_007294.3(BRCA1): c.1612C > T (p.Gln538Ter)	BRCA1	
80357131	46031 NM_007294.3(BRCA1): c.2563C > T (p.Gln855Ter)	BRCA1	
80357223	69350 NM_007294.3(BRCA1): c.2800C > T (p.Gln934Ter)	BRCA1	
80357318	46112 NM_007294.3(BRCA1): c.3937C > T (p.Gln1313Ter)	BRCA1	
104886457	27086 NM_000136.2(FANCC): c.1642C > T (p.Arg548Ter)	FANCC	
137852944	19147 NM_138694.3(PKHD1): c.107C > T (p.Thr36Met)	PKHD1	
180177083	132139 NM_024675.3(PALB2): c.196C > T (p.Gln66Ter)	PALB2	
180177110	152117 NM_024675.3(PALB2): c.2257C > T (p.Arg753Ter)	PALB2	
199475575	108459 NM_000277.3(PAH): c.526C > T (p.Arg176Ter)	PAH	
387906843	39241 NM_002878.3(RAD51D): c.556C > T (p.Arg186Ter)	RAD51D	
529008617	152318 NM_001128425.1(MUTYH): c.1214C > T (p.Pro405Leu)	MUTYH	
587780021	133177 NM_000465.3(BARD1): c.1690C > T (p.Gln564Ter)	BARD1	
34637584	16979 NM_198578.3(LRRK2): c.6055G > A (p.Gly2019Ser)	LRRK2	
78802634	22233 NM_000492.3(CFTR): c.3266G > A (p.Trp1089Ter)	CFTR	
80358809	66611 NM_000059.3(BRCA2): c.581G > A (p.Trp194Ter)	BRCA2	

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
80359011	46678 NM_000059.3(BRCA2): c.7857G > A (p.Trp2619Ter)	BRCA2
104894503	27495 NM_001018005.1(TPM1): c.523G > A (p.Asp175Asn)	TPM1
121908641	21368 NM_000050.4(ASS1): c.1168G > A (p.Gly390Arg)	ASS1
121918593	28009 NM_000540.2(RYR1): c.7300G > A (p.Gly2434Arg)	RYR1
140108514	100191 NM_003494.3(DYSF): c.2643 + 1G > A	DYSF
145138923	98271 NM_000048.3(ASL): c.35G > A (p.Arg12Gln)	ASL
150726175	45795 NM_022787.3(NMNAT1): c.769G > A (p.Glu257Lys)	NMNAT1
267607578	45138 NM_170707.3(LMNA): c.1412G > A (p.Arg471His)	LMNA
376607329	48992 NM_002834.4(PTPN11): c.794G > A (p.Arg265Gln)	PTPN11
587776934	48407 NM_0005027.3(PIK3R2): c.1117G > A (p.Gly373Arg)	PIK3R2
625085888	15630 NM_000277.1(PAH): c.728G > A (p.Arg243Gln)	PAH
62637014	20604 NM_014336.4(AIPL1): c.834G > A (p.Trp278Ter)	AIPL1
80356860	46194 NM_007294.3(BRCA1): c.5117G > A (p.Gly1706Glu)	BRCA1
80357268	70265 NM_007294.3(BRCA1): c.5497G > A (p.Val1833Met)	BRCA1
80357418	70077 NM_007294.3(BRCA1): c.5136G > A (p.Trp1712Ter)	BRCA1
80358145	46229 NM_007294.3(BRCA1): c.5467 + 1G > A	BRCA1
121918166	15994 NM_000275.2(OCA2): c.1327G > A (p.Val443Ile)	OCA2
140342925	150591 NM_001128425.1(MUTYH): c.734G > A (p.Arg245His)	MUTYH
148660051	195093 NM_206933.2(USH2A): c.10073G > A (p.Cys3358Tyr)	USH2A
193922672	45341 NM_004572.3(PKP2): c.1613G > A (p.Trp538Ter)	PKP2
267607144	20039 NM_021625.4(TRPV4): c.806G > A (p.Arg269His)	TRPV4
397516083	51977 NM_000256.3(MYBPC3): c.927 - 9G > A	MYBPC3
397516357	52565 NM_000363.4(TNNI3): c.557G > A (p.Arg186Gln)	TNNI3
587782958	165560 NM_000256.3(MYBPC3): c.3190 + 5G > A	MYBPC3
28934907	26853 NM_004992.3(MECP2): c.316C > T (p.Arg106Trp)	MECP2
28934908	26862 NM_004992.3(MECP2): c.419C > T (p.Ala140Val)	MECP2
28940893	18091 NM_000487.5(ARSA): c.1283C > T (p.Pro428Leu)	ARSA
63751422	96795 NM_000535.5(PMS2): c.1927C > T (p.Gln643Ter)	PMS2
74315366	27817 NM_003000.2(SDHB): c.268C > T (p.Arg90Ter)	SDHB
80338856	34127 NM_001360.2(DHCR7): c.724C > T (p.Arg242Cys)	DHCR7
80357038	69707 NM_007294.3(BRCA1): c.3895C > T (p.Gln1299Ter)	BRCA1
80357136	69533 NM_007294.3(BRCA1): c.3403C > T (p.Gln1135Ter)	BRCA1
80357208	69682 NM_007294.3(BRCA1): c.3817C > T (p.Gln1273Ter)	BRCA1
80357234	69166 NM_007294.3(BRCA1): c.220C > T (p.Gln74Ter)	BRCA1
80357262	69729 NM_007294.3(BRCA1): c.3967C > T (p.Gln1323Ter)	BRCA1
80357305	69822 NM_007294.3(BRCA1): c.4258C > T (p.Gln1420Ter)	BRCA1
80357350	69232 NM_007294.3(BRCA1): c.241C > T (p.Gln81Ter)	BRCA1
104894636	20147 NM_000199.3(SGSH): c.220C > T (p.Arg74Cys)	SGSH
111401431	44742 NM_000138.4(FBN1): c.4588C > T (p.Arg1530Cys)	FBN1
121918624	27928 NM_006920.5(SCN1A): c.664C > T (p.Arg222Ter)	SCN1A
137852981	19794 NM_014795.3(ZEB2): c.2083C > T (p.Arg695Ter)	ZEB2
137854476	31491 NM_000138.4(FBN1): c.1585C > T (p.Arg529Ter)	FBN1
137854480	31500 NM_000138.4(FBN1): c.718C > T (p.Arg240Cys)	FBN1
180177100	133574 NM_024675.3(PALB2): c.1240C > T (p.Arg414Ter)	PALB2
193922109	44392 NM_000053.3(ATP7B): c.3955C > T (p.Arg1319Ter)	ATP7B
200640585	96857 NM_000535.6(PMS2): c.943C > T (p.Arg315Ter)	PMS2
201431517	48426 NM_139242.3(MTFMT): c.626C > T (p.Ser209Leu)	MTFMT
397516037	51905 NM_000256.3(MYBPC3): c.3697C > T (p.Gln1233Ter)	MYBPC3
587780104	133350 NM_002878.3(RAD51D): c.694C > T (p.Arg232Ter)	RAD51D
765123255	181726 NM_001128425.1(MUTYH): c.325C > T (p.Arg109Trp)	MUTYH
63751657	95331 NM_000249.3(MLH1): c.1731G > A (p.Ser577=)	MLH1
75549581	22162 NM_000492.3(CFTR): c.1675G > A (p.Ala559Thr)	CFTR
80338851	16303 NM_194318.3(B3GLCT): c.660 + 1G > A	B3GLCT
80358544	46368 NM_000059.3(BRCA2): c.2979G > A (p.Trp993Ter)	BRCA2
111033178	52388 NM_000260.3(MYO7A): c.3719G > A (p.Arg1240Gln)	MYO7A
121908188	19535 NM_020451.2(SELENON): c.943G > A (p.Gly315Ser)	SELENON
139770721	180483 NM_000051.3(ATM): c.6095G > A (p.Arg2032Lys)	ATM
199476315	40542 NM_001018005.1(TPM1): c.574G > A (p.Glu192Lys)	TPM1
267607004	15310 NM_001134363.2(RBM20): c.1907G > A (p.Arg636His)	RBM20
267608122	94980 NM_000179.2(MSH6): c.4001G > A (p.Arg1334Gln)	MSH6
377349459	150947 NM_000051.3(ATM): c.7913G > A (p.Trp2638Ter)	ATM
387906303	18745 NM_000527.4(LDLR): c.670G > A (p.Asp224Asn)	LDLR
587779227	94719 NM_000179.2(MSH6): c.2057G > A (p.Gly686Asp)	MSH6
587780290	134019 NM_000070.2(CAPN3): c.2243G > A (p.Arg748Gln)	CAPN3
727504317	49251 NM_002755.3(MAP2K1): c.199G > A (p.Asp67Asn)	MAP2K1
5030869	25402 NM_000402.4(G6PD): c.1093G > A (p.Ala365Thr)	G6PD
9332964	18390 NM_000348.3(SRD5A2): c.680G > A (p.Arg227Gln)	SRD5A2
36211723	45266 NM_000256.3(MYBPC3): c.2308G > A (p.Asp770Asn)	MYBPC3
72549410	78547 NM_000335.4(SCN5A): c.1231G > A (p.Val411Met)	SCN5A
80357498	45948 NM_007294.3(BRCA1): c.116G > A (p.Cys39Tyr)	BRCA1
80358079	70118 NM_007294.3(BRCA1): c.5194 - 12G > A	BRCA1
121434529	33201 NM_000262.2(NAGA): c.973G > A (p.Glu325Lys)	NAGA
121908627	21067 NM_005476.5(GNE): c.2086G > A (p.Val696Met)	GNE

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
387906592	38552 NM_001613.2(ACTA2): c.536G > A (p.Arg179His)	ACTA2
397515907	51711 NM_000256.3(MYBPC3): c.1505G > A (p.Arg502Gln)	MYBPC3
397516089	51992 NM_000257.4(MYH7): c.1106G > A (p.Arg369Gln)	MYH7
397516248	52239 NM_000257.4(MYH7): c.5401G > A (p.Glu1801Lys)	MYH7
397516349	52554 NM_000363.4(TNNI3): c.434G > A (p.Arg145Gln)	TNNI3
5030846	15627 NM_000277.3(PAH): c.727C > T (p.Arg243Ter)	PAH
28941784	18134 NM_052845.3(MMAB): c.556C > T (p.Arg186Trp)	MMAB
34126013	181693 NM_001128425.1(MUTYH): c.721C > T (p.Arg241Trp)	MUTYH
62541771	21074 NM_001128227.2(GNE): c.1985C > T (p.Ala662Val)	GNE
62625303	68931 NM_007294.3(BRCA1): c.1471C > T (p.Gln491Ter)	BRCA1
74315379	27453 NM_001001430.2(TNNT2): c.421C > T (p.Arg141Trp)	TNNT2
76687508	108539 NM_000277.3(PAH): c.721C > T (p.Arg241Cys)	PAH
80338794	20654 NM_012434.4(SLC17A5): c.115C > T (p.Arg39Cys)	SLC17A5
80356866	69689 NM_007294.3(BRCA1): c.3841C > T (p.Gln1281Ter)	BRCA1
80357134	69569 NM_007294.3(BRCA1): c.34C > T (p.Gln12Ter)	BRCA1
80357229	69904 NM_007294.3(BRCA1): c.4609C > T (p.Gln1537Ter)	BRCA1
111033260	19972 NM_033056.3(PCDH15): c.733C > T (p.Arg245Ter)	PCDH15
121909398	17403 NM_201548.4(CERKL): c.769C > T (p.Arg257Ter)	CERKL
121913637	29143 NM_000257.4(MYH7): c.2155C > T (p.Arg719Trp)	MYH7
200495564	50200 NM_001128425.1(MUTYH): c.733C > T (p.Arg245Cys)	MUTYH
267670203	20760 NM_194456.1(KRIT1): c.1363C > T (p.Gln455Ter)	KRIT1
587776527	132239 NM_024675.3(PALB2): c.3256C > T (p.Arg1086Ter)	PALB2
587777219	125784 NM_172107.3(KCNQ2): c.794C > T (p.Ala265Val)	KCNQ2
587778617	96774 NM_000535.5(PMS2): c.1261C > T (p.Arg421Ter)	PMS2
587783057	166274 NM_001128425.1(MUTYH): c.1171C > T (p.Gln391Ter)	MUTYH
730880099	178699 NM_000138.4(FBN1): c.1633C > T (p.Arg545Cys)	FBN1
2309689	33868 NM_000018.3(ACADVL): c.1322G > A (p.Gly441Asp)	ACADVL
28933093	29543 NM_170707.3(LMNA): c.481G > A (p.Glu161Lys)	LMNA
28937873	20571 NM_014249.3(NR2E3): c.932G > A (p.Arg311Gln)	NR2E3
59332535	77828 NM_170707.3(LMNA): c.746G > A (p.Arg249Gln)	LMNA
62645748	48213 NM_201253.2(CRB1): c.2843G > A (p.Cys948Tyr)	CRB1
63750828	96748 NM_000251.2(MSH2): c.998G > A (p.Cys333Tyr)	MSH2
80358456	65843 NM_000059.3(BRCA2): c.1689G > A (p.Trp563Ter)	BRCA2
80359101	67273 NM_000059.3(BRCA2): c.8489G > A (p.Trp2830Ter)	BRCA2
80359148	131733 NM_000059.3(BRCA2): c.8969G > A (p.Trp2990Ter)	BRCA2
80359149	67384 NM_000059.3(BRCA2): c.8970G > A (p.Trp2990Ter)	BRCA2
80359211	46791 NM_000059.3(BRCA2): c.9380G > A (p.Trp3127Ter)	BRCA2
111033565	26915 NM_002769.4(PRSS1): c.365G > A (p.Arg122His)	PRSS1
113994205	19482 NM_004937.2(CTNS): c.414G > A (p.Trp138Ter)	CTNS
116840778	23322 NM_033337.2(CAV3): c.80G > A (p.Arg27Gln)	CAV3; SSUH2
118192158	76835 NM_000540.2(RYR1): c.14818G > A (p.Ala4940Thr)	RYR1
121434278	18633 NM_000016.5(ACADM): c.583G > A (p.Gly195Arg)	ACADM
121434346	17058 NM_00103841.2(SLC6A19): c.517G > A (p.Asp173Asn)	SLC6A19
121908011	18814 NM_000372.4(TYR): c.1147G > A (p.Asp383Asn)	TYR
121908638	21365 NM_000050.4(ASS1): c.539G > A (p.Ser180Asn)	ASS1
121912938	32219 NM_001848.2(COL6A1): c.850G > A (p.Gly284Arg)	COL6A1
137853096	22694 NM_000414.3(HSD17B4): c.46G > A (p.Gly16Ser)	HSD17B4
151344631	45847 NM_000218.2(KCNQ1): c.613G > A (p.Val205Met)	KCNQ1
192838388	98283 NM_000050.4(ASS1): c.787G > A (p.Val263Met)	ASS1
267607768	95759 NM_000249.3(MLH1): c.588 + 5G > A	MLH1
376107921	213634 NM_000070.2(CAPN3): c.1319G > A (p.Arg440Gln)	CAPN3
397507981	67234 NM_000059.3(BRCA2): c.8364G > A (p.Trp2788Ter)	BRCA2
398124321	101692 NM_017780.3(CHD7): c.5405 - 7G > A	CHD7
730882246	181441 NM_194279.3(ISCA2): c.229G > A (p.Gly77Ser)	ISCA2
778906552	195186 NM_000016.5(ACADM): c.443G > A (p.Arg148Lys)	ACADM
139428292	39421 NM_005105.4(RBM8A): c.-21G > A	RBM8A
28934891	15165 NM_000071.2(CBS): c.1330G > A (p.Asp444Asn)	CBS
28937316	24408 NM_198056.2(SCN5A): c.4931G > A (p.Arg1644His)	SCN5A
33930165	30165 NM_000518.4(HBB): c.19G > A (p.Glu7Lys)	HBB
35004220	30493 NM_000518.5(HBB): c.93 - 21G > A	HBB
45546039	48043 NM_198056.2(SCN5A): c.665G > A (p.Arg222Gln)	SCN5A
61751402	105177 NM_000350.2(ABCA4): c.4469G > A (p.Cys1490Tyr)	ABCA4
72549387	22776 NM_000104.3(CYP1B1): c.171G > A (p.Trp57Ter)	CYP1B1
75822236	19350 NM_000157.3(GBA): c.1604G > A (p.Arg535His)	GBA
79389353	20821 NM_014270.4(SLC7A9): c.544G > A (p.Ala182Thr)	SLC7A9
80338862	34124 NM_001360.2(DHCR7): c.1228G > A (p.Gly410Ser)	DHCR7
80338892	27366 NM_199292.2(TH): c.698G > A (p.Arg233His)	TH
80356935	68777 NM_007294.3(BRCA1): c.1059G > A (p.Trp353Ter)	BRCA1
80357468	68802 NM_007294.3(BRCA1): c.1116G > A (p.Trp372Ter)	BRCA1
104894365	27628 NM_004985.4(KRAS): c.40G > A (p.Val14Ile)	KRAS
104894639	20153 NM_000199.3(SGSH): c.1339G > A (p.Glu447Lys)	SGSH
111033364	17396 NM_206933.2(USH2A): c.11864G > A (p.Trp3955Ter)	USH2A
119103251	17338 NM_005609.3(PYGM): c.613G > A (p.Gly205Ser)	PYGM

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
119455954	17681 NM_000391.3(TPP1): c.1094G > A (p.Cys365Tyr)	TPP1
121913638	29144 NM_000257.4(MYH7): c.2146G > A (p.Gly716Arg)	MYH7
137854478	31496 NM_000138.4(FBN1): c.3217G > A (p.Glu1073Lys)	FBN1
143353451	179937 NM_001128425.1(MUTYH): c.545G > A (p.Arg182His)	MUTYH
151045328	20182 NM_005709.3(USH1C): c.216G > A (p.Val72=)	USH1C
151344623	24127 NM_001287174.1(ABCC8): c.3992 - 9G > A	ABCC8
193922204	44739 NM_000138.4(FBN1): c.4460 - 8G > A	FBN1
193922219	51564 NM_000138.4(FBN1): c.5788 + 5G > A	FBN1
193922680	33370 NM_005159.4(ACTC1): c.301G > A (p.Glu101Lys)	ACTC1
267608172	96804 NM_000535.5(PMS2): c.2174 + 1G > A	PMS2
397516202	52163 NM_000257.3(MYH7): c.4135G > A (p.Ala1379Thr)	MYH7
397516209	52176 NM_000257.4(MYH7): c.428G > A (p.Arg143Gln)	MYH7
397517159	49176 NM_005633.3(SOS1): c.2536G > A (p.Glu846Lys)	SOS1
587776576	18532 NM_024426.5(WT1): c.1447 + 5G > A	WT1
727503246	175600 NM_000257.4(MYH7): c.4066G > A (p.Glu1356Lys)	MYH7
730881687	181107 NM_007194.4(CHEK2): c.793 - 1G > A	CHEK2
748170941	181727 NM_001128425.1(MUTYH): c.309G > A (p.Trp103Ter)	MUTYH
140583	260073 NM_000138.4(FBN1): c.2581C > T (p.Arg861Ter)	FBN1
2754158	175617 NM_000257.3(MYH7): c.2572C > T (p.Arg858Cys)	MYH7
28931570	33013 NM_001127701.1(SERPINA1): c.187C > T (p.Arg63Cys)	SERPINA1
34424986	22089 NM_004562.2(PRKN): c.823C > T (p.Arg275Ter)	PRKN
61750130	22943 NM_000350.2(ABCA4): c.4139C > T (p.Pro1380Leu)	ABCA4
61750200	22937 NM_000350.2(ABCA4): c.634C > T (p.Arg212Cys)	ABCA4
63750451	24281 NM_000535.5(PMS2): c.1882C > T (p.Arg628Ter)	PMS2
72653706	21598 NM_001171.5(ABCC6): c.3421C > T (p.Arg1141Ter)	ABCC6
74503222	108557 NM_000277.3(PAH): c.745C > T (p.Leu249Phe)	PAH
76296470	15620 NM_000277.3(PAH): c.331C > T (p.Arg111Ter)	PAH
80338860	21826 NM_001360.2(DHCR7): c.1054C > T (p.Arg352Trp)	DHCR7
80356682	29578 NM_000228.2(LAMB3): c.1903C > T (p.Arg635Ter)	LAMB3
80356771	19334 NM_001005741.2(GBA): c.1504C > T (p.Arg502Cys)	GBA
80356904	68978 NM_007294.3(BRCA1): c.1621C > T (p.Gln541Ter)	BRCA1
80356932	69850 NM_007294.3(BRCA1): c.4372C > T (p.Gln1458Ter)	BRCA1
80356947	70087 NM_007294.3(BRCA1): c.514C > T (p.Gln172Ter)	BRCA1
80356992	69906 NM_007294.3(BRCA1): c.4612C > T (p.Gln1538Ter)	BRCA1
80357133	70034 NM_007294.3(BRCA1): c.505C > T (p.Gln169Ter)	BRCA1
80357215	68781 NM_007294.3(BRCA1): c.1066C > T (p.Gln356Ter)	BRCA1
104894419	22712 NM_002312.3(LIG4): c.2440C > T (p.Arg814Ter)	LIG4
113871094	44746 NM_000138.4(FBN1): c.4786C > T (p.Arg1596Ter)	FBN1
118203682	58105 NM_000368.4(TSC1): c.2356C > T (p.Arg786Ter)	TSC1
121908177	19611 NM_031885.3(BBS2): c.823C > T (p.Arg275Ter)	BBS2
121908715	16998 NM_000022.2(ADA): c.986C > T (p.Ala329Val)	ADA
121909122	22411 NM_001083962.1(TCF4): c.1153C > T (p.Arg385Ter)	TCF4
121917901	16740 NM_000124.3(ERCC6): c.2203C > T (p.Arg735Ter)	ERCC6
121964964	15158 NM_000071.2(CBS): c.341C > T (p.Ala114Val)	CBS
137852924	18422 NM_147127.4(EVC2): c.1195C > T (p.Arg399Ter)	EVC2
137854466	31478 NM_000138.4(FBN1): c.8326C > T (p.Arg2776Ter)	FBN1
137854467	31479 NM_000138.4(FBN1): c.364C > T (p.Arg122Cys)	FBN1
137854604	24422 NM_000335.4(SCN5A): c.5126C > T (p.Ser1709Leu)	SCN5A
150518260	51200 NM_000232.4(SGCB): c.341C > T (p.Ser114Phe)	SGCB
200432447	133521 NM_007194.4(CHEK2): c.1555C > T (p.Arg519Ter)	CHEK2
201587138	176561 NM_144612.6(LOXHD1): c.4480C > T (p.Arg1494Ter)	LOXHD1
367543286	70502 NM_002609.3(PDGFRB): c.1681C > T (p.Arg561Cys)	PDGFRB
372827156	54183 NM_004572.3(PKP2): c.1237C > T (p.Arg413Ter)	PKP2
374950566	181683 NM_001128425.1(MUTYH): c.884C > T (p.Pro295Leu)	MUTYH
397514558	48266 NM_000138.4(FBN1): c.2920C > T (p.Arg974Cys)	FBN1
397515992	51839 NM_000256.3(MYBPC3): c.2905C > T (p.Gln969Ter)	MYBPC3
397516456	52796 NM_000364.3(TNNT2): c.304C > T (p.Arg102Trp)	TNNT2
587780082	133292 NM_001128425.1(MUTYH): c.1012C > T (p.Gln338Ter)	MUTYH
587782705	152480 NM_000546.5(TPS3): c.455C > T (p.Pro152Leu)	TPS3
727503974	177432 NM_172107.3(KCNQ2): c.821C > T (p.Thr274Met)	KCNQ2
730881864	180279 NM_002485.4(NBN): c.2140C > T (p.Arg714Ter)	NBN
767215758	188057 NM_002485.4(NBN): c.1030C > T (p.Gln344Ter)	NBN
45517259	27442 NM_000548.3(TSC2): c.2714G > A (p.Arg905Gln)	TSC2
61195471	57234 NM_170707.3(LMNA): c.607G > A (p.Glu203Lys)	LMNA
61753185	18815 NM_000372.4(TYR): c.230G > A (p.Arg77Gln)	TYR
63749869	28021 NM_000540.2(RYR1): c.14582G > A (p.Arg4861His)	RYR1
63749939	32145 NM_000249.3(MLH1): c.200G > A (p.Gly67Glu)	MLH1
63750119	150580 NM_000179.2(MSH6): c.3725G > A (p.Arg1242His)	MSH6
72554308	26053 NM_000531.5(OTC): c.119G > A (p.Arg40His)	OTC
79891110	32671 NM_000719.6(CACNA1C): c.1216G > A (p.Gly406Arg)	CACNA1C
80338707	22758 NM_000303.2(PMM2): c.691G > A (p.Val231Met)	PMM2
80338802	32652 NM_000070.2(CAPN3): c.2306G > A (p.Arg769Gln)	CAPN3
80356700	32571 NM_000083.2(CLCN1): c.689G > A (p.Gly230Glu)	CLCN1

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
80359803	67339 NM_000059.3(BRCA2): c.8754G > A (p.Glu2918=)	BRCA2
81002809	67078 NM_000059.3(BRCA2): c.7805 + 1G > A	BRCA2
104886142	35796 NM_000495.4(COL4A5): c.1871G > A (p.Gly624Asp)	COL4A5
104894423	17048 NM_000231.2(SGCG): c.787G > A (p.Glu263Lys)	SGCG
104894525	22747 NM_000303.2(PMM2): c.385G > A (p.Val129Met)	PMM2
113994049	20984 NM_003907.3(EIF2B5): c.338G > A (p.Arg113His)	EIF2B5
121434372	17127 NM_000159.3(GCDH): c.1198G > A (p.Val400Met)	GCDH
121908099	19299 NM_000784.3(CYP27A1): c.1214G > A (p.Arg405Gln)	CYP27A1
121908192	23730 NM_005262.2(GFER): c.581G > A (p.Arg194His)	GFER
121908753	22237 NM_000492.3(CFTR): c.1055G > A (p.Arg352Gln)	CFTR
121918013	28716 NM_000478.4(ALPL): c.346G > A (p.Ala116Thr)	ALPL
139729994	68418 NM_000492.3(CFTR): c.346G > A (p.Leu115=)	CFTR
142637046	98272 NM_000048.3(ASL): c.446 + 1G > A	ASL
142761835	177782 NM_002225.3(IVD): c.367G > A (p.Gly123Arg)	IVD
146015592	46845 NM_000060.4(BTD): c.470G > A (p.Arg157His)	BTD
150877497	226470 NM_003494.3(DYSF): c.3113G > A (p.Arg1038Gln)	DYSF
199472815	67686 NM_000218.2(KCNQ1): c.1781G > A (p.Arg594Gln)	KCNQ1
199474738	79199 NM_001042492.2(NFL): c.1885G > A (p.Gly629Arg)	NFL
199476112	24747 NC_012920.1: m.11778G > A	MT-ND4
199476317	40544 NM_001018005.1(TPM1): c.688G > A (p.Asp230Asn)	TPM1
201540674	51186 RTEL1: c.2402G > A (p.Arg801His)	RTETL1
267606640	16147 NM_000642.2(AGL): c.3980G > A (p.Trp1327Ter)	AGL
386834233	76679 NM_183050.3(BCKDHB): c.832G > A (p.Gly278Ser)	BCKDHB
397515355	19301 NM_000784.3(CYP27A1): c.1263 + 1G > A	CYP27A1
397515404	48194 NM_020822.2(KCNT1): c.1421G > A (p.Arg474His)	KCNT1
398123787	100221 NM_003494.3(DYSF): c.4253G > A (p.Gly1418Asp)	DYSF
398124641	44139 NM_024531.4(SLC52A2): c.916G > A (p.Gly306Arg)	SLC52A2
587776783	132342 NM_000321.2(RB1): c.1215 + 1G > A	RB1
587776889	39757 NM_015506.2(MMACHC): c.609G > A (p.Trp203Ter)	MMACHC
587777221	165903 NM_014191.3(SCN8A): c.4850G > A (p.Arg1617Gln)	SCN8A
587779818	132798 NM_000051.3(ATM): c.170G > A (p.Trp57Ter)	ATM
587780537	136457 NM_004360.4(CDH1): c.715G > A (p.Gly239Arg)	CDH1
587783050	166264 NM_004360.5(CDH1): c.1137G > A (p.Thr379=)	CDH1
751995154	200340 NM_000018.4(ACADVL): c.1376G > A (p.Arg459Gln)	ACADVL
781404312	186796 NM_000051.3(ATM): c.3G > A (p.Met1Ile)	ATM
786202112	184694 NM_001042492.2(NFL): c.5609G > A (p.Arg1870Gln)	NFL
794727152	191718 NM_021007.2(SCN2A): c.2558G > A (p.Arg853Gln)	SCN2A
796051858	18080 NM_000051.3(ATM): c.496 + 5G > A	ATM
796052505	201880 NM_000816.3(GABRG2): c.316G > A (p.Ala106Thr)	GABRG2
863223408	210238 NM_000020.2(ACVR1L): c.1451G > A (p.Arg484Gln)	ACVR1L
863225082	188114 NM_006245.3(PPP2R5D): c.592G > A (p.Glu198Lys)	PPP2R5D
875989911	228151 NM_000527.4(LDLR): c.938G > A (p.Cys313Tyr)	LDLR
5030852	15638 NM_000277.3(PAH): c.842 + 1G > A	PAH
5030859	15651 NM_000277.3(PAH): c.1223G > A (p.Arg408Gln)	PAH
28930068	32662 NM_000069.2(CACNA1S): c.3716G > A (p.Arg1239His)	CACNA1S
56264519	55267 NM_024022.2(TMPRSS3): c.1276G > A (p.Ala426Thr)	TMPRSS3
61750641	105317 NM_000350.2(ABCA4): c.6089G > A (p.Arg2030Gln)	ABCA4
61751276	104715 NM_000329.2(RPE65): c.11 + 5G > A	RPE65
62507336	108472 NM_000277.3(PAH): c.561G > A (p.Trp187Ter)	PAH
62508613	108291 NM_000277.2(PAH): c.1199 + 17G > A	PAH
72645357	32351 NM_000088.3(COL1A1): c.994G > A (p.Gly332Arg)	COL1A1
80338777	32664 NM_000069.2(CACNA1S): c.1583G > A (p.Arg528His)	CACNA1S
80356908	68776 NM_007294.3(BRCA1): c.1058G > A (p.Trp353Ter)	BRCA1
80357093	69031 NM_007294.3(BRCA1): c.182G > A (p.Cys61Tyr)	BRCA1
80357219	70211 NM_007294.3(BRCA1): c.5345G > A (p.Trp1782Ter)	BRCA1
104886460	99352 NM_001005741.2(GBA): c.115 + 1G > A	GBA
104894129	27501 NM_003289.3(TPM2): c.349G > A (p.Glu117Lys)	TPM2
104894401	32056 NM_004004.5(GJB2): c.428G > A (p.Arg143Gln)	GJB2
104895085	17592 NM_000243.2(MEFV): c.1958G > A (p.Arg653His)	MEFV
111033299	53902 NM_004004.5(GJB2): c.283G > A (p.Val95Met)	GJB2
113994139	33347 NM_139276.2(STAT3): c.1909G > A (p.Val637Met)	STAT3
120074135	18010 NM_000271.4(NPC1): c.2848G > A (p.Val950Met)	NPC1
121909334	23512 NM_007126.4(VCP): c.572G > A (p.Arg191Gln)	VCP
121918491	28307 NM_000141.4(FGFR2): c.1032G > A (p.Ala344=)	FGFR2
137852314	25406 NM_000402.4(G6PD): c.577G > A (p.Gly193Ser)	G6PD
137852327	25425 NM_000402.4(G6PD): c.961G > A (p.Val321Met)	G6PD
137853285	166061 NM_000053.3(ATP7B): c.2128G > A (p.Gly710Ser)	ATP7B
138213197	133488 NM_006361.5(HOXB13): c.251G > A (p.Gly84Glu)	HOXB13
148311934	44907 NM_000162.5(GCK): c.676G > A (p.Val226Met)	GCK
199473684	25807 NM_000169.2(GLA): c.639 + 919G > A	GLA
200482683	131950 NM_014625.3(NPHS2): c.868G > A (p.Val290Met)	NPHS2
371418985	232124 NM_007194.4(CHEK2): c.1232G > A (p.Trp411Ter)	CHEK2
387907281	45778 NM_152296.4(ATP1A3): c.2443G > A (p.Glu815Lys)	ATP1A3

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
397509284	70248 NM_007294.3(BRCA1): c.5445G > A (p.Trp1815Ter)	BRCA1
397514495	152034 NM_000546.5(TP53): c.542G > A (p.Arg181His)	TP53
397514581	48359 NM_172107.3(KCNQ2): c.638G > A (p.Arg213Gln)	KCNQ2
397516101	52008 NM_000257.4(MYH7): c.1358G > A (p.Arg453His)	MYH7
397516264	52270 NM_000257.3(MYH7): c.715G > A (p.Asp239Asn)	MYH7
398122822	48057 NM_001111.5(ADAR): c.3019G > A (p.Gly1007Arg)	ADAR
587777446	141325 NM_022168.4(IFIH1): c.2336G > A (p.Arg779His)	IFIH1
587782962	165566 NM_000257.4(MYH7): c.3158G > A (p.Arg1053Gln)	MYH7
606231435	170985 NM_152296.4(ATP1A3): c.2267G > A (p.Arg756His)	ATP1A3
727504247	172354 NM_001001430.2(TNNT2): c.860G > A (p.Trp287Ter)	TNNT2
730881833	179933 NM_001128425.1(MUTYH): c.857G > A (p.Gly286Glu)	MUTYH
762307622	232266 NM_001128425.1(MUTYH): c.467G > A (p.Trp156Ter)	MUTYH
777759523	17038 NM_199242.2(UNC13D): c.1389 + 1G > A	UNC13D
794728625	197538 NM_130799.2(MEN1): c.784 - 9G > A	MEN1
1060499814	389282 NM_024675.3(PALB2): c.108 + 1G > A	PALB2
25403	51465 NM_000138.4(FBN1): c.184C > T (p.Arg62Cys)	FBN1
28931591	32539 NM_000744.6(CHRNA4): c.851C > T (p.Ser284Leu)	CHRNA4
28942108	18015 NM_000271.4(NPC1): c.2932C > T (p.Arg978Cys)	NPC1
61750152	105192 NM_000350.2(ABCA4): c.4577C > T (p.Thr1526Met)	ABCA4
61750654	105349 NM_000350.2(ABCA4): c.6445C > T (p.Arg2149Ter)	ABCA4
61751404	105219 NM_000350.2(ABCA4): c.4918C > T (p.Arg1640Trp)	ABCA4
61751408	22921 NM_000350.2(ABCA4): c.6079C > T (p.Leu2027Phe)	ABCA4
63751466	24276 NM_000535.5(PMS2): c.2404C > T (p.Arg802Ter)	PMS2
72552255	44374 NM_000053.3(ATP7B): c.2930C > T (p.Thr977Met)	ATP7B
74315369	27822 NM_003000.2(SDHB): c.79C > T (p.Arg27Ter)	SDHB
80338680	16726 NM_000528.3(MAN2B1): c.2248C > T (p.Arg750Trp)	MAN2B1
80356952	68980 NM_007294.3(BRCA1): c.1630C > T (p.Gln544Ter)	BRCA1
80357011	69802 NM_007294.3(BRCA1): c.4186C > T (p.Gln1396Ter)	BRCA1
80357296	69580 NM_007294.3(BRCA1): c.3544C > T (p.Gln1182Ter)	BRCA1
80357367	70140 NM_007294.3(BRCA1): c.5239C > T (p.Gln1747Ter)	BRCA1
80357377	69340 NM_007294.3(BRCA1): c.2761C > T (p.Gln921Ter)	BRCA1
80357471	69016 NM_007294.3(BRCA1): c.178C > T (p.Gln60Ter)	BRCA1
80357497	69389 NM_007294.3(BRCA1): c.2923C > T (p.Gln975Ter)	BRCA1
104893950	18137 NM_005670.3(EPM2A): c.721C > T (p.Arg241Ter)	EPM2A
104894787	26252 NM_004006.2(DMD): c.10108C > T (p.Arg3370Ter)	DMD
111231312	51536 NM_000138.4(FBN1): c.4615C > T (p.Arg1539Ter)	FBN1
112645512	178700 NM_000138.4(FBN1): c.1285C > T (p.Arg429Ter)	FBN1
113001196	51577 NM_000138.4(FBN1): c.6658C > T (p.Arg2220Ter)	FBN1
113249837	51552 NM_000138.4(FBN1): c.5368C > T (p.Arg1790Ter)	FBN1
113812345	51455 NM_000138.4(FBN1): c.1546C > T (p.Arg516Ter)	FBN1
116100695	16552 NM_000298.5(PKLR): c.1456C > T (p.Arg486Ter)	PKLR
118203631	58047 NM_000568.4(TSC1): c.2074C > T (p.Arg692Ter)	TSC1
118203963	16148 NM_025137.3(SPG11): c.6100C > T (p.Arg2034Ter)	SPG11
118204437	15739 NM_000512.4(GALNS): c.1156C > T (p.Arg386Cys)	GALNS
121434526	33315 NM_001613.3(ACTA2): c.445C > T (p.Arg149Cys)	ACTA2
121908547	20943 NM_000334.4(SCN4A): c.3938C > T (p.Thr1313Met)	SCN4A
121912504	29459 NM_000238.3(KCNH2): c.1682C > T (p.Ala561Val)	KCNH2
121913120	31271 NM_000143.3(FH): c.301C > T (p.Arg101Ter)	FH
121913122	31274 NM_000143.3(FH): c.1027C > T (p.Arg343Ter)	FH
121917783	27083 NM_000136.2(FANCC): c.553C > T (p.Arg185Ter)	FANCC
121918775	79496 NM_006920.4(SCN1A): c.2803C > T (p.Arg935Cys)	SCN1A
121964972	15170 NM_000071.2(CBS): c.1058C > T (p.Thr535Met)	CBS
128627256	26327 NM_004006.2(DMD): c.8713C > T (p.Arg2905Ter)	DMD
137854613	24413 NM_198056.2(SCNS5A): c.4867C > T (p.Arg1623Ter)	SCN5A
137886232	39244 NM_002878.3(RAD51D): c.757C > T (p.Arg253Ter)	RAD51D
138996609	181608 NM_003000.2(SDHB): c.688C > T (p.Arg230Cys)	SDHB
144500145	202960 NM_002693.2(POLG): c.2554C > T (p.Arg852Cys)	POLG
180177111	132156 NM_024675.3(PALB2): c.2323C > T (p.Gln775Ter)	PALB2
185492864	99918 NM_001918.3(DBT): c.901C > T (p.Arg301Cys)	DBT
193922185	44706 NM_000138.4(FBN1): c.1948C > T (p.Arg650Cys)	FBN1
199472944	38732 NM_000238.3(KCNH2): c.1841C > T (p.Ala614Val)	KCNH2
199472990	78275 NM_000238.3(KCNH2): c.2254C > T (p.Arg752Trp)	KCNH2
199473161	78626 NM_198056.2(SCNS5A): c.2440C > T (p.Arg814Trp)	SCN5A
199473524	78188 NM_000238.3(KCNH2): c.1838C > T (p.Thr613Met)	KCNH2
273898674	69115 NM_007294.3(BRCA1): c.2059C > T (p.Gln687Ter)	BRCA1
368796923	151096 NM_032043.2(BRIP1): c.1240C > T (p.Gln414Ter)	BRIP1
376128990	215031 NM_052845.3(MMAB): c.571C > T (p.Arg191Trp)	MMAB
397509283	70244 NM_007294.3(BRCA1): c.5431C > T (p.Gln1811Ter)	BRCA1
397515812	51535 NM_000138.4(FBN1): c.4567C > T (p.Arg1523Ter)	FBN1
397516005	51860 NM_000256.3(MYBPC3): c.3181C > T (p.Gln1061Ter)	MYBPC3
397516042	51914 NM_000256.3(MYBPC3): c.3811C > T (p.Arg1271Ter)	MYBPC3
397516127	52044 NM_000257.3(MYH7): c.1987C > T (p.Arg663Cys)	MYH7
397516201	52162 NM_000257.4(MYH7): c.4130C > T (p.Thr1377Met)	MYH7

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
397516435	52758 NM_000546.5(TP53): c.586C > T (p.Arg196Ter)	TP53
397517689	56466 NM_001267550.2(TTN): c.71602C > T (p.Arg23868Ter)	TTN
398123585	99539 NM_001165963.1(SCN1A): c.1837C > T (p.Arg613Ter)	SCN1A
549794342	360820 NM_001271208.1(NEB): c.24094C > T (p.Arg8032Ter)	NEB
574660186	178478 NM_001267550.2(TTN): c.67495C > T (p.Arg22499Ter)	TTN
575822089	227149 NM_001163435.2(TBCK): c.376C > T (p.Arg126Ter)	TBCK
587778618	138806 NM_000535.7(PMS2): c.1687C > T (p.Arg563Ter)	PMS2
587779343	96837 NM_000535.5(PMS2): c.697C > T (p.Gln233Ter)	PMS2
587780088	133302 NM_001128425.1(MUTYH): c.55C > T (p.Arg19Ter)	MUTYH
587781269	150486 NM_007194.4(CHEK2): c.283C > T (p.Arg95Ter)	CHEK2
587781756	151166 NM_002878.3(RAD51D): c.451C > T (p.Gln151Ter)	RAD51D
672601370	171771 NM_001244008.1(KIF1A): c.946C > T (p.Arg316Trp)	KIF1A
727505006	176130 NM_000138.4(FBN1): c.3373C > T (p.Arg1125Ter)	FBN1
794728165	197808 NM_000138.4(FBN1): c.1090C > T (p.Arg364Ter)	FBN1
794728228	197690 NM_000138.4(FBN1): c.4621C > T (p.Arg1541Ter)	FBN1
794728283	197585 NM_000138.4(FBN1): c.8038C > T (p.Arg2680Cys)	FBN1
879255678	247653 NM_144997.5(FLCN): c.1429C > T (p.Arg477Ter)	FLCN
886041116	263863 NM_015339.4(ADNP): c.2188C > T (p.Arg730Ter)	ADNP
1553547838	512805 NM_001172509.1(SATB2): c.1375C > T (p.Arg459Ter)	SATB2
45507199	59122 NM_000548.3(TSC2): c.5228G > A (p.Arg1743Gln)	TSC2
60458016	29564 NM_170707.3(LMNA): c.1072G > A (p.Glu358Lys)	LMNA
61672878	29534 NM_170707.3(LMNA): c.1130G > A (p.Arg377His)	LMNA
61750173	24396 NM_000180.3(GUCY2D): c.2513G > A (p.Arg838His)	GUCY2D
61753180	18833 NM_000372.4(TYR): c.140G > A (p.Gly47Asp)	TYR
61754375	18835 NM_000372.4(TYR): c.896G > A (p.Arg299His)	TYR
62636275	20778 NM_001253.2(CRB1): c.3307G > A (p.Gly1103Arg)	CRB1
63750453	95615 NM_000249.3(MLH1): c.304G > A (p.Glu102Lys)	MLH1
63750604	95363 NM_000249.3(MLH1): c.1790G > A (p.Trp597Ter)	MLH1
63751632	95404 NM_000249.3(MLH1): c.1896G > A (p.Glu632=)	MLH1
74315205	19563 NM_006005.3(WFS1): c.2590G > A (p.Glu864Lys)	WFS1
74503330	22256 NM_000492.3(CFTR): c.3752G > A (p.Ser1251Asn)	CFTR
80282562	57854 NM_000492.3(CFTR): c.532G > A (p.Gly178Arg)	CFTR
80356702	32581 NM_000083.2(CLCN1): c.950G > A (p.Arg317Gln)	CLCN1
80358543	131539 NM_000059.3(BRCA2): c.2978G > A (p.Trp993Ter)	BRCA2
80358810	46556 NM_000059.3(BRCA2): c.582G > A (p.Trp194Ter)	BRCA2
80358997	67062 NM_000059.3(BRCA2): c.7721G > A (p.Trp2574Ter)	BRCA2
80359205	67482 NM_000059.3(BRCA2): c.9317G > A (p.Trp3106Ter)	BRCA2
81002873	67120 NM_000059.3(BRCA2): c.7976 + 1G > A	BRCA2
104894317	18840 NM_000372.4(TYR): c.1336G > A (p.Gly446Ser)	TYR
104894590	16599 NM_000263.3(NAGLU): c.2021G > A (p.Arg674His)	NAGLU
111033270	19955 NM_022124.5(CDH23): c.5237G > A (p.Arg1746Gln)	CDH23
111436401	226974 NM_000540.2(RYR1): c.10347 + 1G > A	RYR1
112406105	200333 NM_000018.4(ACADVL): c.1097G > A (p.Arg366His)	ACADVL
113560320	15440 NM_017841.2(SDHA2F): c.232G > A (p.Gly78Arg)	SDHA2F
113690956	16661 NM_000018.2(ACADVL): c.1182 + 1G > A	ACADVL
113994171	33871 NM_000018.3(ACADVL): c.1679 - 6G > A	ACADVL
113994207	19490 NM_004937.2(CTNS): c.589G > A (p.Gly197Arg)	CTNS
114925667	260377 NM_024818.4(UBA5): c.1111G > A (p.Ala371Thr)	UBA5
118192122	76888 NM_000540.2(RYR1): c.7361G > A (p.Arg2454His)	RYR1
118192176	28015 NM_000540.2(RYR1): c.6502G > A (p.Val2168Met)	RYR1
118203982	16396 NM_001080.3(ALDH5A1): c.612G > A (p.Trp204Ter)	ALDH5A1
119462987	18289 NM_007171.3(POMT1): c.2005G > A (p.Ala697Thr)	POMT1
120074190	18179 NM_000218.2(KCNQ1): c.1766G > A (p.Gly589Asp)	KCNQ1
121434544	32653 NM_000070.2(CAPN3): c.1715G > A (p.Arg572Gln)	CAPN3
121434548	32661 NM_000070.2(CAPN3): c.1469G > A (p.Arg490Gln)	CAPN3; POMT1
121908153	19416 NM_00124313.1(NLRP3): c.907G > A (p.Asp303Asn)	NLRP3
121908185	19531 NM_020451.2(SELENON): c.1397G > A (p.Arg466Gln)	SELENON
121908419	20395 NM_014384.2(ACAD8): c.1129G > A (p.Gly377Ser)	ACAD8
121908759	44497 NM_000492.3(CFTR): c.1865G > A (p.Gly622Asp)	CFTR
121908889	21460 NM_003060.3(SLC22A5): c.506G > A (p.Arg169Gln)	SLC22A5
121909013	22181 NM_000492.3(CFTR): c.1651G > A (p.Gly551Ser)	CFTR
121909019	22197 NM_000492.3(CFTR): c.3197G > A (p.Arg1066His)	CFTR
121909092	22321 NM_001005360.2(DNM2): c.1102G > A (p.Glu368Lys)	DNM2
121918009	28711 NM_000478.5(ALPL): c.1001G > A (p.Gly334Asp)	ALPL
121918592	28008 NM_000540.2(RYR1): c.1021G > A (p.Gly341Arg)	RYR1
137852871	17416 NM_000709.3(BCKDHA): c.868G > A (p.Gly290Arg)	BCKDHA
141158996	22214 NM_000492.3(CFTR): c.2490 - 1G > A	CFTR
141554661	208401 NM_004287.4(GOSR2): c.336 + 1G > A	GOSR2
148032587	194820 NM_000303.2(PMM2): c.442G > A (p.Asp148Asn)	PMM2
193922503	44492 NM_000492.3(CFTR): c.1585 - 8G > A	CFTR
199472687	77968 NM_000218.2(KCNQ1): c.421G > A (p.Val141Met)	KCNQ1
201016593	245339 NM_000527.4(LDLR): c.11G > A (p.Trp4Ter)	LDLR
267606997	21861 NM_058216.2(RAD51C): c.773G > A (p.Arg258His)	RAD51C

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
267607914	96367 NM_000251.2(MSH2): c.212 - 1G > A	MSH2
369560930	98197 NM_000018.4(ACADVL): c.520G > A (p.Val174Met)	ACADVL
370523609	227889 NM_000016.5(ACADM): c.600 - 18G > A	ACADM
370950728	186993 NM_000152.3(GAA): c.655G > A (p.Gly219Arg)	GAA
374143224	187013 NM_000152.3(GAA): c.1979G > A (p.Arg660His)	GAA
397508045	67476 NM_000059.3(BRCA2): c.92G > A (p.Trp31Ter)	BRCA2
397508200	67910 NM_000492.3(CFTR): c.1393 - 1G > A	CFTR
397509418	75098 NM_021942.5(TRAPP11): c.1287 + 5G > A	TRAPP11
397515330	76388 NM_00109851.2(PRKG1): c.530G > A (p.Arg177Gln)	PRKG1
398122711	97208 NM_000059.3(BRCA2): c.8633 - 1G > A	BRCA2
398123139	98311 NM_000060.4(BTD): c.626G > A (p.Arg209His)	BTD
398123763	100162 NM_003494.3(DYSF): c.1053 + 1G > A	DYSF
587777057	77012 NM_020988.2(GNAO1): c.607G > A (p.Gly203Arg)	GNAO1
587777570	150453 NM_004522.2(KIF5C): c.709G > A (p.Glu237Lys)	KIF5C
587778777	76741 NM_000784.3(CYP27A1): c.1184 + 1G > A	CYP27A1
587779110	96248 NM_000251.2(MSH2): c.1760 - 1G > A	MSH2
587780639	139490 NM_000051.3(ATM): c.7788G > A (p.Glu2596=)	ATM
587781894	151348 NM_000051.3(ATM): c.9023G > A (p.Arg3008His)	ATM
587782719	152505 NM_000051.3(ATM): c.8122G > A (p.Asp2708Asn)	ATM
727503030	176785 NM_00127893.1(ELN): c.1150 + 1G > A	ELN
730881581	180665 NM_000059.3(BRCA2): c.8174G > A (p.Trp2725Ter)	BRCA2
730882035	180121 NM_000051.3(VHL): c.482G > A (p.Arg161Gln)	VHL
750663117	234071 NM_000051.3(ATM): c.3078 - 1G > A	ATM
756039188	243266 NM_000527.4(LDLR): c.12G > A (p.Trp4Ter)	LDLR
796053216	202741 NM_014191.3(SCN8A): c.4423G > A (p.Gly1475Arg)	SCN8A
876661242	231905 NM_000059.3(BRCA2): c.9381G > A (p.Trp3127Ter)	BRCA2
879254600	245669 NM_000527.4(LDLR): c.626G > A (p.Cys209Tyr)	LDLR
1057519632	362622 NM_003718.4(CKD13): c.2149G > A (p.Gly717Arg)	CDK13
10250779	15457 NM_000290.3(PGAM2): c.233G > A (p.Trp78Ter)	PGAM2
28928905	29469 NM_000238.3(KCNH2): c.1468G > A (p.Ala490Thr)	KCNH2
28931593	32066 NM_004004.5(GJB2): c.224G > A (p.Arg75Gln)	GJB2
28937318	24429 NM_198056.2(SCNSA): c.1100G > A (p.Arg367His)	SCNSA
61749397	15329 NM_000552.4(VWF): c.3946G > A (p.Val1316Met)	VWF
61751403	105220 NM_000350.2(ABCA4): c.4919G > A (p.Arg1640Gln)	ABCA4
62514907	15633 NM_000277.3(PAH): c.442 - 1G > A	PAH
62514956	98659 NM_000277.3(PAH): c.912 + 1G > A	PAH
62516146	108608 NM_000277.1(PAH): c.842 + 5G > A	PAH
62642939	98657 NM_000277.2(PAH): c.890G > A (p.Arg297His)	PAH
62644503	108560 NM_000277.3(PAH): c.755G > A (p.Arg252Gln)	PAH
63749856	21618 NM_001171.5(ABCC6): c.3904G > A (p.Gly1302Arg)	ABCC6
63750783	30442 NM_000518.5(HBB): c.47G > A (p.Trp16Ter)	HBB
66555264	414003 NM_000088.3(COL1A1): c.1821 + 1G > A	COL1A1
72645321	414022 NM_000088.3(COL1A1): c.769G > A (p.Gly257Arg)	COL1A1
74315368	27820 NM_003000.2(SDHB): c.725G > A (p.Arg242His)	SDHB
74315471	18113 NM_000487.5(ARSA): c.739G > A (p.Gly247Arg)	ARSA
78973108	19367 NM_001005741.2(GBA): c.887G > A (p.Arg296Gln)	GBA
80338735	33917 NM_000156.5(GAMT): c.327G > A (p.Lys109=)	GAMT
80338857	34128 NM_001360.2(DHCR7): c.725G > A (p.Arg242His)	DHCR7
80338864	21831 NM_001360.2(DHCR7): c.1342G > A (p.Glu448Lys)	DHCR7
80338944	32040 NM_004004.5(GJB2): c.231G > A (p.Trp77Ter)	GJB2
80356914	70276 NM_007294.3(BRCA1): c.551G > A (p.Trp183Ter)	BRCA1
80357212	70255 NM_007294.3(BRCA1): c.5467G > A (p.Ala1823Thr)	BRCA1
80357307	70275 NM_007294.3(BRCA1): c.5510G > A (p.Trp183Ter)	BRCA1
80358252	18013 NM_000271.4(NPC1): c.530G > A (p.Cys177Tyr)	NPC1
104894103	19470 NM_175073.2(APTX): c.837G > A (p.Trp279Ter)	APTX
104894415	20583 NM_006783.4(GJB6): c.31G > A (p.Gly11Arg)	GJB6
104894519	21096 NM_004862.3(LITAF): c.334G > A (p.Gly112Ser)	LITAF
104894727	27461 NM_000363.4(TNNI3): c.586G > A (p.Asp196Asn)	TNNI3
104894828	25754 NM_000169.2(GLA): c.902G > A (p.Arg301Gln)	GLA
111683277	175150 NM_000256.3(MYBPC3): c.3190 + 1G > A	MYBPC3
111984349	258823 NM_000138.4(FBN1): c.782G > A (p.Glu2610Lys)	FBN1
113403872	16550 NM_000298.5(PKLR): c.1529G > A (p.Arg510Gln)	PKLR
121434249	18383 NM_000348.3(SRD5A2): c.682G > A (p.Ala228Thr)	SRD5A2
121908216	23534 NM_001127221.1(CACNA1A): c.4982G > A (p.Arg1661His)	CACNA1A
121908551	20948 NM_000334.4(SCN4A): c.3877G > A (p.Val1293Ile)	SCN4A
121908552	20949 NM_000334.4(SCN4A): c.1333G > A (p.Val445Met)	SCN4A
121908557	20958 NM_000334.4(SCN4A): c.2024G > A (p.Arg675Gln)	SCN4A
121908716	16996 NM_000022.2(ADA): c.632G > A (p.Arg211His)	ADA
121908723	17007 NM_000022.3(ADA): c.646G > A (p.Gly216Arg)	ADA
121909768	21834 NM_001360.2(DHCR7): c.1055G > A (p.Arg352Gln)	DHCR7
121913039	31702 NM_001953.4(TYMP): c.622G > A (p.Val208Met)	TYMP
137853050	22116 NM_006009.3(TUBA1A): c.1265G > A (p.Arg422His)	TUBA1A

TABLE 34-continued

Disease Targets for Base Editing			
RS#	AlleleID Name		GeneSymbol
137853283	166064 NM_000053.3(ATP7B): c.2336G > A (p.Trp779Ter)		ATP7B
137854612	24434 NM_198056.2(SCN5A): c.4222G > A (p.Gly1408Arg)		SCN5A
139751448	187031 NM_000271.4(NPC1): c.1211G > A (p.Arg404Gln)		NPC1
150038620	187049 NM_004646.3(NPHS1): c.2335 – 1G > A		NPHS1
180177122	132185 NM_024675.3(PALB2): c.2718G > A (p.Trp906Ter)		PALB2
181087667	40103 NM_007055.3(POLR3A): c.2617 – 1G > A		POLR3A
193922110	44393 NM_000053.3(ATP7B): c.4058G > A (p.Trp1353Ter)		ATP7B
199473565	78528 NM_198056.2(SCN5A): c.1066G > A (p.Asp356Asn)		SCN5A
199474703	40437 NM_000258.2(MYL3): c.281G > A (p.Arg94His)		MYL3
199971687	216058 NM_052845.3(MMAB): c.291 – 1G > A		MMAB
201188361	40345 NM_014714.3(IFT140): c.634G > A (p.Gly212Arg)		IFT140
202160208	75126 NM_013334.3(GMPPB): c.860G > A (p.Arg287Gln)		GMPPB
281875334	38553 NM_001101.3(ACTB): c.587G > A (p.Arg196His)		ACTB
386134249	45185 NM_000244.3(MEN1): c.1277G > A (p.Cys426Tyr)		MEN1
387906623	38652 NM_000138.4(FBN1): c.5284G > A (p.Gly1762Ser)		FBN1
387906905	39430 NM_021625.4(TRPV4): c.947G > A (p.Arg316His)		TRPV4
397507479	48850 NM_004333.5(BRAF): c.1595G > A (p.Cys532Tyr)		BRAF
397514494	48018 NM_021625.4(TRPV4): c.557G > A (p.Arg186Gln)		TRPV4
397515854	51599 NM_000138.4(FBN1): c.760G > A (p.Gly2536Arg)		FBN1
397515982	51820 NM_000256.3(MYBPC3): c.2670G > A (p.Trp890Ter)		MYBPC3
397516031	51898 NM_000256.3(MYBPC3): c.3627 + 1G > A		MYBPC3
397516471	52818 NM_001001430.2(TNNT2): c.518G > A (p.Arg173Gln)		TNNT2
398122853	38917 NM_004006.2(DMD): c.9G > A (p.Trp3Ter)		DMD
483352809	65656 NM_006087.3(TUBB4A): c.745G > A (p.Asp249Asn)		TUBB4A
515726205	40114 NM_001031726.3(C19orf12): c.205G > A (p.Gly69Arg)		C19orf12
564069299	200114 NM_000255.3(MMUT): c.1106G > A (p.Arg369His)		MMUT
574673404	182906 NM_002485.4(NBN): c.37 + 1G > A		NBN
587780345	134590 NM_000162.3(GCK): c.544G > A (p.Val182Met)		GCK
606231324	136674 NM_000257.3(MYH7): c.1573G > A (p.Glu525Lys)		MYH7
727504382	49283 NM_030662.3(MAP2K2): c.619G > A (p.Glu207Lys)		MAP2K2
730880850	29166 NM_000257.3(MYH7): c.732 + 1G > A		MYH7
730882175	181517 NM_002238.3(KCNH1): c.1405G > A (p.Gly469Arg)		KCNH1
751604696	425943 NM_001360.2(DHCR7): c.1337G > A (p.Arg446Gln)		DHCR7
753288303	216044 NM_000255.3(MMUT): c.1280G > A (p.Gly427Asp)		MMUT
767399782	213656 NM_006087.3(TUBB4A): c.763G > A (p.Val255Ile)		TUBB4A
794728208	197723 NM_000138.4(FBN1): c.371G > A (p.Asp1238Asn)		FBN1
796756333	410338 NM_024422.4(DSC2): c.943 – 1G > A		DSC2
797044872	205316 NM_004977.2(KCNC3): c.1268G > A (p.Arg423His)		KCNC3
797045586	207083 NM_032682.5(FOXP1): c.1541G > A (p.Arg514His)		FOXP1
863223403	209408 NM_002140.4(HNRNPK): c.257G > A (p.Arg86His)		HNRNPK
876658367	232176 NM_003000.2(SDHB): c.587G > A (p.Cys196Tyr)		SDHB
105751785	358911 NM_024675.3(PALB2): c.3G > A (p.Met1Ile)		PALB2
155558205	431537 NM_014233.3(UBTF): c.628G > A (p.Glu210Lys)		UBTF
140630	197685 NM_000138.4(FBN1): c.4930C > T (p.Arg1644Ter)		FBN1
28940869	19031 NM_017739.3(POMGNT1): c.1324C > T (p.Arg442Cys)		POMGNT1
34451549	30497 NM_000518.5(HBB): c.316 – 197C > T		HBB
41556519	31832 NM_000400.3(ERCC2): c.2047C > T (p.Arg683Trp)		ERCC2
45611033	175462 NM_000257.4(MYH7): c.3133C > T (p.Arg1045Cys)		MYH7
55832599	151478 NM_000546.5(TP53): c.799C > T (p.Arg267Trp)		TP53
59616921	18036 NM_000226.3(KRT9): c.487C > T (p.Arg163Trp)		KRT9
60399023	29651 NM_000526.4(KRT14): c.373C > T (p.Arg125Cys)		KRT14
61749409	104973 NM_000350.2(ABCA4): c.1804C > T (p.Arg602Trp)		ABCA4
61749423	105003 NM_000350.2(ABCA4): c.2041C > T (p.Arg681Ter)		ABCA4
61750645	105327 NM_000350.2(ABCA4): c.6229C > T (p.Arg2077Trp)		ABCA4
61751383	22946 NM_000350.2(ABCA4): c.6088C > T (p.Arg2030Ter)		ABCA4
61752871	28154 NM_000329.2(RPE65): c.271C > T (p.Arg91Trp)		RPE65
61757582	21827 NM_001360.2(DHCR7): c.1210C > T (p.Arg404Cys)		DHCR7
61816761	31358 NM_002016.1(FLG): c.1501C > T (p.Arg501Ter)		FLG
62507344	15662 NM_000277.2(PAH): c.1066 – 3C > T		PAH
72559722	186816 NM_001287174.1(ABCC8): c.2509C > T (p.Arg837Ter)		ABCC8
72646846	56340 NM_001256850.1(TTN): c.56953C > T (p.Arg1898Ter)		TTN
72648250	225057 NM_001256850.1(TTN): c.88243C > T (p.Arg2941Ter)		TTN
72650700	39295 NM_001171.5(ABCC6): c.1552C > T (p.Arg518Ter)		ABCC6
72651642	271557 NM_000088.3(COL1A1): c.2089C > T (p.Arg697Ter)		COL1A1
72653170	32386 NM_000088.3(COL1A1): c.3040C > T (p.Arg1014Cys)		COL1A1
74315348	20408 NM_014625.3(NPHS2): c.871C > T (p.Arg291Trp)		NPHS2
74315391	22425 NM_172107.3(KCNQ2): c.619C > T (p.Arg207Trp)		KCNQ2
74315442	23435 NM_000100.3(CSTB): c.202C > T (p.Arg68Ter)		CSTB
74315472	18114 NM_000487.5(ARSA): c.827C > T (p.Thr276Met)		ARSA
75166491	108429 NM_000277.3(PAH): c.472C > T (p.Arg158Trp)		PAH
75949023	39947 NM_144612.6(LOXHD1): c.4714C > T (p.Arg1572Ter)		LOXHD1
78635798	16299 NM_032193.3(RNASEH2C): c.205C > T (p.Arg69Trp)		RNASEH2C
80338652	18848 NM_000081.3(LYST): c.3310C > T (p.Arg1104Ter)		LYST

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
80338826	29117 NM_002473.5(MYH9): c.2104C > T (p.Arg702Cys)	MYH9
80338934	17522 NM_024577.3(SH3TC2): c.3325C > T (p.Arg1109Ter)	SH3TC2
80338957	20935 NM_000334.4(SCN4A): c.2111C > T (p.Thr704Met)	SCN4A
80356680	29580 NM_000228.2(LAMB3): c.124C > T (p.Arg42Ter)	LAMB3
80356779	76552 NM_001876.3(CPT1A): c.1436C > T (p.Pro479Leu)	CPT1A
80356973	69370 NM_007294.3(BRCA1): c.2869C > T (p.Gln957Ter)	BRCA1
80356982	69227 NM_007294.3(BRCA1): c.2410C > T (p.Gln804Ter)	BRCA1
80357067	69840 NM_007294.3(BRCA1): c.4339C > T (p.Gln1447Ter)	BRCA1
80357089	69512 NM_007294.3(BRCA1): c.3331C > T (p.Gln1111Ter)	BRCA1
80357352	69958 NM_007294.3(BRCA1): c.4810C > T (p.Gln1604Ter)	BRCA1
80357485	69485 NM_007294.3(BRCA1): c.3286C > T (p.Gln1096Ter)	BRCA1
80359818	31157 NM_006516.3(SLC2A1): c.376C > T (p.Arg126Cys)	SLC2A1
80359826	201142 NM_006516.3(SLC2A1): c.988C > T (p.Arg330Ter)	SLC2A1
104894003	33314 NM_001101.4(ACTB): c.547C > T (p.Arg183Trp)	ACTB
104894261	31727 NM_130799.2(MEN1): c.1579C > T (p.Arg527Ter)	MEN1
104894267	31731 NM_130799.2(MEN1): c.1378C > T (p.Arg460Ter)	MEN1
104894364	27627 NM_004985.4(KRAS): c.173C > T (p.Thr58Ile)	KRAS
104894621	23472 NM_000304.3(PMP22): c.215C > T (p.Ser72Leu)	PMP22
104894714	19826 NM_181882.2(PRX): c.2857C > T (p.Arg953Ter)	PRX
104894797	26321 NM_004006.2(DMD): c.9568C > T (p.Arg3190Ter)	DMD
11103297	53892 NM_004004.5(GJB2): c.169C > T (p.Gln57Ter)	GJB2
11103538	17382 NM_032601.3(MCEE): c.139C > T (p.Arg47Ter)	MCEE
111687884	51571 NM_000138.4(FBN1): c.643C > T (p.Arg215Ter)	FBN1
112901682	76366 NM_001141945.2(ACTA2): c.115C > T (p.Arg39Cys)	ACTA2
114368325	38634 NM_000782.4(CYP24A1): c.1186C > T (p.Arg396Trp)	CYP24A1
118192226	34614 NM_172107.3(KCNQ2): c.1342C > T (p.Arg448Ter)	KCNQ2
118192251	34269 NM_004519.3(KCNQ3): c.988C > T (p.Arg330Cys)	KCNQ3
118203427	58245 NM_000368.4(TSC1): c.682C > T (p.Arg228Ter)	TSC1
118203434	58253 NM_000368.4(TSC1): c.733C > T (p.Arg245Ter)	TSC1
118203542	57958 NM_000368.4(TSC1): c.1525C > T (p.Arg509Ter)	TSC1
118203999	16285 NM_024675.3(PALB2): c.2962C > T (p.Gln988Ter)	PALB2
118204429	15511 NM_000035.4(ALDOB): c.178C > T (p.Arg60Ter)	ALDOB
121907916	18505 NM_000280.4(PAX6): c.607C > T (p.Arg203Ter)	PAX6
121908212	23527 NM_001127221.1(CACNA1A): c.1997C > T (p.Thr666Met)	CACNA1A
121908427	20365 NM_133647.1(SLC12A6): c.3031C > T (p.Arg1011Ter)	SLC12A6
121908489	20807 NM_003919.2(SGCE): c.289C > T (p.Arg97Ter)	SGCE
121912708	33034 NM_001182.4(ALDH7A1): c.328C > T (p.Arg110Ter)	ALDH7A1
121913344	151858 NM_000546.5(TP53): c.916C > T (p.Arg306Ter)	TP53
121917784	27085 NM_000136.2(FANCC): c.37C > T (p.Gln13Ter)	FANCC
121918167	15995 NM_000275.2(OCA2): c.2228C > T (p.Pro743Leu)	OCA2
121918244	16869 NM_001023570.3(IQCB1): c.1381C > T (p.Arg461Ter)	IQCB1
121918257	16926 NM_000255.3(MMUT): c.322C > T (p.Arg108Cys)	MMUT
122445105	26774 NM_000489.4(ATRX): c.736C > T (p.Arg246Cys)	ATRX
122445108	26781 NM_000489.4(ATRX): c.109C > T (p.Arg37Ter)	ATRX
122453121	26733 NM_004484.3(GPC3): c.1159C > T (p.Arg387Ter)	GPC3
128626235	26264 NM_004006.2(DMD): c.433C > T (p.Arg145Ter)	DMD
137852897	17803 NM_024312.4(GNPTAB): c.3565C > T (p.Arg1189Ter)	GNPTAB
137852994	19999 NM_018136.4(ASPM): c.9178C > T (p.Gln3060Ter)	ASPM
137853229	21102 NM_004260.3(RECQL4): c.2269C > T (p.Gln757Ter)	RECQL4
138049878	171163 NM_000257.4(MYH7): c.2608C > T (p.Arg870Cys)	MYH7
138119149	39897 NM_0020745.3(AARS2): c.1774C > T (p.Arg592Trp)	AARS2
139675596	40180 NM_023073.3(CPLANE1): c.7477C > T (p.Arg2493Ter)	CPLANE1
140511594	39892 NM_024753.4(TTC21B): c.626C > T (p.Pro209Leu)	TTC21B
143343083	169011 NM_004004.5(GJB2): c.298C > T (p.His100Tyr)	GJB2
148865119	210450 NM_000071.2(CBS): c.146C > T (p.Pro49Leu)	CBS
180177091	132277 NM_024675.3(PALB2): c.751C > T (p.Gln251Ter)	PALB2
199422209	33004 NM_001127701.1(SERPINA1): c.1178C > T (p.Pro393Leu)	SERPINA1
199473556	78702 NM_198056.2(SCNS5A): c.361C > T (p.Arg121Trp)	SCNS5A
200075782	39327 NM_003560.3(PLA2G6): c.109C > T (p.Arg37Ter)	PLA2G6
200287925	151917 NM_002485.4(NBN): c.127C > T (p.Arg43Ter)	NBN
200309328	176122 NM_000138.4(FBN1): c.8080C > T (p.Arg2694Ter)	FBN1
200440128	205749 NM_012160.4(FBXL4): c.64C > T (p.Arg22Ter)	FBXL4
201632198	55279 NM_024022.2(TMPRSS3): c.325C > T (p.Arg109Trp)	TMPRSS3
267606919	21912 NM_004646.3(NPHS1): c.3478C > T (p.Arg1160Ter)	NPHS1
267607143	20038 NM_021625.4(TRPV4): c.943C > T (p.Arg155Ter)	TRPV4
267607258	46918 NM_002437.5(MPV17): c.293C > T (p.Pro98Leu)	MPV17
375699023	223602 NM_024675.3(PALB2): c.1042C > T (p.Gln348Ter)	PALB2
387906799	39125 NM_001244008.1(KIF1A): c.296C > T (p.Thr99Met)	KIF1A
387906904	39429 NM_021625.4(TRPV4): c.694C > T (p.Arg232Cys)	TRPV4
387907329	51081 NM_007075.3(WDR45): c.700C > T (p.Arg234Ter)	WDR45
397507215	46080 NM_007294.3(BRCA1): c.3352C > T (p.Gln118Ter)	BRCA1
397507447	47625 NM_024312.4(GNPTAB): c.1123C > T (p.Arg375Ter)	GNPTAB
397509002	69322 NM_007294.3(BRCA1): c.2713C > T (p.Gln905Ter)	BRCA1

TABLE 34-continued

Disease Targets for Base Editing		
RS#	AlleleID Name	GeneSymbol
397509151	69806 NM_007294.3(BRCA1): c.4201C > T (p.Gln1401Ter)	BRCA1
397509330	70405 NM_007294.3(BRCA1): c.850C > T (p.Gln284Ter)	BRCA1
397514477	40113 NM_001031726.3(C19orf12): c.32C > T (p.Thr11Met)	C19orf12
397515848	51592 NM_000138.4(FBN1): c.7180C > T (p.Arg2394Ter)	FBN1
397516463	52805 NM_001001430.2(TNNT2): c.388C > T (p.Arg130Cys)	TNNT2
398123061	76995 NM_012160.4(FBXL4): c.1444C > T (p.Arg482Trp)	FBXL4
398123168	98367 NM_000143.3(FH): c.952C > T (p.His318Tyr)	FH
398123832	100328 NM_004006.2(DMD): c.10171C > T (p.Arg3391Ter)	DMD
398123929	100476 NM_004006.2(DMD): c.3151C > T (p.Arg1051Ter)	DMD
398124478	102281 NM_138694.3(PKHD1): c.2341C > T (p.Arg781Ter)	PKHD1
536907995	137626 NM_007194.4(CHEK2): c.58C > T (p.Gln20Ter)	CHEK2
587776407	153707 NM_024675.3(PALB2): c.451C > T (p.Gln151Ter)	PALB2
587776935	48413 NM_005465.4(AKT3): c.1393C > T (p.Arg465Trp)	AKT3
587780062	133253 NM_000535.5(PMS2): c.823C > T (p.Gln275Ter)	PMS2
587780226	133611 NM_032043.2(BRIP1): c.1315C > T (p.Arg439Ter)	BRIP1
587781948	151416 NM_000465.3(BARD1): c.1921C > T (p.Arg641Ter)	BARD1
587783685	168920 NM_003482.3(KMT2D): c.12592C > T (p.Arg4198Ter)	KMT2D
587784339	169779 NM_003560.3(PLA2G6): c.1903C > T (p.Arg635Ter)	PLA2G6
724159971	172085 NM_152778.2(MFSD8): c.1444C > T (p.Arg482Ter)	MFSD8
727503504	176073 NM_000363.4(TNNI3): c.508C > T (p.Arg170Trp)	TNNI3
727503513	172503 NM_001001430.2(TNNT2): c.280C > T (p.Arg94Cys)	TNNT2
727504136	177069 NM_001165963.1(SCN1A): c.3733C > T (p.Arg1245Ter)	SCN1A
730881422	179951 NM_000465.3(BARD1): c.1996C > T (p.Gln666Ter)	BARD1
730882029	180988 NM_000546.5(TP53): c.1024C > T (p.Arg342Ter)	TP53
747604569	185305 NM_032043.2(BRIP1): c.484C > T (p.Arg162Ter)	BRIP1
750621215	184806 NM_002878.3(RAD51D): c.898C > T (p.Arg300Ter)	RAD51D
753330544	195505 NM_206933.2(USH2A): c.13316C > T (p.Thr4439Ile)	USH2A
761494650	185659 NM_007194.4(CHEK2): c.85C > T (p.Gln29Ter)	CHEK2
763091520	197655 NM_000138.4(FBN1): c.6169C > T (p.Arg2057Ter)	FBN1
768933093	226933 NM_024685.4(BBS10): c.145C > T (p.Arg49Trp)	BBS10
773770609	264863 NM_17750.4(SLC13A5): c.997C > T (p.Arg333Ter)	SLC13A5
778989252	236615 NM_007194.4(CHEK2): c.1315C > T (p.Gln439Ter)	CHEK2
786202064	184902 NM_007294.3(BRCA1): c.4834C > T (p.Gln1612Ter)	BRCA1
786203821	184272 NM_024675.3(PALB2): c.940C > T (p.Gln314Ter)	PALB2
794726710	187772 NM_001165963.1(SCN1A): c.3637C > T (p.Arg1213Ter)	SCN1A
794726730	187817 NM_001165963.1(SCN1A): c.2134C > T (p.Arg712Ter)	SCN1A
794728195	197755 NM_000138.4(FBN1): c.2645C > T (p.Ala882Val)	FBN1
796051885	199890 NM_003239.4(TGFB3): c.898C > T (p.Arg300Trp)	TGFB3
797044883	205286 NM_019066.4(MAGEL2): c.1912C > T (p.Gln638Ter)	MAGEL2
869312892	226683 NM_139276.2(STAT3): c.2147C > T (p.Thr716Met)	STAT3
876658461	232175 NM_003000.2(SDHB): c.640C > T (p.Gln214Ter)	SDHB
886037684	248861 NM_177438.2(DICER1): c.2062C > T (p.R688*)	DICER1
886038001	249129 NM_007294.3(BRCA1): c.2599C > T (p.Gln867Ter)	BRCA1
886039480	260102 NM_024675.3(PALB2): c.2368C > T (p.Gln790Ter)	PALB2
886040218	261660 NM_007294.3(BRCA1): c.4225C > T (p.Gln1409Ter)	BRCA1
886041222	264422 NM_000280.4(PAX6): c.781C > T (p.Arg261Ter)	PAX6
1057521083	366251 NM_015265.3(SATB2): c.1165C > T (p.Arg389Cys)	SATB2

#### Example 6: Demonstration of Gene Editing Activity in Plant Cells

**[0805]** Base-editing activity of an RGN-deaminase fusion protein of the invention is demonstrated in plant cells using protocols adapted from Li, et al., 2013 (*Nat. Biotech.* 31:688-691). Briefly, an expression vector comprising an expression cassette capable of expressing in plant cells an RGN-deaminase fusion protein operably linked to a SV40 nuclear localization signal (SEQ ID NO: 43) and a second expression cassette encoding a guide RNA targeting one or more sites in the plant PDS gene that flank an appropriate PAM sequence are introduced into *Nicotiana benthamiana* mesophyll protoplasts using PEG-mediated transformation. The transformed protoplasts are incubated in the dark for up to 36 hr. Genomic DNA is isolated from the protoplasts using a DNeasy Plant Mini Kit (Qiagen). The genomic region flanking the RGN target site is PCR amplified, products are purified, and the purified PCR products are analyzed using Next Generation Sequencing on Illumina

MiSeq. Typically, 100,000 of 250 bp paired-end reads (2×100,000 reads) are generated per amplicon. The reads are analyzed using CRISPResso (Pinello, et al. 2016 *Nature Biotech.* 34:695-697) to calculate the rates of editing. Output alignments are analyzed for INDEL formation or introduction of specific adenine mutations.

#### Example 7: Testing mRNA Delivery

**[0806]** To determine if the base editors are capable of delivery in different formats, mRNA delivery was tested with primary T-cells. Purified CD3+ T-cells or PBMCs were thawed, activated using CD3/CD28 beads (ThermoFisher) for 3 days, then nucleofected using the Lonza 4D-Nucleofector X unit and Nucleocuvette strips. The P3 Primary Cell kit was used for both mRNA and RNP delivery. Cells were transfected using the EO-115 and EH-115 programs for mRNA and RNP delivery respectively. Cells were cultured in CTS OpTimizer T cell expansion medium (ThermoFisher) containing IL-2, IL-7, and IL-15 (Miltenyi Bio-

tec) for 4 days post nucleofection before being harvested using a Nucleospin Tissue genomic DNA isolation kit (Macherey Nagel).

[0807] Amplicons surrounding the editing sites were generated by PCR using primers identified in Table 35 and subjected to NGS sequencing using the Illumina Nextera platform using 2×250 bp paired end sequencing. The estimated base editing rate was determined by calculating the overall substitution rate for each sample. The average and number of samples for each guide tested are shown below.

TABLE 35

Average Editing rate for LPG50148-nAPG07433.1 via mRNA delivery		
SGN	Average % Edit	N
SGN002352	7.84	2
SGN002364	29.79	2
SGN002367	0.1	2
SGN001061	0.37	1
SGN001062	71.81	1
SGN001064	3.99	1
SGN002254	8.92	2
SGN002255	5.26	2

TABLE 35-continued

Average Editing rate for LPG50148-nAPG07433.1 via mRNA delivery

SGN	Average % Edit	N
SGN002256	8.32	2
SGN002290	2.88	2
SGN002293	9.68	2
SGN002299	27.05	2
SGN002132	29.11	2
SGN002137	7.77	2
SGN002139	6.00	2
SGN001770	1.22	2
SGN001773	0.49	2
SGN002212	29.63	2
SGN002216	2.58	2
SGN002218	36.13	2
SGN002230	14.32	2
SGN002231	33.18	2
SGN000753	6.84	2
SGN000754	26.41	1
SGN001856	0.5	2
SGN002248	9.91	2
SGN002249	40.19	2

## SEQUENCE LISTING

```

Sequence total quantity: 564
SEQ ID NO: 1      moltype = AA  length = 169
FEATURE          Location/Qualifiers
REGION           1..169
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..169
note = source = /note="APG09982 protein sequence"
source            1..169
mol_type = protein
organism = synthetic construct
SEQUENCE: 1
MSDLELNHEY WMRHALQLAK RARDEGEVPV GAVLVVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC SGAMVHSRIG TLVFGVRNEK TGAAGSLMDV 120
LRHPGMNHQV QIIDGVLAPE CSGLLCRFFR MPRRVFNQQK AESTSSPGD 169

SEQ ID NO: 2      moltype = AA  length = 164
FEATURE          Location/Qualifiers
REGION           1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..164
note = source = /note="APG03724 protein sequence"
source            1..164
mol_type = protein
organism = synthetic construct
SEQUENCE: 2
MSNPELTHEH WMRYALTTLAK RAREEGEVPV GAVLVVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG QLVFGVRNAK TGAAGSLMDV 120
LHHPGMNHRI EFTEGVLRDE CAAMLCRFFR QPRRVFNALK TGNA 164

SEQ ID NO: 3      moltype = AA  length = 169
FEATURE          Location/Qualifiers
REGION           1..169
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..169
note = source = /note="APG09949 protein sequence"
source            1..169
mol_type = protein
organism = synthetic construct
SEQUENCE: 3
MSIPELNHDV WMRHALQLAK RAREEGEVPV GAVLVLNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG QLVFGVRNAK TGAAGSLIDV 120

```

---

-continued

---

LHHPGMNHGV AITEGVLRREE CAAMLCRFFR QPRRVFNALK KPAGDPTAF	169
 SEQ ID NO: 4 moltype = AA length = 172	
FEATURE Location/Qualifiers	
REGION 1..172	
note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION 1..172	
note = source = /note="APG08196 protein sequence"	
source 1..172	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 4	
MSNPELNHEY WMRYALTAK RARDEGEVPV GAVLVLNDQV IGEGNRAIG LHDPTAHEI 60	
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG RLVFGVRNAK TGAAGSLIDV 120	
LHHPGMNHHI EMEEGVLRDE CAAMLCRFFR QPRRVFNALK KSPPDSPNLQ AR 172	
 SEQ ID NO: 5 moltype = AA length = 169	
FEATURE Location/Qualifiers	
REGION 1..169	
note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION 1..169	
note = source = /note="APG06333 protein sequence"	
source 1..169	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 5	
MSNPELTHDH WMRHALTLAQ RARNEGEVPV GAVLVLNGQV IGEGNRAIG LHDPTAHEI 60	
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG QLVFGVRNAK TGAAGSLIDV 120	
LHHPGMNHGV EIIEGVLRDE CAAMLCRFFR HPRRVFNALK KNAGTSPTQ 169	
 SEQ ID NO: 6 moltype = AA length = 166	
FEATURE Location/Qualifiers	
REGION 1..166	
note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION 1..166	
note = source = /note="APG06489 protein sequence"	
source 1..166	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 6	
MSDTELNHEY WMRHALMLAK RARDEGEVPV GAVLVLKNQV IGEGNRAIG LHDPTAHEI 60	
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG NLVFGVRNAK TGAAGSLIDV 120	
LHHPGMNHGV EIAEGVLADE CSAMLCRFFR HPRRVFNALK QAAKHD 166	
 SEQ ID NO: 7 moltype = AA length = 171	
FEATURE Location/Qualifiers	
REGION 1..171	
note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION 1..171	
note = source = /note="APG08449 protein sequence"	
source 1..171	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 7	
MSDIELNHEY WMRHALMLAK RAREEGEVPV GAVLVLNNQV IGEGNRAIG LHDPTAHEI 60	
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG HLVFGVRNAK TGAAGSLIDV 120	
LHHPGMNHRI EFTEGVLADE CSGMLCRFFR YPRRVFNTLK QAAKANPPAA Q 171	
 SEQ ID NO: 8 moltype = AA length = 173	
FEATURE Location/Qualifiers	
REGION 1..173	
note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION 1..173	
note = source = /note="APG05174 protein sequence"	
source 1..173	
mol_type = protein	
organism = synthetic construct	
 SEQUENCE: 8	
MSIPELNHDV WMRHALTLAK RAREEGEVPV GAVLVLNGQV IGEGNRAIG LHDPTAHEI 60	
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG QLVFGVRNAK TGAAGSLMDV 120	
LHHPGMNHGV EITEGVLRDE CAAMLCRFFR QPRRVFNALK KPAGDPSALQ NNR 173	

---

---

-continued

---

```

SEQ ID NO: 9          moltype = AA  length = 168
FEATURE
REGION
1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..168
note = source = /note="APG09102 protein sequence"
1..168
mol_type = protein
organism = synthetic construct
SEQUENCE: 9
MSNPFPFTHEY WMRHALTLAR RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLYDTTL YSTFEPVCVMC SGAMVHSRIG TLVFGVRNEK TGAAGSLMDV 120
LGHPGMNHQV KTIGGVLAPE CSGLLCRFFR MPRRVFNQQK AELKSSGD 168

SEQ ID NO: 10         moltype = AA  length = 167
FEATURE
REGION
1..167
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..167
note = source = /note="APG05723 protein sequence"
1..167
mol_type = protein
organism = synthetic construct
SEQUENCE: 10
MSDAELTHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLYDTTL YSTFEPVCVMC AGAMVHSRIG RLIFGVRNAK TGAAGSLIDV 120
LHHPGMNHRV EVVEGILRDE CAGMLCRFFR QPRRVFNAALK KGATDVL 167

SEQ ID NO: 11         moltype = DNA  length = 507
FEATURE
misc_feature
1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..507
note = source = /note="mammalian codon optimized APG09982"
1..507
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 11
atcagggcac tggaaactgaa tcacggatgtc tggatgaaatc acggccctgca gctggccaag 60
cggggccagag atgaggcgca ggtgcggctgtc ggccgtgtc tggatgtc caaccaggatg 120
atcggccaaag gttggaaacat agccattgtt ctggatgttca caacatgtt cggccaaatc 180
atggccctgc ggcaggccgg cctggatgttca caacatgtt cggccaaatc 240
tatagcacat ttcggccatc cttggatgttca ttggatgttca ttggatgttca 300
accctggatgtt cttggatgttca ttggatgttca ttggatgttca ttggatgttca 360
ctggggcatc ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 420
ttttctggcc ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 480
ggcaatcca ccaggccggcc ttggatgttca ttggatgttca ttggatgttca ttggatgttca 507

SEQ ID NO: 12         moltype = DNA  length = 492
FEATURE
misc_feature
1..492
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..492
note = source = /note="mammalian codon optimized APG03724"
1..492
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 12
atcagggcaccc ccggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 60
cggggccagag aggaaggaga agtggatgttca ttggatgttca ttggatgttca ttggatgttca 120
atcggccaaatc gttggaaatag agccattgtt ctggatgttca ttggatgttca ttggatgttca 180
atggccctgc gacaggccgg cttggatgttca ttggatgttca ttggatgttca ttggatgttca 240
tactccacct ttcggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 300
caatgttccgg ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 360
ctggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 420
ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 480
acaggcaacg cc ttggatgttca ttggatgttca ttggatgttca ttggatgttca ttggatgttca 492

SEQ ID NO: 13         moltype = DNA  length = 507
FEATURE
misc_feature
1..507

```

---

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..507
note = source = /note="mammalian codon optimized APG09949"
source        1..507
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 13
atgagcatcc ccgagctgaa tcacgatgtt tggatgcggc acgcccgtac cctggccaaa 60
agagccagag aggaaggcga ggtgcctgtg gggtgcgtgc tggtgcgtgaa cgcccgatgtg 120
atcgaggaaag gtcggaaatag agccatggta ctgcgtatgacc ctacagccca cgctgaaatc 180
atggccctga gacaggccgg cctggccctc cagaactaca gactgtacga caccacctg 240
tactctacct tcgagccttg cgtgtatgtgc gccggcgcca tggtgcactc cagaatccgc 300
caagctggtgt tcggcgatgtc gaacggccaaag acaggcgctg ctggcgactc gatcgacgtg 360
ctgcgtatcacc tcggcatgaa ccacagggtt gccaatcaccg aggaggatgtg gccggaaagag 420
tgccgcgcoca tgctgtgttag attctcaga caaccttagac gggcttcaa cgccctgaaag 480
aaggccatgtt gcgaccggcac agccttt 507

SEQ ID NO: 14      moltype = DNA length = 516
FEATURE           Location/Qualifiers
misc_feature      1..516
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..516
note = source = /note="mammalian codon optimized APG08196"
source            1..516
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 14
atgagcaacc ccgagctgaa tcacgagttt tggatgcggc acgcccgtac actggccaa 60
cgccgcctggg acgaggccggc agtgcctgtg gggtgcgtgc tggtgcgtgaa cgcccgatgtg 120
atcgaggaaag gtcggaaatag agccatggta ctgcgtatgacc ctacagccca cgcccgatgtc 180
atggccctgc gacaggccgg cctggccctc cagaactaca ggctgtacga cacaacccctg 240
tattccacct tcgagccttg cgtgtatgtgc gccggcgcca tggtgcacag cagaatccgc 300
caagctggtgt tcggcgatgtc gaacggccaaag acaggcgctg ctggcgactc gatcgacgtg 360
ctgcgtatcacc tcggcatgaa ccacacatc gatggaaag aaggcgatgtg gagatgtgag 420
tgccgcgcota tgctgtgtcc gttttcaga caaccttagaa gatgttcaa cgccctgaaag 480
aaatctccac ctgatagcccc taatctgcag gccaga 516

SEQ ID NO: 15      moltype = DNA length = 507
FEATURE           Location/Qualifiers
misc_feature      1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..507
note = source = /note="mammalian codon optimized APG06333"
source            1..507
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 15
atgagcaacc ctgagctgac acacgaccac tggatgcggc acgctctgac cctggcccg 60
cgccgcctggg acgaggccggc agtgcctgtg gggtgcgtgc tggtgcgtgaa cgcccgatgtg 120
atcgaggaaag gtcggaaatag agccatggta ctgcgtatgacc ctacagccca cgcccgatgtc 180
atggccctga gacaggccgg cctggctgtg cagaactaca gactgtacga caccgtgtc 240
tacagccacct tcggcgatgtc cgtgtatgtgc gccggcgcca tggtccactc tagaatccgc 300
caagctggtgt tcggcgatgtc gaatggccaaag acaggcgctg ccggcgactc gatcgacgtg 360
cttcatcacc ccggaaatgaa ccacagatgtt gatggcgatgtg gccggatgaa 420
tgcgcgcota tgctgtgtcc gttttcaga cacccaaagaa gggatgttcaa cgccctgaaa 480
aagaacgccc gcaccggccc cacccag 507

SEQ ID NO: 16      moltype = DNA length = 498
FEATURE           Location/Qualifiers
misc_feature      1..498
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..498
note = source = /note="mammalian codon optimized APG06489"
source            1..498
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 16
atgagcgaca ccgagctgaa ccacgagttt tggatgcggc acgcccgtat gctggctaa 60
cgccgcctggg atgaggccggc agtgcctgtg gggtgcgtgc tggtccctgaa gaaccaggat 120
atcgaggaaag gtcggaaatag agccatggta ctgcgtatgacc ctacagccca cgcccgatgtc 180
atggccctga gacaaggccgg cctggctgtg cagaactaca gactgtacga cacaacccctg 240
tattccacct tcgagccttg cgtgtatgtgc gccggcgcca tggtgcactc tagaatccgc 300

```

---

-continued

---

```

aatctggtgt tcggcgtgcg gaacgccaag accggcgctg ctggcgact gatcgacgtg 360
ctccatcacc ctggaaatgaa ccacagatgtg aaatcgccg aaggagtgtt ggccgacgaa 420
tgcagcgccca tgctgtgcgg ttttcaga caccaaggc gggtgtttaa cgcctgaaa 480
caggccgcta agcacgac 498

SEQ ID NO: 17 moltype = DNA length = 513
FEATURE Location/Qualifiers
misc_feature 1..513
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..513
note = source = /note="mammalian codon optimized APG08449"
source 1..513
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 17
atgtctgata tcgagctgaa tcacgactac tggatgcggc acggcctgat gctggccaag 60
cgggcccacag aggaaaggcgaa agtgcgttg ggcccgctgc tggtgtgaa caaccagggtg 120
atcgagaag gctgaaatag agccatcgcc ctgcgtatgc ctaccgccc cgccgagatc 180
atggccctgaa gacaggcgccg actgggtgtc cagaactacc ggctgtacga caccacctg 240
tacagcacat tcgagccttg tggatgtgc gccggagccca tggtgcacag cagaatccgc 300
caacctggttt ttggcgatgcgaa gaacgccaag accggcgctg ctggcgact gatcgacgtc 360
ctgcatcacc ctggcatgaa ccacagaatgaa attcacaacg agggcggtgtt cgccgacgag 420
tgctccggca tgctgtgcgg gttttcaga tttcttagaa gggtgttcaa caccctgaaa 480
caggccgcta aaggcaaccc ccccgccgtt caa 513

SEQ ID NO: 18 moltype = DNA length = 519
FEATURE Location/Qualifiers
misc_feature 1..519
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..519
note = source = /note="mammalian codon optimized APG05174"
source 1..519
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 18
atgtctatcc ctgagctgaa ccacgatgtg tggatgcggc acggcctgac actggccaag 60
cgggcccacag aagaggaggaa agtgcgttg ggcccgctgc tggtgtgaa tgcccgatgtg 120
atcgccgcacag gctgaaatag agccatcgcc ctgcgtatgc ctaccgccc cgccgagatt 180
atggccctgcg ggcaggcgccg actggtcgtc caaaattaca gactgtacga caccacactg 240
tacagcacatc ttggatccttg tggatgtgc gccggcgcttga tggtgcacag cagaatccgc 300
caacctggttt ttggcgatgcgaa gaacgccaac accggcgccg ctggcgccct gatggacgtg 360
ctgcatcacc ctggcatgaa ccacatcgatgaa attcacaacg agggcggtctt cagatgag 420
tgctccggca tgctgtgcgg gttttcaga tttcttagaa gggtgttcaa caccctgaaa 480
aaggccgctggcg cgcggccacaccc cggccgtt cggccgtt aacaaccgg 519

SEQ ID NO: 19 moltype = DNA length = 504
FEATURE Location/Qualifiers
misc_feature 1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..504
note = source = /note="mammalian codon optimized APG09102"
source 1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 19
atgagcaacc cccaaatccac ccacgactac tggatgcggc acggcctgac actggctaga 60
aggggccggg acggggcgca ggtgcgttg ggcccgctgc tggtgtgaa caaccagggtg 120
atcgagaag gctgaaatag agccatcgcc ctgcgtatgc ctaccgccc cgccgaaatc 180
atggccctgaa gacaggcgccg cttgtgtgtc cagaactacc ggctgtacga caccacactg 240
tatagcacatc tcgagccttg cttgtgttg agcggcgacta tggtgcacag cagaatccgc 300
accctggttt ttggcgatgcgaa gaacgagaag accggcgccg ctggcgccct gatggacgtg 360
ctcgccggccaccc ccggcatgaa ccacccggatc aacccatcgatc gggcggtgtt ggccctgaa 420
tgtagccggcc tgctgtgcgg gttttcaga atgccttagaa gagtgtttaa tcaacagaaa 480
cccgagatgtgaa agtctccgg 504

SEQ ID NO: 20 moltype = DNA length = 501
FEATURE Location/Qualifiers
misc_feature 1..501
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..501
note = source = /note="mammalian codon optimized APG05723"
source 1..501

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct

SEQUENCE: 20
atgagcgcac ccgagctgac acacgagtac tggatgcggc acgcctgac cctggccag 60
cgccgcagat atgaggaga agtgcctgt ggcgcgtgc ttgtgtgaa caaccagtg 120
atcgccgaag gtggaatag agccatcgga ctgcgttacca ccacccgccta cgctgaaatc 180
atggccctga gacaggcggg cctggccatc cagaactaca gactgtacca caccacccgt 240
tattccacct tcggatctt tggtatgtgc gcggagacta ttgtgcacag cagaatccgc 300
agactgattt tcggcgtgcg gaacgccaag acaggcgcctt ctggatctt gatcgacgtg 360
ctccatcacc cccgcattttt gacaggatctt gagggtggtgg aaggcatctt gcgggacgag 420
tcggccggca tgctgtgcg ttgttcaga caaccttaggc gggtttaa cgcctgaaag 480
aaaggcgcta cagatgtgtc g 501

SEQ ID NO: 21      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="SGN000930 target sequence"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 21
gaacaactca aatggaaatg aatat 25

SEQ ID NO: 22      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000186 target sequence"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 22
ggacagtgcg catctccctg 20

SEQ ID NO: 23      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000194 target sequence"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 23
ggcccacagc attcaggteg 20

SEQ ID NO: 24      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000143 target sequence"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 24
catggcagta cattagagca 20

SEQ ID NO: 25      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000139 target sequence"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 25

```

---

-continued

---

aggtttaat ggcccgccct	20
SEQ ID NO: 26	moltype = RNA length = 135
FEATURE	Location/Qualifiers
misc_feature	1..135
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..135
	note = source = /note="SGN000930 sgRNA sequence"
source	1..135
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 26	
gaacaactca aatggaaaatg aatatgtcat agttccatga aagccaaaag tggctttatg 60	
gtttctatga taagggtttc ggcccggtgc gtcggggatc gcctccccat tccgatggc 120	
tttcccccattt ttatt	135
SEQ ID NO: 27	moltype = RNA length = 130
FEATURE	Location/Qualifiers
misc_feature	1..130
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..130
	note = source = /note="SGN000186 sgRNA sequence"
source	1..130
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 27	
ggacagtgcg catctccctg gtcatagttc cattaaagcc aaaagtggct ttgatgttgc 60	
tatgataagg gtttcgaccc gtggcgctgg ggatcgcttg cccattgaaa tgggcttctc 120	
cccatttattt	130
SEQ ID NO: 28	moltype = RNA length = 130
FEATURE	Location/Qualifiers
misc_feature	1..130
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..130
	note = source = /note="SGN000194 sgRNA sequence"
source	1..130
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 28	
ggccgcacagc attcaggctcg gtcatagttc cattaaagcc aaaagtggct ttgatgttgc 60	
tatgataagg gtttcgaccc gtggcgctgg ggatcgcttg cccattgaaa tgggcttctc 120	
cccatttattt	130
SEQ ID NO: 29	moltype = RNA length = 130
FEATURE	Location/Qualifiers
misc_feature	1..130
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..130
	note = source = /note="SGN000143 sgRNA sequence"
source	1..130
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 29	
catggcagta cattagagca gtcatagttc cattaaagcc aaaagtggct ttgatgttgc 60	
tatgataagg gtttcgaccc gtggcgctgg ggatcgcttg cccattgaaa tgggcttctc 120	
cccatttattt	130
SEQ ID NO: 30	moltype = RNA length = 130
FEATURE	Location/Qualifiers
misc_feature	1..130
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..130
	note = source = /note="SGN000139 sgRNA sequence"
source	1..130
	mol_type = other RNA
	organism = synthetic construct
SEQUENCE: 30	
aggtttaat ggcccgccct gtcatagttc cattaaagcc aaaagtggct ttgatgttgc 60	
tatgataagg gtttcgaccc gtggcgctgg ggatcgcttg cccattgaaa tgggcttctc 120	
cccatttattt	130

---

-continued

---

```

SEQ ID NO: 31      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000930 FWD primer"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 31
gacagccaag aggtttgcc                                     20

SEQ ID NO: 32      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000930 REV primer"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 32
ctgtcccttg cagcttctgt                                     20

SEQ ID NO: 33      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000186 FWD primer"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 33
tggcccstat gtggagatca                                     20

SEQ ID NO: 34      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000186 REV primer"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 34
ggcagagtc agcctcatag                                     20

SEQ ID NO: 35      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000194 FWD primer"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 35
atgacattca ggccacagtg                                     20

SEQ ID NO: 36      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="SGN000194 REV primer"
1..20
mol_type = other DNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 36
ttccctccta ttcaggccca                                         20

SEQ ID NO: 37          moltype = DNA  length = 19
FEATURE
misc_feature
1..19
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..19
note = source = /note="SGN000143 FWD primer"
source
1..19
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 37
acatggacg agcagcgaa                                         19

SEQ ID NO: 38          moltype = DNA  length = 22
FEATURE
misc_feature
1..22
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..22
note = source = /note="SGN000143 REV primer"
source
1..22
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 38
agggccccctg gagagggttt aa                                         22

SEQ ID NO: 39          moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..20
note = source = /note="SGN000139 FWD primer"
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 39
ctttagctg gaggtccatc                                         20

SEQ ID NO: 40          moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..20
note = source = /note="SGN000139 REV primer"
source
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 40
tgttggcaa tctagtcg                                         20

SEQ ID NO: 41          moltype = AA  length = 1071
FEATURE
REGION
1..1071
note = source = /note="Bacillus sp. APG07433.1"
source
1..1071
mol_type = protein
organism = Bacillus sp.

SEQUENCE: 41
MRELDYRIGL DIGTNSIGWG VIELSWNKDR ERYEKVRIVD QGVVRMFRAE MPKTGASLAE  60
PRRIARSSRR RLNRKSQRKK NIRNLVQHG VITQEELDSL YPLSKKSMDI WGIRLDGLDR 120
LILNHFEWRL LIHLAQRRGF KSNRKSELKD TETGKVLSII QLNEKRLSLY RTVGBEMWMKD 180
PDFSKYDRKR NSPNEYVFVSV SRAELEKEIV TLFAAQRRFQ SPYASKDLQE TYLQIWTQHQL 240
PFASGNAILN KVGYCSLLKG KERRIPKATY TFQYFSALDQ VNRTTRLGPDF QPFTKEQREI 300
ILNNMFMQRTD YYKKKTIPEV TYYDIRKWLE LDETIQFKGL NYDPNEELKK IEKKPFINLK 360
AFYEINKVVA NYINSERTNETF STLDYDGIGY ALTVYKTDKD IRSYLNKSSHN LPKRCYDDQL 420
IEELLSLSYT KFGHLSLKA1 NHVLSIMQKG NTYKEAVDQL GYDTSGLKKE KRSKELPPIS 480
DEITNPPIVKR ALTQARKVNV AIIIRRHSPPH SVHIELAREL SKNHDERTKI VSAQDENYKK 540
NIKGAISLSE HGILNPTGYD IVRYKLWKEQ GERCAYSLLKE IPADTFNNEL KKERNNGAPIL 600
EVDHILPYSQ SFIDSYHNKV LVYSDENRKK GNRIPYTYFL ETNKDWEAFA RYVRSNKFSS 660
KKKREYLLKR AYLPRESELI KERHLNDTRY ASTFLKNFIE QNLQFKEAED NPRKRRVQTV 720
NGVITAHFRK RWGLEKDRQE TYLHHAMDAI IVACTDHMMV TRVTEYYQQIK ESNKSVKPY 780

```

-continued

---

FPMPWEGFRD	ELLSHLASQP	IAKKISEEELK	AGYQSLDYIF	VSMPKRSIT	GAAHKQTIMR	840
KGGIDKKGKT	IIERLHLKD	IKFDENGDFK	MVGKEQDMAT	YEAIKQRYLE	HGKNSSKAFE	900
TPLYKPSKKG	TGNLIKRVKV	EGQAQSFVRE	VNGGVAQNGD	LVRVDLFEKD	DKYYMVPYIV	960
PDTVCSELPK	KVASSKGYE	QWLTLDNSPT	FKPSLYPVDL	VRLVKGDEDR	FLYFGTLDID	1020
SDRLNFKDVN	KPSKKNEYRY	SLKTIEDLEK	YEVGVLGDLR	LVRKETRRNF	H	1071

SEQ ID NO: 42	moltype = AA length = 1071					
FEATURE	Location/Qualifiers					
REGION	1..1071					
	note = source = /note="Description of Artificial Sequence:					
	Synthetic polypeptide"					
REGION	1..1071					
	note = source = /note="nAPG07433.1"					
source	1..1071					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 42						
MRELYRIGL	AIGTN SIGWG	VIELSWNKDR	ERYEKVIRD	QGVRMF DRAE	MPKTGASLAE	60
PRRIARSSRR	RLNRKSQRKK	NIRNL LVHQH	VITQEELDSL	YPLSKKSM DI	WGIRLDGLDR	120
LINHFEWRL	LIHLAQRRGF	KSNRKSELD K	TETGKVLS SSI	QLNEKRLS LY	RTVGBMWMD	180
PDFSKYDRKR	NSPNEYVF SV	SRAELEKEIV	TLFAAQR RFQ	SPYASKDLQ E	TYLQIWT HQL	240
PFASGNAILN	KVGYCSLLKG	KERRIPKAT Y	TFQYFSALD Q	VNRTRLGP DF	QPFTKEQRE I	300
ILNNMFQRTD	YYKKKTIPEV	TYYDIRK WLE	LDETIQFK GL	NYDQNEEL KK	IEKKP FINLK	360
AFYEINKVVA	NYSER TNETF	STLDYDGIV Y	ALTVYKTD KD	IRS YLKS SHN	LPKRCYDDQ L	420
IEELLSLSYT	KFGHLSLKAI	NHVLSIMQKG	NTYKEAVD QL	GYDTSGL KKE	KRSKFL PPI S	480
DEITNP IVKR	ALTQARKVV N	AIIRRHGS PH	SVHIELARE L	SKNHDERT KI	VSAQDENYKK	540
NKGAI SILSE	HGILNPTGYD	IVRYKLWKE Q	GERCAYS LKE	IPADTFFN EL	KKERNGA PIL	600
EVDHILPSQ	SPIDS YHN KV	LVYSDENR KK	GNRIPYTF L	ETNKDWEA FE	RYVRSN KFFS	660
KKKREYLLKR	AYL PRESELI	KERHLNDTRY	ASTFLKNF IE	QNLQFKEA ED	NPRKRRV QT V	720
NGVITA HFRK	RWGLEKDRQ E	TYLHHAMDA I	IAVACTDH MM	TRVTEYYQIK	ESNKSVK PY	780
FPMPWEGFRD	ELLSHLASQP	IAKKISEEEL K	AGYQSLDY IF	VSMPKRSIT	GAAHKQTIM R	840
KGGIDKKGKT	IIERLHLKD	IKFDENGDF K	MVGKEQDM AT	YEAIKQRY LE	HGKNSSK AF E	900
TPLYKPSKKG	TGNLIKRVKV	EGQAQSFVRE	VNGGVAQNG D	LVRVDLFEKD	DKYYMVPYIV	960
PDTVCSELPK	KVASSKGYE	QWLTLDNSPT	FKPSLYPVD L	VRLVKGDED R	FLYFGTLDID	1020
SDRLNFKD VN	KPSKKNEYRY	SLKTIEDLE K	YEVGVLGDL R	LVRKETRRNF	H	1071

SEQ ID NO: 43	moltype = AA length = 7					
FEATURE	Location/Qualifiers					
REGION	1..7					
	note = source = /note="Description of Artificial Sequence:					
	Synthetic peptide"					
REGION	1..7					
	note = source = /note="SV40 NLS"					
source	1..7					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 43						
PKKKRKV						7

SEQ ID NO: 44	moltype = AA length = 22					
FEATURE	Location/Qualifiers					
REGION	1..22					
	note = source = /note="Description of Artificial Sequence:					
	Synthetic peptide"					
REGION	1..22					
	note = source = /note="3X Flag tag"					
source	1..22					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 44						
DYKDHGDYK DH DIDYK DDD DK						22

SEQ ID NO: 45	moltype = AA length = 16					
FEATURE	Location/Qualifiers					
REGION	1..16					
	note = source = /note="Description of Artificial Sequence:					
	Synthetic peptide"					
REGION	1..16					
	note = source = /note="peptide linker"					
source	1..16					
	mol_type = protein					
	organism = synthetic construct					
SEQUENCE: 45						
SGSETPGTSE SATPES						16

SEQ ID NO: 46	moltype = AA length = 16
---------------	--------------------------

-continued

---

```

FEATURE          Location/Qualifiers
REGION          1..16
                note = source = /note="Description of Artificial Sequence:
                                Synthetic peptide"
REGION          1..16
                note = source = /note="Nucleoplasmin NLS"
source           1..16
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 46
KRPAAATKKAG QAKKKK                                         16

SEQ ID NO: 47          moltype = DNA length = 723
FEATURE          Location/Qualifiers
misc_feature     1..723
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
misc_feature     1..723
                note = source = /note="GFP-stop coding sequence"
source           1..723
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 47
atgggtgttta agggcgagga actgttcacc ggcgtgggtc ccatacctgggt ggaactggac 60
ggggatgtga acggccacaa gttcagcgtg tccggcgagg gagagggcga cgccacatac 120
ggcaagcttg caacctgaaatt catctgcacc acaggaaacg tcccctgtcc ttgacctacc 180
ctggtcacca cccctgacgtg cgccgtgcac tgttttagccg ctatccccca ccacatgaaa 240
cagcacgact ttttcaaaag cggcatgcctt cggggctacg tgcaagagccg gaccatcttc 300
ttcaaggacg acggaaatttta caagaccaga gcccgggtga agttcgaggg cgacacccctg 360
gtgaatagaa tcgagctgaa gggcatcgac ttcaaggaaat atggcaacat cctggccac 420
aagcttggat acaaactacaa cagggccacaaat gtgtacatca tggccgcacaa gcagaagaaac 480
ggcatcaagg tggaaatccaa tagcagacac aacatcgagg acggcagcgt gcaactggcc 540
gtcattacc agcagaacac ccctatcgcc gatggcttgc tgctgtgcc tgacaacccac 600
tacctgagca cccagagcgc cctgtctaaa gatcctaagc agaagcggga ccacatggtc 660
ctgcttggat tcgtgaccgc cgctggcata acactcgccg tggacgact gtacaagttaa 720
tga                                              723

SEQ ID NO: 48          moltype = RNA length = 141
FEATURE          Location/Qualifiers
misc_feature     1..141
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
misc_feature     1..141
                note = source = /note="GFP-stop guide RNA"
source           1..141
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 48
gggtccccgtt gccttgaccc accctggtca tagttccatg aaagccaaaa gtggcttga 60
tgtttctatg ataagggtttt cggccctgtgg cgtcgggat cgcctgcacca ttccgtatgg 120
tttctcccca ttatttttt t                                         141

SEQ ID NO: 49          moltype = AA length = 9
FEATURE          Location/Qualifiers
REGION          1..9
                note = source = /note="Description of Artificial Sequence:
                                Synthetic peptide"
REGION          1..9
                note = source = /note="meganuclease motif"
source           1..9
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 49
LAGLIDADG                                         9

SEQ ID NO: 50          moltype = DNA length = 318
FEATURE          Location/Qualifiers
misc_feature     1..318
                note = source = /note="Homo sapiens Human RNA pol III U6
                                promoter"
source           1..318
                mol_type = genomic DNA
                organism = Homo sapiens

SEQUENCE: 50
tgtacaaaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggtc 60
ggcaggaag agggcctatt tcccatgatt ccttcatatc atacaaggct 120

```

-continued

---

gttagagaga taattagaat	taatttgact gtaaacacaa	agatattagt acaaaatacg	180
tgcgtgaa agtaataatt	tcttggtag tttcgtt	taaaattatg ttttaaatg	240
gactatcata tgcttaccgt	aacttgcagg tatttcgatt	tcttggctt atatatctt	300
tggaaaggac gaaacacc			318
SEQ ID NO: 51	moltype = DNA length = 32701		
FEATURE	Location/Qualifiers		
misc_feature	1..32701		
	note = source = /note="Homo sapiens CFTR gene"		
source	1..32701		
	mol_type = genomic DNA		
	organism = Homo sapiens		
SEQUENCE: 51			
aatttggaaac aatgacatc	acaggcggtc agagaaaaag	ggttgagccg caggcacca	60
gagtagtagg tctttggcat	taggacttg agccccagac	gcctcagcag ggaccggc	120
gccccggaga ccatgcagag	gtcgctctg qaaaaggcca	gcgttgctc caaactttt	180
ttcagggttag aagggtggca	accggagctt gggaaagac	gtgcccacga aagaggagg	240
cgtgtgttat ggtgggttt	gggttaaagg aataaggc	ttttaaaaag atgcgtatc	300
attcattgtt ttgaaagaaa	atgtggat	tgtgaaataa aacagaaagc	360
gatggaaagaa tgaactgaag	ctgatgtat	agagagccac atctacttc	420
tttagaatctc aagactcaag	tacgtacta	tgcacttgtt ttatccatt	480
actaaaaata ctgttataa	agactcta	tatggttat tgggtttcc	540
ttggacactt gattgtctc	ttggacatac	cagggtccat gcttgcata	600
cagaaaaacat ttcttgactg	aattcagca	acaaaaattt tgggttagt	660
tgottaaagt atttattgtt	atgagactgg	atatacttag tatttgcac	720
tttttcaaaa attgaaagca	aatttgcata	ttgaaaaaaag ttacttcaca	780
agotataat tttaaaagc	ataggatata	ataccgaatg tatatccaa	840
taaaattgtat tcatagcc	atgtgatgag	tgacattttt	900
ttttttaaaat gcatctaagt	tcgaaatctt	aggcaaaatg ttgttagat	960
tatttgaagt gtttttgga	tattgcattt	actttgtcc tggattata	1020
tgaatgaata ggtactgtc	tctcttggga	cattacttg	1080
aagcatactg aggtatccaa	aaagtccaaat	atgttataaa tagctcatat	1140
ggggggaaagga atttagctt	cacatcttc	tatgtttt	1200
tccttggaaact ccgggtctaa	ggagagactg	ttggccctt	1260
gatggagag aaggacttta	ctctttggaa	ttatctttt	1320
ttgttactcc accttataaa	tcgggtttt	tattgtatc	1380
tcaaaaatgtt aatttggcata	aatttgcata	tttttttgc	1440
aatgccaacc agtctaaaaa	tgcagtttc	agagaatga	1500
aataactaattg aactttaaa	tagctacta	atatttcatg	1560
tttctttttaa caaatccat	ttgtatctgc	aaagtgggtt	1620
ttgcaaaagaa aatgttattac	gacacatttta	tataggat	1680
ctccaaaata tttcataagg	tcgtttaaga	ataggatgt	1740
ctatattatg gaactgcata	tcacatccaa	taccaatgt	1800
tctggactgc aatttctaaa	ttgttataaa	ttatgtttt	1860
gcttagttt caaagggtta	ggctttggag	tcagatttt	1920
actgttttagt agctctgtt	cctggggcag	gtcccttac	1980
ctttaaaattt tggactgt	catatgggtt	atctctgtt	2040
agcctaattgt taactgtat	tattatggat	tacatattt	2100
gcacctgtt aatataagat	gctcaatca	tccttgcata	2160
tggatatgc aatgatggc	ctgtatgaga	atthaattt	2220
ggcccttacca gatatacagg	aaacacgtca	catgttttca	2280
tagaattttt ctttctgaat	ttgtatgtt	ttttatgtt	2340
ttaggtatagt agttatgtat	aggatccctt	tcgttgcata	2400
gaaaacccaa aagtggaa	aaatccat	gaaatgtat	2460
gatttcacat caagtataaa	ttttttttt	ttttttttt	2520
ataaaaaataa cataatctaa	ttttttttt	ttttttttt	2580
atcagagacaa acaagttacat	ttttttttt	ttttttttt	2640
aagctatttt taaactagaa	ttttttttt	ttttttttt	2700
ggattgttata aataatggc	ttttttttt	ttttttttt	2760
aaatttacat ttcttcacaa	ttttttttt	ttttttttt	2820
ctctcatgcc ctgtttttt	ttttttttt	ttttttttt	2880
gtcacatgac ataactgtaa	ttttttttt	ttttttttt	2940
aggaactgtg ggaacccac	ttttttttt	ttttttttt	3000
tgtgggtct cctgttact	ttttttttt	ttttttttt	3060
tttaggtatca aggtggaaat	ttttttttt	ttttttttt	3120
taagataattt tgatgtactt	ttttttttt	ttttttttt	3180
aaacaaaatac acctgtat	ttttttttt	ttttttttt	3240
ataaaaaaccacacttctt	ttttttttt	ttttttttt	3300
attatgtatccat aataatccat	ttttttttt	ttttttttt	3360
cattattttt tttttttt	ttttttttt	ttttttttt	3420
ccagactgtt ctcttgcact	ttttttttt	ttttttttt	3480
cactctgtttt atataaagat	ttttttttt	ttttttttt	3540
taatttttttagt tttttttt	ttttttttt	ttttttttt	3600
tttttttttgggtttttttt	ttttttttt	ttttttttt	3660
aatcataact cactgtcaacc	ttttttttt	ttttttttt	3720
cgagtagctg ggttttcagg	ttttttttt	ttttttttt	3780

-continued

ttttttaatg atatacggac agtatctcac tatgttgcctt aggctgtctc tggagactcc 3840  
ggccctaagt tatccctcca ctcagtcctt caaaatgtctg ggattacagg cgttaaccac 3900  
ctcacccctgt cagccctaag acatgtctta atgaagaaaa atataatgtc tttagacaat 3960  
ggaagataaa ttaaaaattt actatgaaag atttataaaatg atgaccattt tgaatggac 4020  
cacacttatt tggttatata aattatgata cacttataaa aattatcatat gatgttttg 4080  
tatttacatt ttatccatc gtggcaatt tggagaaaa gctaaaaattt tggctaaagcc 4140  
ataaaatattt ttgcagttt ttgagggtt ttgtaaaaatg ttgcacaagg aagaccagg 4200  
ggctacccaa acaagggtt agtcttagtc tgatcaatc atacacattt tctcgatgtt 4260  
gtctatcaga aaaaccttgg tttatccaaa tcaaaaatgg atagatgtc aaaaacaaagg 4320  
ccaatatgtt ttgaaacattt atattgtatc atacaactgc caagcattcc cgattaccat 4380  
gactccattt agtcagtc tggccaaatg ccatcaatg ggcacggccca gggttccat 4440  
atttcctctt ggctttcatc cctatagaa ttggggggc ccacccctgg gatagggac 4500  
cttcgtctt gaacaatgt gtctgaacatc taacaatgt tgacttctt caccagtccc 4560  
tcaatagttt tttatccatc ttctttgtt gaccatgtt tggatttaca cttctcaggat 4620  
tttcagctg ggaatctgtt taaattttt attagctta ctctcaggat 4680  
tctaaaatgg cctccctgtac ctgtatatc aaaaaatttta taatttacatttattata 4740  
taaaaatctt ttgtatatgtt aaaaatattt tttatataatttattataat tggttttttt 4800  
aattttgcata aattttaaaaa gttccctttt gtttgaatg tttatctt atgtttttt 4860  
atgctatgtt aatttataat cacttgattt aagtaatattt ttatataatc tataaggat 4920  
atgtagttttaatattttttaatccctgtca aagagaaaaa tggaaatctt ttttttttttgc 4980  
tacttcatca atattttatc ttgtatgtt cttgttgc aacttccatc atccctcaga gctaactgtt 5040  
acacttaggg aatcagggtt ttccagggtt ccattatgtt gttatggag gggatggac 5100  
tttagttaat aatattcaat acataatgtt taacactgtt tttatctt atgtttttt 5160  
actgtatataat aatgtatgtt tttatctt gtcatttcc aagggccatc aatttctgt 5220  
tgatagatc aataatataat tttatctt gtttgc aaaaaccatc aggaagggtt tttatctt 5280  
agcatcgtt tttatctt gtttgc aacttccatc atccctcaga gtttttttttgc 5340  
ttggcaattt tgatctgtt gcatctgtt agggcatctt cttggagaaa aaccccttgag 5400  
atgcaatgg gtcggaaagg gaaaacagac tatgataaaatc atcaatgtt ttggagatct 5460  
tgttagaaagg ttaatccatc aatgttgc tttatctt gtcatttcc atggggaaa aatctgtt 5520  
ttctgttgc tttatctt gtcatttgc aatatttgc caatatttgc atttatctt tttatctt 5580  
attagcattt aagtttgc tttatctt gtttgc aacttccatc atccctcaga gtttttttttgc 5640  
aggcttctt tagcaagaag cttgtaaaaa tttatctt tttatctt tttatctt tttatctt 5700  
agaatattatc ttccatcaggat cttgtaaaaa tttatctt gtttgc tttatctt 5760  
caatggca ggcctgtcaat cttgtatattt tttatctt gtttgc tttatctt 5820  
tttttgcata taaatatttgc tttatctt tttatctt gtttgc tttatctt 5880  
tttgcataat tttatctt gtttgc tttatctt tttatctt tttatctt 5940  
tttttgcataat tttatctt gtttgc tttatctt tttatctt tttatctt 6000  
atcgatgaca ttcaacatgc ttccatcaggat cttgtaaaaa tttatctt gtttgc tttatctt 6060  
aagttggact ttggccatgtt tttatctt tttatctt gtttgc tttatctt 6120  
tgtcccttc gccttagctt ttgttagata gcatgttgc tttatctt tttatctt 6180  
aaggaaatgtt agatgttgc ttgttgc tttatctt gtttgc tttatctt 6240  
atgecttgc atagccaggat aacttccatc tagtttgc tttatctt tttatctt 6300  
gtacttctt tttatctt gtttgc tttatctt gtttgc tttatctt 6360  
tatttctatgtt gtttgc tttatctt gtttgc tttatctt 6420  
atggaaacaaa ttatcttgc tttatctt gtttgc tttatctt 6480  
geactttaacc cttgttgc tttatctt gtttgc tttatctt 6540  
tcacacttacca aggaggatattt tttatctt gtttgc tttatctt 6600  
aaaatattaaat cttgttgc tttatctt gtttgc tttatctt 6660  
atccctggaaa gttccctatc acggggacat aaaaatgtt acagaggattt ttggggatgt 6720  
ggctggaaaatc ctcttagcc gggccatgtt tttatctt gtttgc tttatctt 6780  
gtttggctt tttatctt gtttgc tttatctt 6840  
atcgatgttcat atactgttgc agtttgc tttatctt 6900  
gttggggatgtt gtttgc tttatctt 6960  
aaagaggatc tttatctt gtttgc tttatctt 7020  
gaaagaggatc tttatctt gtttgc tttatctt 7080  
gggtgtctt tttatctt gtttgc tttatctt 7140  
acctatgttca tttatctt 7200  
taggtatattt tttatctt 7260  
gatataatgtt gtttgc tttatctt 7320  
tatttgcataat tttatctt 7380  
actgtttaat tttatctt 7440  
tgtctgttgc tttatctt 7500  
acatatgttca tttatctt 7560  
gatcatatgtt gtttgc tttatctt 7620  
tgccttcgtt gtttgc tttatctt 7680  
tttagtgcataat tttatctt 7740  
cctccatctt gtttgc tttatctt 7800  
aggggacttca gtttgc tttatctt 7860  
catcacttca tttatctt 7920  
atatttgcataat tttatctt 7980  
accacatgttca tttatctt 8040  
tatttgcataat tttatctt 8100  
catcacttca tttatctt 8160  
atatttgcataat tttatctt 8220  
catggctgtt tttatctt 8280  
ttacatgttca tttatctt 8340

-continued

---

tttttagtct ttcttagttt	aaaataaaat	ctagaaaaaa	tttaagtct tctataactt	8400
ttttcggtt actgaataat	ttaaaaatgt	agtgaaacat	ttagacatgc aaaatggact	8460
tttcagaaga agaaaatggt	agcttaacag	ttattagatt	atgtccaga ataatttttg	8520
acttataatg ctctgttgc	catttcattt	cctctttttt	tggaatatgc atcttttaat	8580
gtgtccttca aggcaaaggc	tctatcttat	ctatcttgc	tcttgcattt tcccaggcca	8640
atgttttca caattttttt	aaaaaaacaat	actgtaatct	atttcaaat aaaattttcc	8700
atggggaccgc agtgtatata	aatagcagt	acaataaaag	ataataactc tcccatataat	8760
acaagaaac agtaaccta	gtgtctaaa	gtaaaggcta	cagtgttgc ttatacacat	8820
tatatgtataat ttcttgatc	ctatcttgc	gtgttttca	cagtgttgc ttatgtataat	8880
cgagatgc ttataattt	atgtttttt	aaacttgc	gccacaaggc ccaggaaataa	8940
gtttaatgtt gggaaacttag	caatacataat	aggatataat	actcttgc cctcagctga	9000
atttatttgc tggttgc	ccccaaagg	tgactcttaa	ataaattttag tttgttgc	9060
gaaattttctt gggataaatt	actttgtat	gtagttttga	aaaaaaaaaca ggtatattt	9120
agtctgaaatg ttgttgc	actaaggaa	tgttattttt	gtagaagtcg cctaagctca	9180
gcacttttatt atgcctgaa	attatactgc	ctgtccttca	ggtaagggtg ttatgtatgc	9240
agtttgcac tgaactcta	ttcatagtc	tgaaggctg	agagtactc agaqaataat	9300
ttttgtcttg aatataatg	acgcttagac	taaaacttta	attacgatgc tgaagaagaa	9360
agtggtaggt gattgc	atataatgt	aatatttttgc	atttttttttgc	9420
gttataatgt tgctttttt	ttggaaaggct	atttttttttgc	tttttttttgc	9480
aatatcacct taaagcaatg	acgtcatgata	aagtattttt	aaaccatgat aatatacatat	9540
gtggcttattt atttgc	tttttttttgc	tttttttttgc	tttttttttgc	9600
gcaaacagggt atgtcaaaaga	ttatgtggg	tttttttttgc	tttttttttgc	9660
agtgcacatgc tcataatgc	tttttttttgc	tttttttttgc	tttttttttgc	9720
ttatgtcaag gagatgtc	tttttttttgc	tttttttttgc	tttttttttgc	9780
agccaggat catgttgc	tttttttttgc	tttttttttgc	tttttttttgc	9840
gagaatgtgtt gatttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	9900
caaatcagaa gtttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	9960
atagtgtgaa aaccatgc	tttttttttgc	tttttttttgc	tttttttttgc	10020
gtgtcccccattt tccagccatt	tttttttttgc	tttttttttgc	tttttttttgc	10080
atcttcatgtc tcaactgtca	tttttttttgc	tttttttttgc	tttttttttgc	10140
agacagacatc ggcttgc	tttttttttgc	tttttttttgc	tttttttttgc	10200
tacttaatctt ctctgaaatc	tttttttttgc	tttttttttgc	tttttttttgc	10260
catggcacag tattttatgc	tttttttttgc	tttttttttgc	tttttttttgc	10320
aagggttcttca agagaatgt	tttttttttgc	tttttttttgc	tttttttttgc	10380
acacacacaa acatgttgc	tttttttttgc	tttttttttgc	tttttttttgc	10440
caaatctgc agtatgggtc	tttttttttgc	tttttttttgc	tttttttttgc	10500
gtctgaaggc agtctgttgc	tttttttttgc	tttttttttgc	tttttttttgc	10560
atacaggcct tcaaccgatt	tttttttttgc	tttttttttgc	tttttttttgc	10620
accaaaggatc atgttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	10680
aaaattaacc atactcttct	tttttttttgc	tttttttttgc	tttttttttgc	10740
agacataggc cgttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	10800
aatttagcacc ttttcatctt	tttttttttgc	tttttttttgc	tttttttttgc	10860
tatcccttc ttcctatgtt	tttttttttgc	tttttttttgc	tttttttttgc	10920
ctacaaatgc cccttcattt	tttttttttgc	tttttttttgc	tttttttttgc	10980
ctagcttttta aattttatgc	tttttttttgc	tttttttttgc	tttttttttgc	11040
gtctttaggc aattttatgc	tttttttttgc	tttttttttgc	tttttttttgc	11100
gtctgtatgc ctactacaca	tttttttttgc	tttttttttgc	tttttttttgc	11160
acctgtacgg ttgttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	11220
atatcgaaac atgtctaaac	tttttttttgc	tttttttttgc	tttttttttgc	11280
gggaccacca ttgttatatgc	tttttttttgc	tttttttttgc	tttttttttgc	11340
atcttataacc tccaaatgtt	tttttttttgc	tttttttttgc	tttttttttgc	11400
gttggattttc tctgttttgc	tttttttttgc	tttttttttgc	tttttttttgc	11460
taccacaaat ttgggtgtt	tttttttttgc	tttttttttgc	tttttttttgc	11520
aagtacaaga tcaagggtt	tttttttttgc	tttttttttgc	tttttttttgc	11580
catatccctt cccttcattt	tttttttttgc	tttttttttgc	tttttttttgc	11640
agggcatctc cctgtatctc	tttttttttgc	tttttttttgc	tttttttttgc	11700
tctgtgttca aacttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	11760
acttttagtgc ctcatttttgc	tttttttttgc	tttttttttgc	tttttttttgc	11820
cacatccctt gtagccggg	tttttttttgc	tttttttttgc	tttttttttgc	11880
caacccatataa caatttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	11940
accatataatg atatacatgt	tttttttttgc	tttttttttgc	tttttttttgc	12000
ctctatgtata gtggccata	tttttttttgc	tttttttttgc	tttttttttgc	12060
tacaggaaca atccaagaag	tttttttttgc	tttttttttgc	tttttttttgc	12120
gaatgtatgc ccatttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12180
acttagaaag aatatttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12240
atttaataatc ttgttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12300
agaaacatca ttaaaaatgtt	tttttttttgc	tttttttttgc	tttttttttgc	12360
tttactatag aatttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12420
aaagatgttca gctaaggatc	tttttttttgc	tttttttttgc	tttttttttgc	12480
atccatgttttgc tgaatttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12540
ggcactgtac aacatgtatgc	tttttttttgc	tttttttttgc	tttttttttgc	12600
gttagtcttgc tgaggggaaag	tttttttttgc	tttttttttgc	tttttttttgc	12660
tccatgttgc gacaggggaa	tttttttttgc	tttttttttgc	tttttttttgc	12720
cttcttgttgc accttacatgc	tttttttttgc	tttttttttgc	tttttttttgc	12780
atacacattt ttttttttgc	tttttttttgc	tttttttttgc	tttttttttgc	12840
tggggatattt atcttagactt	tttttttttgc	tttttttttgc	tttttttttgc	12900

-continued

---

caatgcattt	gaggttctt	ggaaatagag	gttaggttt	attnaagga	agttaccatt	12960
ttttttcca	gtgtatgtg	gttacccaa	agaatgcatt	gcccagtat	gtaatagga	13020
taatctttt	aaaatataag	agccaccta	taaatcata	gttcattca	gcgggagctc	13080
ctgcagagg	aaaaaagaag	agaatctgg	acacggttc	ctttaaagg	catttccta	13140
gagtgtgaat	ggaagcaaga	gattataaca	ttttgaggct	aaaaaaatc	tgaatgcct	13200
ataaaaaattt	ttttctcaa	attatcatca	tttgtgttt	taatgacctg	attgcaaga	13260
tgaacatttt	gaattctta	attgttatt	aggattgggt	aatgaatcaa	ttatctatta	13320
ctgtatgtt	tgctatgg	aaaaatagca	acttaatgt	tttgacagg	tttactttag	13380
tatatgtgc	tttatgaaa	aaaaagatgt	aaatattaag	taaaaggat	ttaagcaag	13440
gttttgagg	tagatgttca	ttatccctt	ggtaaaccc	gagccaaatg	ttgtctatgt	13500
tcttcgcct	tgcttgc	cttcctct	ggatcactg	tgggatgcg	ggattgtta	13560
tctggggatg	ctgtccatc	ctgcctct	caagcttgc	tattgtatc	cctcccaatg	13620
ataataaaagc	ttgaagaaaa	tgaaaatgc	gttagtattt	gtccctaaac	tcaagaacag	13680
gatggaaacct	aaatcttg	tcatacata	gtgtctat	actgtcccc	aaaaagagaa	13740
gtaaagaaga	tgctactt	ccctttat	ttgcgtact	tagcaattt	tttcttgag	13800
ggtaaqtga	taacactg	ggaaaaatgg	gttaaaatgc	caccaaaac	ccaattccat	13860
gttttagttt	aaagtggaa	atcagctgc	actggaaatg	ctgaatccaa	tgccatgtat	13920
ttttttgtat	cttcttg	aataatcat	ttgtgttt	acataccctg	ataacagagc	13980
agagaacata	aacaatgg	ggtaaggg	agatataag	agaagagaaa	aattccagaa	14040
tcgactgtatc	atttttatct	ttttatgtat	tttcaggcg	aatcttagag	accaacttta	14100
tcacaactga	attttaaaa	tcaccagtt	tgtcattgt	atgcagocat	agttttagt	14160
ttatccctgg	agtattattt	cttacatct	tttcatctt	aacattttt	aggtcactt	14220
tagtctctat	ttcaccatgt	aagaaacaaa	aatccccaa	ctatatcagg	ttgaatttaca	14280
cagtattttt	tttttaattt	ttggggaaatg	cgattcaagg	cagtaactt	caagtcgt	14340
tttagaaagga	ttaataaaat	agtggtttt	ctgtacat	agtggaggt	cattacatca	14400
tttgggttt	gaaagtca	aggatgttca	gtatgcgtt	tgccatgtat	ggttcatgcc	14460
aggcagat	ctgtacttca	taacccagag	cttacatgt	cattatgtc	cccaagagaa	14520
aatgtcacct	ccatcttca	ataaaacatt	tagcaaaagaa	aatcaatgt	ctttaatcc	14580
aaatcttg	ttatccatg	aataacaat	atggctcg	aaaatatgg	tatttctgtc	14640
aaaggacaga	gaaacatgt	agagatgtt	tactttgggt	cctagtgt	gtatctgaac	14700
aagcttaggt	aacaaatgg	ctcaataagg	gattttgg	tctagaaaaa	gagggaaat	14760
accaaaaata	tggataattt	ataaaaaaaa	taccagcaa	gttaaaatca	tatattatgt	14820
ggggagat	tttatatcac	tcatgttatt	tctatttt	tcctatattt	ggccaaggag	14880
aggtggaaact	tggtttctt	tttccctct	agctacat	ggacatactt	aaaactgttt	14940
ctctgttct	gttctctaa	atgtgttatt	ctaaacatgt	ccgtatgt	gttttgacag	15000
ttgcacaatgt	ttttttttt	aaatgttataa	atggccagc	agtaaccagg	ttggcatttt	15060
actataaaat	cttaaggcc	atccattcc	cccttccctt	attttcttg	tttcaatata	15120
atttttatgt	ccatgttca	taaaaaatct	aaatttagaga	gcaatggat	cccttgcatt	15180
gtgaatataa	agctcttca	tgtgtactt	taacggatgt	agcaatttt	agggttagtt	15240
atcattcttc	ttcttagcc	tttgttcc	ctacatgtt	ttggctctt	ttatgtgtt	15300
cattccaga	tagataaaaa	atacaaccaa	aatccgtt	aggcactct	aggatgtt	15360
tctaaatgt	atgggcatca	actataatc	acaatgttt	gtttccctt	ttatgtttca	15420
agatgggtt	gatttttttt	gtatgttca	ctctcaatgt	cttgcattt	ttatgttca	15480
tgaatataa	gtatgttca	ttgtgttca	ttgtgttca	ttatgtgtt	ttcaatgtca	15540
atgttacaca	ttcccttgg	tctttacat	actattgtt	ttacaactt	tatctgtt	15600
cataaaatca	taggttccc	agaagttt	ttccctgtt	ctagaggaca	gaaaacaaac	15660
aatttacaca	actatgtgt	catgttcc	gtccctttagat	ctcatgtat	atgtatgtt	15720
tcttaatata	tttgcgtt	ccacccagg	atttgcatt	ttgtgttca	ttatgttca	15780
aaagtcaacta	taatgtaaaat	tttgatttca	ccagttct	gagtggaaa	tcaatgttt	15840
agtaaaaactt	tctaaatata	acactgac	ttgtttct	tatactgtt	ttcttaatata	15900
tagctttat	gagatataat	tcatattca	aaacactt	ccattttaa	catacaatcc	15960
aatgatttt	tagtatctt	aaagatgt	ctatccat	aaccaattt	agaacacttt	16020
catcaactgt	aaaagaaact	ccatccat	tagcaatgt	tccttattcc	aaatccccct	16080
gctcgccct	gacaactaca	aatgtactt	ccatctat	agatttgct	gttctggaaa	16140
ttttatgtt	atagaacaaa	gtgttctt	gtactgtt	tatttactt	agcattttt	16200
ttcaaaatgt	catccctt	gtatgttca	tcatgtt	attttttt	attttttag	16260
agacaggccc	ttgtctgtt	gcccagg	ttgtgttca	ttgtgttca	atgggtact	16320
atagtttga	agtataggc	gaaagcggt	ctcccacct	agtcctccga	gtatgttca	16380
ctacaggott	gttccatgt	actgttca	tttataattt	ttttagagac	agggtctt	16440
atgttgc	aggctgtct	ccaaatcc	ggctcaatgt	tttcccttcc	cacagcatcc	16500
taaaatgtt	gttgcgtt	ttgtgttca	ttgtgttca	tttataattt	ttttaggtt	16560
tttgcataata	ttccacttgt	aaagatatgt	attttattt	ttctttcccc	agttatata	16620
tatttcattt	gttccatatt	ataaaaaatg	gtgtatgtt	cattttttt	ttttttttt	16680
caagttttt	ttgtgttca	ttttttttt	ttttttttt	ttttttttt	ttttttttt	16740
gatggccat	atgtatgtt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	16800
agtggtgtca	tcattttt	ttcccttcc	cattttt	ttgtgttca	ttttttttt	16860
atttgcacca	tttacttta	tctacttta	aaaatttac	ccatctact	ggccatgtt	16920
ttgttatttca	ttgtgttca	ttttttttt	ttttttttt	ttttttttt	ttttttttt	16980
tggattttt	tttgcatttt	ttgtgttca	ttttttttt	ttttttttt	ttttttttt	17040
taattttca	tttattttt	ttgtgttca	ttttttttt	ttttttttt	ttttttttt	17100
tgctacttta	aaaaattatgt	tgtatgtt	ttttttttt	ttttttttt	ttttttttt	17160
cacaatctt	atttttttt	tctacttca	ttttttttt	ttttttttt	ttttttttt	17220
tagcctcgt	ttctcaact	taatgttgg	ttttttttt	ttttttttt	ttttttttt	17280
gtggggatca	gttgcataaa	tgtatgtt	ttttttttt	ttttttttt	ttttttttt	17340
gaataactat	tttagctt	ttttttttt	ttttttttt	ttttttttt	ttttttttt	17400
aaatataaacc	aaagccaa	agttgtatgt	ttttttttt	ttttttttt	ttttttttt	17460

-continued

---

gctttataga caacattgaa tttggctc atgaacatca ggaatgtgg tcatggtagt	17520
tattatccg accaggaaact gtatgtgaga gatcttcaga gcttttcca aggcgatata	17580
actggaaat aatagagaca aggttacaag ctatggctgt gttttctt taaaatctt	17640
agttcagttt tttcaataa cagatttgta gttaggcata ggtgactggg gattcgatt	17700
cttcaagttg aaatattacc ttgttgagaa agaaaccatg tgtgagacaa ccatgttgag	17760
aaagaaaaag tgattttata gaaaattaat attgtatagt agcattat gaaaatcatg	17820
aagttgaac atatttgcc agaaaattta cattaatagt taccatagc attaatgc	17880
ttataattac acataccctt tccttaatga aaaagaattt tttcccttca aagttatgc	17940
tgctattgtt aaacattaga gaatatacg aagcaaaaaaaa gaaaatatct ttttgat	18000
tttcttaaca tacgttgtt cctaataatg tttatgtt agaagcatg catgaaatgg	18060
gtagatcaat ttcttattta atgttggat tcattaggtt cgaagtttgc aaattaattt	18120
ccattagggt gctgtatgg ttgttaatcc tggacccgtca gaagattttt cagatttgg	18180
ttgttagtctt ttgttagca gcaaaaattt agttctccag agttctgaa attatttgac	18240
cactttaatgt tggttacctt accttagagaa agaaaaaaagaa ctctccaaag tcccttggt	18300
aaattaagcc tcatgacaa ttaactcaaa tatacacaag gttgttctt agcgagacata	18360
tactccctaa agttgatcaa gctgaccaag tgattactgc ttataaattt accattttat	18420
ggagaagaag caaacactgc taaataccctt gtggaatcag aggaggggaa atttagtaact	18480
tgaccccaat actcgatattt taaattgtat ttttgcggcc tacaagtttt acacaggact	18540
tttagagact ggtatgtatc accttgcata gtcctactt tactatgtt ctttgagaaa	18600
aatacatctg actaaataacatc tctgatataa aattggatata aataatgtg acattcaaaa	18660
tgttatttat gattttagaa aaatatccctt atagacacta gatgagttt agtctcaat	18720
caatccccc tatcatatgc acattatcaaa ataactaaag caaagtgtt gagctgtgt	18780
ctagaagttt gggattttatc atcacaatgt tttccatgtt gtcctctt ccctctgcct	18840
gttcttcaaca ttgtttttt ttttttttggacttccatgtt gtcctctt ccctctgcct	18900
tcaaactcat ggcaataaca ttggatgtatc agaaattttc taatgttattt gaatttgct	18960
acacaaacta gagaatttgc tattatccatc tcaagtttgcatttccatgtt gcaaaaagga	19020
aaaaggcttat tagtctttaa gtttttttttgcatttttgcatttttgc ctgtcaat	19080
tttagtgggtt gaaaaatttttgcatttttgcatttttgc ctgtcaat	19140
aaaaaaatgtt gtttttttttgcatttttgc atttttttttgc atttttttgc atttttttgc	19200
ttgaaaatgtt ctgatattca tggatgtatc ctgttttttttgcatttttgc atttttttgc	19260
gegtcttttgg ttttttttttgcatttttgc atttttttttgc atttttttgc atttttttgc	19320
atatgacttttgg gggatgtatc ttatgttttttgcatttttgc atttttttgc atttttttgc	19380
agttcttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	19440
agaaatgaaa ttactatag tttcatctgt gttgtgtgtt gttgtgtgtt gttgtgtgtt	19500
ttaaaaaaatc aatgtatgtt gcttttttttgcatttttgc atttttttgc atttttttgc	19560
gaacagaacaa gtttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	19620
catagtaaaa ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	19680
tagtttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	19740
tcaggactt gggggggccca aggaggaggatc atcttcttgc cccaggattt tgcaaccagg	19800
ttggggcaggc tggtgtatc ctatcttcataaaaacaaacaaacaaaggacataaa	19860
gaaaagaaaaa aatagccca gcttttttttgcatttttgc atttttttgc atttttttgc	19920
gtcgtcaata ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	19980
ctagagggtt gggggggccca aggaggaggatc atcttcttgc cccaggattt tgcaaccagg	20040
tttttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	20100
tagtggaaat aatgtatgtt gtttttttttgcatttttgc atttttttgc atttttttgc	20160
tcagggtttt ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20220
gtttaatgtt atatgttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20280
gagaatttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	20340
tgagacggatc ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20400
tgcaaccccttccatgtt gtttttttttgcatttttgc atttttttgc atttttttgc	20460
attacaggca ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20520
tttctccatgtt gtttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20580
gcttcccaaa gtttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20640
ttttcattca attttttgcatttttgc atttttttgc atttttttgc atttttttgc	20700
gttgcatttgc ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20760
ccaattttgtt ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20820
aagtgtatctt aatgtatgtt gtttttttttgcatttttgc atttttttgc atttttttgc	20880
tttgatgtatc ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	20940
aacttattaa aatgggcaca gataacacaa gaaatgtttt ataaaacgc ctcacaaat	21000
gttgcatttgc ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21060
aaaaagaggatc acgttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21120
gaaaatgcctt ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21180
tagtgcata ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21240
gcacatttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21300
ccatttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21360
tcttttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21420
ggccatttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21480
agtttatttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21540
tgggagaccg aggccatgtt gtttttttttgcatttttgc atttttttgc atttttttgc	21600
tgggagaccg ctgttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21660
tgttatttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21720
gttgcatttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	21780
gtctcaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa	21840
acaaacaaaaaa aacaaaaat ttttttttttgcatttttgc atttttttgc atttttttgc	21900
ctttaaacgt ttttttttttgcatttttgc atttttttgc atttttttgc atttttttgc	21960
atttttttttgcatttttgc atttttttgc atttttttgc atttttttgc atttttttgc	22020

-continued

---

tatatttttgt	tgctttaaaa	tgatgacgt	gagagtattt	atatcataga	aattggcaaa	22080
tgccgtaaatg	cagtttttgt	ttttgtttt	gttttccgga	gaggggatg	ttaaatattt	22140
gcctgcatgc	aacaccacta	catgcagtct	gctatcttt	gttcttcctg	ctttcaggtct	22200
cctctcccg	ctgtctgtct	agcacaaacc	agcataccaa	atttcttaa	ataggaaag	22260
ttgaacatgg	aaaagaatg	aatgaagtca	aaagaatgtg	gaaagaccta	ggcttggca	22320
tttagtaaag	tttagcatct	ctaagccctc	atctctttat	caataaaatt	gagaatgtat	22380
cccttttagt	tctaccatt	taagaagatt	ttcaaatgaa	aaccacaacc	tgctcatgtt	22440
tatgaaggca	ctttggaaag	cgctaaatac	acgggtttt	attagtagta	aacacttaact	22500
tcacctttt	cacttctga	ctttagttt	caagggtct	taatctaaat	tatatcataa	22560
attgctgtcc	catagttttt	tacagctta	ttgcccaccc	tatgttcac	tttcttctg	22620
ttctttatgt	tagatactgg	gatagatgc	accaagggtgg	tgtgcacata	ctttctcaga	22680
tgtatgtccac	tgaagacctt	gcatgtat	ggcatttcatt	ttcctgtgt	attcagactg	22740
gcctcaacta	ttttctttat	tgctctccag	gaaaaattac	aatgaatca	gactggcaaa	22800
tgaagggtaa	acctaattat	cgctttttgt	taaagacagc	tcttgttaaa	atgcggat	22860
tgcaaaattaa	tggaaaaat	atgacatgt	aaacacatact	cacttattaa	tatcttagta	22920
aggaataatt	gatgaagtta	cttaacctta	gagccctaat	tcagtaatgt	tttaatgaag	22980
gacaaggtgt	agagatatcg	agaacccagg	gcaggtgcct	actgaagaag	ttccagacca	23040
aggaagtata	aaaaggacc	tgggtggag	cagttagat	ggatgtgg	gccactggca	23100
aagttttgcc	cggaaacagt	gtcääaatgt	ttgcattttg	catagccct	tcttttttgc	23160
ttctgaatgg	ctttgtctaga	atatcttc	tataatgaat	ttatctgt	tctcagat	23220
tgctaaagca	ctcccccttgc	aattttgggt	ctttaacatg	cattttgata	cattacccaa	23280
taagggtctga	atgacacaaa	ttttagaact	ctccagagaa	aagaagatg	ctgagggaaa	23340
aagcataggt	ttggggactca	ctaaatccca	gttcaattcc	ttctttaat	aaatataattc	23400
aatttttagt	gagaaagctc	tcgtgtctc	gaattttat	ttagaaatttc	tctttgtaca	23460
tgattgtattt	cacaatccct	cttctgcctc	ctttctact	ttcttcttc	tagatttcc	23520
tatctttatg	aagattttc	tgcccttatec	tcaacagtt	gaaacaat	ttttgaaaat	23580
cactacgta	toctgtcat	tgatccca	tgccaaactt	actaattttc	attataattt	23640
attatttttt	gatgcctaga	ggggcattg	gtgttagtgc	tatggagtgc	ggagacaaaa	23700
cataagaag	ttatgtatcc	accctcagg	aatgatttc	acatgataat	taatgtcaaca	23760
aattgtataga	aactaattac	taactctctg	gctatagtca	ttctttcaat	gaatagctca	23820
ttactgtgat	tgcatgtctc	agtaacaaaa	ttatataagg	ctgtgtat	aatgttgatt	23880
aagtgcattgt	cttattcaga	gttttttat	atttggaaatg	gaagggctg	gacttcagta	23940
atttgcata	aactgtatgt	atatgattat	tttggggcag	ttatttttt	aagaataatt	24000
taaatatgg	atgttttagca	gtttgtttt	ttccctggaa	aaaccataact	attattccct	24060
cccaatccct	ttgacaaatgt	gacagtccaa	tttagttcag	gatattgtat	ttttatcacag	24120
gtgttagctg	taagagatga	agcctggat	ttttagaaat	tgacttattt	tatttcata	24180
tttacatgtg	cataattttc	catatggccag	aaaatgttga	tagatcaga	ttccaaatct	24240
gtatggagac	caaataactgt	gaatatctgt	ttccctctc	tttattttat	ctggaccaga	24300
ccaaattttg	ggaaaaggata	cacagcggc	ctggaaattt	cagacatata	ccaaatccct	24360
tctgttgatt	ctgtgtgacaa	tctatctga	aaatggaa	gtttagttca	tgtacattgt	24420
tttagttttag	agagaaattt	attattttaa	ttattaga	agagaaagc	aaacatattt	24480
taagtttaat	tcttataattt	aaaatagga	gccaagttat	gtggctaat	cctgtatcc	24540
caactattt	ggagggtcaag	atgagaggat	tgcttgat	caggagttt	ataccaggct	24600
ggcaacata	gcaagatgtt	atcttcatac	aaaataaaaaa	agtttagtgc	ggaggttagt	24660
gcattgttgt	atcccaact	atcaggagg	ctgaaagg	agggtttactt	gagccaggaa	24720
gtttgaggtt	gcagtgtat	atgattgtgc	cactgcact	cagttgggt	gacacagc	24780
aaccctctt	cttctaaaaaa	aaaaaaa	aggacatct	cattttcaca	ctgaaatgtt	24840
gactgaaatc	attaaacaaat	aaaatctaa	aaagaaaaaa	atcgtttcc	taagaaatga	24900
tttttttcc	tggaaaaat	acattttgg	tcagagaa	tgtcttata	gagaccatga	24960
gatggatttt	gtgaaaacta	aagtaacacc	attatgtat	aaatgtgt	tatttgcatt	25020
caaaacccctt	atattttat	acaaatgtat	ttccctggaa	gtcttaatgt	aatggctact	25080
ggtttatcaaa	caaataatgt	aattgtat	ttttgat	ttttgat	ccaggtagaa	25140
tatctcttct	cagccactct	gatgttggaa	catcattatc	tctatttatc	agaaaaagca	25200
actgaggctc	agagagataa	tatactttgc	cagttaatga	atgtggagc	catgttcca	25260
gtctggatct	gtattgcctt	gtcttcttgc	aatggatgt	ccccccat	agaatctctc	25320
agtttccctt	ccaaatccaa	ggtttagat	ttttgtat	ccagtgcac	aaacccatt	25380
taactgttta	agtaaataaa	aggaacgtat	ttatgttgc	atgaagctt	aactatgtga	25440
agaccttagt	ggagaacttgg	cctttaggaac	tcaatggac	caaggacta	aatgccac	25500
ggttggcattt	gtcttcttgc	ggttttttat	tcttcagacc	gaccgtt	ctacataaag	25560
ttgggtccctt	gtttagactc	ttttgttctt	tctttttag	ccacggaa	aggagccctt	25620
tgtcccttgc	taatgttgc	aaatccca	gactcttgc	tcatagtgt	tacccttgc	25680
gccactcata	gtcttagat	aaatggctg	agtctgtgc	caacagcaca	ggcctgtat	25740
cagataaaag	ggtttagtgc	ggggataaa	aaataagaca	tagtactaa	attattgcac	25800
caaagtaaaa	acattttgtt	gtcttgcatt	ttttagttat	ttttttat	tttttttt	25860
ttttggcattt	ttttaggc	ttttttttat	ttttttttat	ttttttttat	ttttttttat	25920
gaattaatgt	agttggatgt	ggcatttgc	ttttttttat	ttttttttat	ttttttttat	25980
ttgaacactt	gggtgtgtca	aatgttgc	ggtagggata	ttttttttat	ttttttttat	26040
tttattttttt	tcgtatctt	ttttagttat	tgatccca	agtgccttgc	atagggcct	26100
gcataatgtt	tatacttcaat	aaatattttgt	tgaatgtat	catgttgc	ttttttat	26160
ctagcattac	atagaaaccc	gttgcatttgc	ttttttat	ttttttat	ttttttat	26220
ttgcaggat	atgttgcatt	ttttttat	ttttttat	ttttttat	ttttttat	26280
tgtttttttt	gttttttttt	ttttagat	ttttttat	ttttttat	ttttttat	26340
atgggtcaat	tttgcctc	ttttttat	ttttttat	ttttttat	ttttttat	26400
cagcctctcg	atgttgc	ttttttat	ttttttat	ttttttat	ttttttat	26460
tttagatgt	gtggatgtt	ttttttat	ttttttat	ttttttat	ttttttat	26520
tgatctgccc	acccatggcc	ttttttat	ttttttat	ttttttat	ttttttat	26580

-continued

---

gctttgaata	tctaagttt	aattggatgc	tgagggaaat	attaatcaga	gtaggcgctgg	26640
gttaattgaa	aatgtgata	cattttgtt	tatggccaga	tagagaacat	gaatctgaat	26700
ttgcagaattt	atctcggtta	acatttttt	cttcccggtt	ttcaactgt	cccccatgtt	26760
gattcaattt	aaaaaatata	cctatttac	tcaattcaa	caatgtatg	ccagtcacaa	26820
cccatacgtt	ctattatttt	tgttttgtt	tgttttgtt	tctccaccct	gttacttctt	26880
ttcttataaa	atggatattt	gaaatttt	gaaatattt	ggaagagtg	cataccatt	26940
tttgtacttt	gtacccctgc	accctggga	agtgaccctg	gcttcacatt	tcataactgc	27000
cttgcacca	tggccctcaa	gtggtgca	gatggttgaa	gaacattaaac	ctatctggct	27060
caattttgt	accatggatt	gaatccctta	cataactgca	gtgtgcaaa	cacacatccg	27120
ttccaagatt	gtagtcagga	tatgaaactt	ttaagaataa	aacttctcc	cttctgatct	27180
gggcctggta	tgtgtctca	ctagaaacc	atcacact	cttggcgta	acaatttgg	27240
gcaccaagtt	gttcaagttt	cacccattaa	agaatccc	cgacccgtcc	ttctccctcag	27300
gtaactaccc	cattctat	tttcttcat	agctaacatt	ctctgetctc	ctggctctc	27360
tacttcaact	tcatatccat	ctcgcctc	gaagttatgtt	ttccaccat	ttccctaaaac	27420
tacattggcc	agggtcaacta	gagacccctt	atgaaatata	acaacaccc	tctacattac	27480
ttccgtgtgg	accactttt	cacattgtaa	ccatttgtt	ggtttatgt	cacaccctt	27540
ccttggctt	ccatctgtat	ccatttctcc	tttgatggag	aaggtagtc	tgctccat	27600
ttagctttt	actctcgat	accaaatgtt	atggatggaa	ggttagct	gtgtgtgaga	27660
gaaagggtg	gaagcatgt	ggggggaaa	tagatggaa	aaggtaat	ggctttatag	27720
aagggccttc	atagcaagc	ttctagggg	tgccaaagatc	catgcttgc	gattggcagg	27780
cttgccttc	aatctcagct	gtgttattact	cctttatgtt	ttttgtttt	ttgtgtgtt	27840
tgtttttgag	acagagtctc	gtctgtc	ccaggctgg	gtgttagt	gtgtctcag	27900
ctcaactgca	actctcgctc	ctgggtca	gcaatctca	gtctctgt	tagctgggac	27960
tacaggcatg	caccacagg	cctggctaa	ttttgttag	acggggttt	gctatgtcg	28020
ccaggctgtt	cttgcactcc	tgacccca	tgatctgc	gccttggct	cccaaagtgt	28080
tgggattagt	ggcgctgagcc	actgccccgg	cctattactc	cttttagtgc	atttagagcc	28140
atgttactt	atggtaactt	gacagatgt	ggaataacc	ctgatgaa	gtaaaggcctt	28200
tgtcttaat	tttacatgt	tcttcctgt	gggtcatgaa	attttattc	tctgtacagt	28260
ttgaaaat	agatgataat	attttagat	attttattc	tttgtgaga	aaaaaaaggc	28320
tttcattaa	agaaatctgt	ggcaataact	taataat	aatcagctt	ttgtcctata	28380
gtatattttt	gaaaacagaa	agtttactat	atttcagoc	ttttcagac	tatthaatgt	28440
ctatttttac	tataatgaa	atataat	tttgcattgt	tctcatttt	caaagactt	28500
taatacatg	tttttagt	tgaaaatgaa	gtttaatgt	agtttatg	tctactttt	28560
aaaacaaatg	ctaacagatt	tttctcatgt	taatcaca	aaaaagccac	ctgacattt	28620
aacttgg	tgatgtgca	gtgaaat	ataaatctc	cacagttca	aaaaaataaa	28680
gatcaaggta	taaggggaaa	atgtgaaat	tttgcgtt	tatttttcc	acccgttct	28740
aagcacacg	atgaggatcc	gtaaaaggct	tactttt	gtccaccctt	ttcattgtt	28800
tttagaagcc	caacactttt	ctttaacaca	tacaatgtt	cctttcat	aatcaattc	28860
cctgcacagt	gatatatggc	agaggatgg	attctgc	atatctgct	gagtgtttg	28920
tgttgtatgg	tctccatg	attttgc	tataatact	gggttaat	cttggat	28980
acttgcgtg	atcaaactat	gttaagggg	ataggacaac	taaaatattt	gcacatgca	29040
cttattggtc	ccactttttt	ttctttgc	gagaatggg	tagagatgt	gttcaaga	29100
aaatctttaa	actcatat	gccctcg	gtgtttttt	ctggatatt	atgttctat	29160
gaatctttat	atatttaggg	gtaaaggat	cattttgt	tccat	atcacataa	29220
tatatttatt	tttgcattt	tgaaaagact	acgaaatct	gtgatagt	gtaaaat	29280
aaaggatgaa	tccaaactcc	aacactaaga	aaccac	aactctag	aggataagta	29340
aaatcttta	gaaactaaa	tgtctgg	cacgggtgg	aatttacat	ctcaatggc	29400
tcagcaaaaat	aaattgtt	ctttaaaaat	tattttgt	tatgttca	aatcacattt	29460
tcttactat	acatgaggat	actgggt	ttattttgt	gttcaaca	ggagatgt	29520
aggagacaat	gtcagcagaa	tttagtca	tgca	tatcatat	aatgttgt	29580
atattttgaa	atcatatctg	catgttga	tgttcaaa	aaaacacta	aaaatttaaa	29640
gtatagcgc	ttaataactat	aaataat	tactaaaat	ttaaaatgtt	cttgcataat	29700
attttttttaa	tatcttcat	ctcatcgt	tgaaaatgtt	cacacgtt	aatccaggct	29760
tttgtgtgtt	taagtgcctt	gtatgttcc	cagttgt	ccatgtg	tctgtatt	29820
tatttttat	atcatgaa	cattttt	atcttgc	gtaaacct	aaaggagaca	29880
gattcaagac	ttgttataatc	ttcttgc	agctgtgc	aaatattgt	ttggggcg	29940
tacttattat	atggatgt	ttgttataat	atggtctt	atgcctc	aggat	30000
gtgtctccca	gagaatggcc	atttgtt	cattgtt	aaaatttgc	ttcatacacc	30060
cccataaaa	atacat	aacttatt	aaacaaatgt	agtacactt	ggccca	30120
gttctctat	gtcttgcata	atttc	agaaat	ctgactt	aaatata	30180
tccataat	atattttat	ggaaatctg	aaccatttgc	gcattttgg	gtaaaaatgt	30240
ttttattat	aaatttaaat	gaggtagt	gataat	ttactttt	gttacctt	30300
agatgat	tctcaatc	agcaccaccc	agagctt	gaaacaat	tattcac	30360
tctgtat	atttgtat	atttttag	aaataatgtt	gctgggt	ttgaatcagg	30420
gtatggat	cagtccact	tgatc	atataat	tgtaaat	taacattt	30480
aataatgt	tggtggat	aatgtat	tatttgc	aattgtt	tcatggcc	30540
gatttcgt	aaactcaaaa	tttctcttgc	ttgtgttct	cattgtat	tgcttctt	30600
gattaaccta	agcactact	agtagaa	ggagaggg	tctaaat	aggcccc	30660
ctatgcctc	cttggctgtt	aaaataat	atttctct	atccac	catagt	30720
tcatgtat	aaaaacaccc	accttatt	caaagt	atgat	acat	30780
tagaattttt	tttgcatttgc	tgcatt	aaaataaa	tgtat	gat	30840
tttattgt	acctaatact	ttccatgt	atata	gttgc	attttt	30900
ttaatttcc	tagtacttcc	atggg	aaaattat	tct	acaggaa	30960
gagcgtgagg	caaattaat	aacttccca	aatgtt	acaca	tttgc	31020
tcccaatctc	tttagtcc	aagcc	ccctt	cacc	atacc	31080
ttaatataat	attataat	taatgtt	attagt	tca	acacaa	31140

-continued

---

```

gtaaccagcc ttggttctgc tgagtttctc tgcgcttctc caggagacac aggctacaga 31200
gcttgaaggc tgaggattct tccagggtca ctccaggggc aaatctgaaa ctttcttcag 31260
gacaggaatc aacgagatct tctcaacttac ttatacctgg gggaggaact gtatgaardt 31320
caccgaagaa ccagtcatgc taaggccaa acctatagac aaaaaaaaaa ataggagaat 31380
ggagtagatgt tggagaaaaga ctaaatgtt cttaaacttc tcaagttaa aaatatccca 31440
gcaaaaagaga tcgtaaaaagc ccttcatggc gtatattaatttccatgcata ggggtgagtg 31500
gaaaggtaact ctggagcccg aggtacago tttggaaacttgcagcacctt gtaaggggaa 31560
agcgtgttcc catcatctca actctactg ataaccaatg gaatatttgtt gagaaggaa 31620
tcctggggaa agaaggagct gaaatgttga ggtgagaagg cagagagaag aatatttata 31680
tttgggatgg cacaatgttgc atgaggctgc aggtttttcc ccttgcata gagaaaaaaa 31740
ccacgctgac accatgttgc ttacatgttgc gagaatgttgc caaatgttgc gatcttaat 31800
aatttagata aacatgttgc ccattttatgttgc ttttcaggat atctgtatctc 31860
tttgcatttc tcttttttgc tcttataaga ataaaaagggg gggagaaaaat tttagcatta 31920
tagtattttctt ctacatcttc tctgtccctt tacataactt acaccatgtc cttcttattt 31980
atggtttatatttgc ttcttcttctt ctttgcata gcaaggataa atgagccagg 32040
gattcttgc aactatgttgc cactttctt ctttgcata ggtcataactt tcagaaatctc 32100
tacacattctt tagtccctctt aaacaatgttgc agttgtggca taaaatattt tgcttgggtt 32160
caggactgtt agagaaaaat ttgtgttgc ttgtgttgc gtaatcttgc ttgtgttgc taaaatattt 32220
ggaggttgc ttcatgttgc gtaatcttgc ttgtgttgc gtaatcttgc ttgtgttgc taaaatattt 32280
tagctatgttgc ttcttccatgttgc ttaactatgttgc ttgtgttgc ttcatgttgc ttgtgttgc 32340
agatagatgtt aacatgttgc ttaacatgttgc agactcttagt caagtccggg cacgggggct 32400
cacacgttgc atccacgttgc ttggggggggc cgagggtggc aaatcacaag gtcaggagtt 32460
tgagaccaggc ctggccataa ttgtgttgc ttgtgttgc gtaatcttgc ttgtgttgc taaaatattt 32520
aggtgtgttgc gggtgtgttgc gggccatgttgc ttgtgttgc ttgtgttgc taaaatattt 32580
agaatagtttgc gaaatgttgc ggcagagggtt gcaatgttgc gatctgttgc cactgttgc 32640
cagcctgggtt gacatgttgc gactccatctt ctttgcata gaaaatattt aagactcttgc 32700
9 32701

```

```

SEQ ID NO: 52 moltype = AA length = 1060
FEATURE Location/Qualifiers
REGION 1..1060
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION 1..1060
note = source = /note="nAPG00969"
source 1..1060
mol_type = protein
organism = synthetic construct

```

```

SEQUENCE: 52
MKQTTSYSLGL AIGIASVGYG LIDNDENIVD AGVRLFPEAN SKNNDRRK RGSKRLIRRK 60
RHHIARVKHL LKESGIDVSY ENTVLTPYE IRCKGLTLPN TNHELSIALL HLAKRRGVHN 120
VKSMMDQEKVKM GNELESTKEQL TINDNLLEEK FICELQLERL NKEGIVRSHS NRFKTADYIK 180
EIKNLLETQA KQNTLVTEEF IEKYIEIFSG RRKYYEGPGG ESKYGWKGDI EKWEYGLMGK 240
CTYFPKELRC VKHAYSAALF NLLNDLNNLS INREEDTKLS QYEKEQIIIEK IFKVRKTPTL 300
TQIAKLLKVD PTNIKGFRTK ANGTPDFLSI KIYHDLKGII DDKQLLDDIA FLDNVQAQLT 360
VWQDSQS1QE KLKTLNKNLD DKTKEISEL KKYTQTHSLS LKLNINVLLPE LWETTKNQMT 420
ILSELKLKPR KIDLHNCNEI PVNMINDLV SPVVRSLTQ SIEMINQIHK DYGHPREIVI 480
ELAREKNSEE KNKFIKSLENE KNKQINDEVI EKLNASNHRD NKGMFNKVKL WILODGHCLY 540
SLKPIRLEDL LNNPNHYEID HIIPIKSVSPD DSMSNKVLVY QIENSKKGNR TPYQYLTSA 600
KTITYEKFKA NITQLAKSNH KISKKKLDYL LEERDINRFH IKKEFINRNQ VDTRYATRSL 660
INNLKYYFSE KDINVKVKST NGSFTDYLRK LWNPFPKDREF YHKHHAEDEL IIAMANKIFT 720
TRKIFKEQNS VFSDEQILDG EVTNQFTEKFY KVQAIKKYDK YKYSHRVDKK 780
PNRQLFDLTT YSTREFEGEE YYIGKIDYI NLKDKRLKKI FTKSPEKILM YQHDSQTFKK 840
LKQIMRSYED EVNPLAKYHK ETGEYLRKEC KKGNGPIVKS LKYRVTKLGV HKDITHKYEN 900
SKNKVVILSL KPFRMDVFKE NGVYKFFITIR YCDLKETVNS YTISEHLYKA KLKAKDIKSM 960
DSFKWSFYKN DLLEYNGELC TFKGVNDDKK NKIEVNWVEK NMFAIYAEKKN LKSQQLVKSI 1020
TKSTVKSLLK YTTDILGNRY PVRNEKLKLM IRKQTFRGDGL 1060

```

```

SEQ ID NO: 53 moltype = AA length = 1092
FEATURE Location/Qualifiers
REGION 1..1092
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION 1..1092
note = source = /note="nAPG06646"
source 1..1092
mol_type = protein
organism = synthetic construct

```

```

SEQUENCE: 53
MRYSIGLAIG TTTSIGNAVIN KDLQRFEHLG VRIFDAEAPN KDGSSLSAPR RLARSSRRRL 60
RRRKHRVERT KQLLINKGLL TKTQVKNLYN SKNINLDIWD IRVSGIDRKL FNNEFARVLI 120
HFSKNRGFKS NRKSELKEDD NGAILSAVKE NRELMDEKGY RTIAEMLVSD EKYEGTKRKN 180
GGDYSHVVAR SDIENEICLL FQKQREYGHF FATEENEEAF LSIWSSQRPF STKDDIVKKI 240
GNCCTLEPKEK RAPKSTYTTFE RFRALDKLNR LRILSTTAPS RPLTNNEERKS ILSSLFSKKE 300
VKYKELRKL KLTDDQRFNE IYYSPDETIE KTEENRTFLSL ESQYKIKKII EKTESKNMQS 360
SYHPIDYDTI GYALTIVFKDD KDIQHYLQNS YIDSKGKAIP NMANREYNLE LIEELLGLSF 420

```

-continued

AKFGHLSLKA	LNNILPYMEE	GEPYHIACEM	ASYQFSQRRLS	KEKHRLLPPI	PVDEIPNPVV	480
VRALTQVRKV	LNSIIKKYGP	PSDIYIELAR	EMSKPFKBRK	SLEREFNENR	QINEKAKAHI	540
SELYRIPNDP	RPHDILFKFL	WNEQNGICPY	SLKPISIEYL	FNIGYAEVDH	IIPYSRSFDD	600
SNGNKVLVLT	RENQNKLNR	PYEWFGHEEN	RWEDFVSFIR	TMVKGKKKN	MLLKNPDEE	660
QEEQIILSRNL	NDTRYITRYI	KSFIEDNLEF	RTEENKEQYV	HTVNGAYTSL	MRKRWGLNKD	720
RGRNDLHHAV	DAAIITAVSLP	FKNKVNAYFP	RQETGLSKLL	NNKKDIFPEP	WRNFIKELEA	780
RMIQDPKMK	RALESLELET	YGEIFLNKLK	PIFVSRMPK	SIKGQIHEET	IRRVRGFTEE	840
GFLVTVKTR	LDQIIPFDKNG	DFFPMYGKETD	IKYTYMAIKQR	YLEYGQDKQK	AFAVPLRKPS	900
KNPKNAPIVR	SVKIEGKANR	VVMLDDKAAA	DNASIVRTEV	FRHKKTGEYY	LTPVYVADIL	960
SNKIPDRLLT	IKKSYSDWDR	ITDEHEYLF	LYNNDLVKII	LPKEKETKVK	TGGNHLWQEG	1020
FFYFKGVDS	NAGIKIINHL	NSFEARIGTK	RLIAFEKYQV	NPLGEINKVH	GEKRPGEMLN	1080
KEEIKENRKN	IS					1092

SEQ ID NO: 54	moltype = AA	length = 1108
FEATURE	Location/Qualifiers	
REGION	1..1108	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1108	
	note = source = /note="nAPG09748"	
source	1..1108	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 54						
MAIRSIKLKL	KTRTGPEAQN	LRKGIVRTHR	LLNEGVAYYM	KMLLLFRQES	TGGQTKKELQ	60
EELVRIHIREQ	QOKNRADKNT	QALPLDKAFA	ALRQLYELLV	PSSIGQSGDA	QIISRKFLLSP	120
LVDPNSEGK	GTSKAGAKPT	WQKKKIEANDP	TWEQDYEKWK	KRREEDPTAS	VITLLEEYGI	180
RPIFPLYTNT	VADIawlPLQ	SNQFVRTWDR	DMLQQAAIERL	LWSWESWNKRV	QEEYSKLQEK	240
MTQLNEQLEG	GQEWISSLLEQ	YEEQREQELI	ENMTAANDKY	RITKRQMKW	NELYEQWSTV	300
LPNASHEQYR	EALKRVCQRL	RGRFCDAHF	QYLMKEEHHH	IWKGNPORIH	YFVARNELKK	360
RLEEAKQNAT	MTLPDARKH	LWVRFDARGG	NLQDYYLTAE	ADNPRSRFFV	TFSQLIWPN	420
SGWMEKQDVE	VELALSKQFY	QQTTLQKNDK	GKQKIEFKDK	GSGSTFSGHL	GGAKLQLERG	480
DLEKEEKDFE	GGEIGSVYLN	IIVIDFEPLQE	VKNGRQLQSPY	GQVLQLVRRP	NEFPKVTYK	540
SEELVEIWS	STKDSAGVES	LESGFGRVMSI	ALGLRTAAAT	SIFSVEESEND	ANAAGFSYWI	600
EGTPLVAVHK	RSYMLKLPGE	QVEKQVREKR	DERQDQQRV	RFQIRILSQV	IRMAKKQNRE	660
RADELDHLSQ	ALEKQKSLLD	QTDTFWNGI	VCDLTDALE	KEGGHQEQAVV	QIHRKAEHHV	720
GKVVQAWRK	PDADERKGIA	GLSMWSIEEL	DSLRKLL1SW	SRRTRNPREI	NCFEQGHSH	780
QLRLTHIQNV	KEDRLKQLSH	AIVMTALGYV	YDEKKLEWFA	KYPACQVILF	ENLSQYRSNM	840
DRSTKENSL	MKWAHRSIPIK	YVHMQAEPYQ	IQIGDVRAEY	SSRFHAKTGT	PGIRCKMVSG	900
HDLQGRRFEN	LQKRLISEQF	LTEEQVQLR	PGDIVPDDG	EWFMTLSDGS	EGKEVVFQ	960
DINAAQNLQK	RFWQRYNELF	KVSCRVLIRG	EEEYLIPKTK	SVQAKLGKGL	FVKKTDVMK	1020
DVYVWDSQAK	LKGKTTFTEE	SESPEQLED	QEIIIEEAAA	KGTYRTLFRD	PSGVFFPEFV	1080
WSTQKDFWSE	VKRRLYGKLR	ERFLMKTR				1108

SEQ ID NO: 55	moltype = AA	length = 1150
FEATURE	Location/Qualifiers	
REGION	1..1150	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1150	
	note = source = /note="nAPG09882"	
source	1..1150	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 55						
MYSIGLAGI	SSVGWSVIDE	RTGNVIDLGI	RLFSAKNSEK	NLERRTNRGG	RRLIRRKTNR	60
LKDAKKILAA	VGFYEDKSLK	NSCPYQLRVK	GLTEPLSKGE	IYKVTLHLIK	KRGISYLED	120
DTEAAKESQD	YKEQVRENAQ	LLTKYTPGQI	QLQLRKENN	VKTGINAQGN	YQLNVFKVSA	180
YANELATILK	TQQAFTYPNEL	TDDWIALFVQ	PGIAEEAGLI	YRKRPYHHG	GNEANNSPYG	240
RWSDFKFTGQ	PATNIFDKLI	GKDFQGELRK	SGLSLSAQOQ	NLLNDLTLNK	IDGEIPLSPE	300
QKEYIILAEML	TKEFTTRFGVN	DVVKLLGVKK	ERLSGWRLLDK	GKGKPEIHTLK	GYRNWRKIFA	360
ESGIDLATLP	TETIDCLAKV	LTLNTEREGI	ENTLAFELSE	LAESVKLLVL	DRYKELSOSV	420
STQAWHRSFL	KTLHLLIPEL	MNATSEQNTL	LEQFQLKSDV	RKRYSEYKKL	PTKDVLTIEY	480
NPFTVNKTVSQ	AFKVIDALLV	KYGEQIRYI	TIEMPRDDNE	EDEKKRIKEL	HAKNSQRKND	540
SQSIFYFMQKSG	WSQEKFQTTI	QKNRNRFLAKL	LYYYEQDGIC	AYTGLSISPE	LLVSDSTEID	600
HIIPISTISLD	DSINNKVLV	SKANQVKGQQ	TPYDAWMGDS	FFKINGKFSN	WDDYQKWVES	660
CHFSHKKENN	LLETRNIFDS	EQVEKFLARN	LNDTRYASRL	VLNTLQSF	NQETKVRVVN	720
GSPTHTLRKK	WGADLDKTRE	THHHHAVIDAT	LCAVTPFVKV	SRYHYAVKEE	TGEKVMREID	780
FETGEIVDEM	SYREFPKSKK	YERKTYQVKW	PNFREQQLKPV	NLHPRIKFSH	QVDRKANRKL	840
SDATIYSVRE	KTEVKTLKSG	KQKITTDETY	IGKIKDIYTV	DGWEAFKKQ	DKLLMKDLD	900
KTYERLLSIA	ETTPDFQVE	EKNGKVKRVK	RSPFAVYCEE	NDIPAIRKYA	KNNNGPLIRS	960
LKYYDGKLNK	HINITKDSQG	RPVEKTKNGR	KVTLQSLKPY	RYDIYQDLET	KAYYTQLY	1020
SDLRFVCEKY	GITEKEYMKK	VAEQTKGQV	RFCFSLQKND	GLEIEWKDSQ	CYDVRFYNFQ	1080
SANSINFKGL	EQEMMPAENQ	FKQKPYNNGA	INLNIAKYKG	EGKKLRKFNT	DILGKKHYLY	1140
YEKEPKNIK						1150

-continued

---

```

SEQ ID NO: 56      moltype = AA length = 1068
FEATURE          Location/Qualifiers
REGION           1..1068
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..1068
note = source = /note="nAPG03850"
source            1..1068
mol_type = protein
organism = synthetic construct
SEQUENCE: 56
MKYVLGLAIG IASCGWAVIN QEKHRIEDLG VRIFDKAENP KDGKSLATPR RDARSTRRTL 60
RRKKHMRMQR1 KILLVKHGLL SKTEIDHLYE SATEIDVWYL RLNALERRLN PKEFARVLIH 120
LAKRGRGFSN RKEETLSENG QILENISENQ QIMEQQNYRT VGEMLKDKK FENHKRKNKDG 180
TYIGTVTRQQ LKEEIQMIFN AQRLYKNDYA TEEFESSYLE IWASQRPYAS KDQIEKMIGY 240
CTLEPKEKRV PKASWSFQYF VALQTINNL RLINKDRIEL SFEEKNQIMN LALEKSIVKY 300
IDIRKLSSIP NEFHFNLLY SADTVDTAVE NKKCIEFKY HSINKLYKQI YGKSPVNLLP 360
IDYDTIACGL TIFKDDKDIL AYLQNQKVNA KGKPISNLAK KTYDDTFIQA LLTLNFSKMG 420
HLSFKALKNI IPFLEEGLSY DKACEKAGYN FKGTSHAEKT QKLPVIPQNT NPVVHRLSQ 480
TRKVINAIIK KYGSPSAIHI ETARELSKTF QERKEIDSMY QDNSKKNEHA IHKLKELGLI 540
NPSPGINIVKF KLWNEQDGRC MYSGKYIEPH RLFEETYEV DHLIPYSRSL DDSYNNKALT 600
LGIENQRKGN KTPYBYIGKT SIWHEFETRV QSNSKRINKKK QOKLLLQYFS YTREQEFIGR 660
NLNDTRYATI YLSSLIQQH1 IFSESSRKKK VHTVSGIITS HLRSRWGFNK DRKEGHIIHHA 720
LDAVIVAVTS DHMIQRVTKY YKLKELNRNL QAKRMPFPPE WEGFRLELEA RISPNTQQYL 780
KRILFKNYAD VNLSEIKPIF VSRRMPKRSIT GELHQETIRK LIGYNEKGKV LTAIKTKLED 840
IPFDANGDFP MYGKETDLYT YNAIKERYLHS HKKDKRKSFQ DPLYKPTKSG EIGPLIKSIK 900
IMDTRTIVNP VNQGKGVVYN SKIARTDVFK KDDKYYLIPI YTIDLKMN1 PQKAITAGKG 960
YEDWTTIDPS FTFLFSLFPN DLQIVPSKN KTIKARTTWS KKEVLLPSLT GFKGVHSGT 1020
AGITVETHDG SVIANVGSKQ LLLFEKYQVD VLGHYTKIKE EKRIGMVI 1068

SEQ ID NO: 57      moltype = AA length = 1081
FEATURE          Location/Qualifiers
REGION           1..1081
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..1081
note = source = /note="nAPG07553"
source            1..1081
mol_type = protein
organism = synthetic construct
SEQUENCE: 57
MQYVLGLAIG IASCGWAVIN QEKERIEDLG VRIFDKAENP KDGKSLAAPR RDARSTRRTL 60
RRKKHMRMQR1 KILLVKHGLL SQTEIDHLYE SATEVDWNL RLDATIERKLN PKEPTRVLIH 120
LAKRGRGFSN SKETTLSENG QILESISENQ QIMEQQNYKT IGEMILKDKK FENHKRKNKDG 180
TYIGTVTRQQ LQDEIQLIIFN AQRLYKNNYA TKEFESSYLE IWASQRPYAS KDQIEKMIGY 240
CTLEQKEKRV PRASWSFQYF VALHTINNL RISKDRIEL SFKEKKQIMN LALEKPIVKY 300
IDIRKLSSIP NELHFNSLLY SADTVDTVE NRKCIELKEY HSINKVYKQI YGKNALNLLP 360
IDYDTIAYGL TIFKDDKDIL EHLNKVVA KGKPINNLAK KTYDDTFIQA LLTLNFSKMG 420
HLSFKALKNI IPFLEEGLSY DKACEKAGYN FKGTSYTQQT QKLPVIPQNT NPVVHRLSQ 480
TKKVINAIIK KYGSPNAIHI ETARELSKTF QERKEIDSMY QDNSKKNEHA IHKLKELGNI 540
NPSPGINIVKF KLWNEQDGRC MYSGKYIEPH RLFEETYEV DHLIPYSRSL DDSYNNKLT 600
LGIENQRKGN KTPYEMGN1 SIWHEEIRV QSNKKINKKK QOKLLLQHFS YAREQEFIGR 660
NLNDTRYATI YLSSLIQQH1 IFSESSRKKK VHTVSGIITS HLRSRWGFNK DRKEGHIIHHA 720
LDAVIVAVTS DHMIQRVTKY YKLKELNRNL QAKRMPFPPE WEGFRLELEA RISPNTQQYL 780
KGLRFKNYAD VNLCEIKPIF VSRRMPKRSIT GELHQETIRK FIGYNEKGKV LTAIKTKLED 840
IPFDANGDFP MYGKETDLYT YNSIKERYLHS HKKDKRKSFQ EPLYKPTKSG GIGPLIKSIK 900
IMDTRTIVNP VNQGKGVVYN SKIARTDVFK KDDKYYLIPI YTIDLKMN1 PQKAITAGKG 960
YEDWTTIDHS FTFLFSLFPN DLKIVPSKN KEIKARSTSS KKEILLPSLI GFKGSVHSGT 1020
AGITVETHDG RFIANVGSKQ LLLFEKYQVD VLGHYTKIKE EKRIGMATCN DNKKSTAFGS 1080
L 1081

SEQ ID NO: 58      moltype = AA length = 1150
FEATURE          Location/Qualifiers
REGION           1..1150
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..1150
note = source = /note="nAPG05586"
source            1..1150
mol_type = protein
organism = synthetic construct
SEQUENCE: 58
MSIGLGLI SSVGWSVIDE ETGKIVDLGV RLPSAKNSEK NLERRRTSRA RRLIRRKTNR 60
LKDAKKLLEA IGFYEDKALK NVCPYQLRK GLTEGLTKGE LYKVVVLHIVK KRGISYLDDED 120
DAAEAKESQD YKEQVRKNAQ LLTKYTPGQI QLQLRKENNVR VKTGINGQGH YQLNVFKVSA 180
YADELATILK TQQALYPNEL TDDWIALFVQ PGIAENAGLI YRKRPYHGP GNEANNSPYG 240

```

---

-continued

---

RWSDFQKTCQ	PAANIFDKLI	GKDFQGELRA	SGLSLSAQQY	NLLNDLTLNK	IDGEVSLSP	300
QEKFILTELM	TKEFARFGVN	DIAKLLGVKK	EQLSGWRLDK	KKGPEIHTLK	GYRNWRKIFA	360
EAGIDLATLP	TETIDCLAKV	LTLNTEREGV	ENTLAFELPE	LAEPVKSLVL	DHYKELSQSI	420
STQAWHRFLS	KTLHLLIPEL	IKSTSEQNTL	LEQFQLKAGV	RKRYSDYKKL	PTKEVLAEIY	480
NPTVNVKTSQ	AFKVMDALLE	KYGKDQIHYI	TVEMPRDNE	EEERKRIKEL	QTKNQRKND	540
SQQYFLQKSG	WSQEKFQATTI	HKNRFLAKL	LYFFEQDGCV	AYTGNPISPL	LLVSDSTEID	600
HIIPISISLD	DSINNKVLVL	SHANQVKGQQ	TPYDARMAGA	FNKINGKFSN	WDEYQKWVES	660
RPFPSRKVNN	LLETTRNIFDS	EVQVKFLSRN	LNDTRYASRL	VNLNTQSFFE	NQDTIVRVVN	720
GSFTHTLRKK	WGADLDKTRE	THHHHAVDAT	LCAVTPFKV	SRHYHYAVNEE	TGEKFMREID	780
VEFGEILDEI	PYREYKAKK	YERKTYQVKG	SNPREQLKPI	TIHPKIKFSH	QVDRKANRKL	840
SDATIYVSRE	KTEVKTLKSG	KEKITTDIYT	IGKIKDITYV	DGWEAPKKQ	DKLIMKEFDE	900
KTYELLVTIA	ATTPDFQVE	EKNGKVKVRK	RSPFAVYCEE	NGIPAIRKYA	KKNNGPVIIRS	960
LKYYDGKLNK	HINITKDEKG	RPVEQTKNGR	KVTLQSLKPY	RYDIYQDLET	KAYYTQVLYY	1020
SDLRFVGESEY	GITEKEYMKK	VAEQTKGQVV	RCFCFSLQKND	GLEIEWKDSQ	RYDVRFYNFQ	1080
SANSINFKGL	EQEMIPIAENQ	FQKQPKYNNGA	INLNIAKYKG	EGKKLRKFNT	DILGKKHLS	1140
YEKEPKNIK						1150

SEQ ID NO: 59	moltype = AA	length = 1052
FEATURE	Location/Qualifiers	
REGION	1..1052	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1052	
	note = source = /note="nAPG01604"	
source	1..1052	
	mol_type = protein	
	organism = synthetic construct	

SEQUENCE: 59						
MVTKYILGLA	IGITSVGYGI	INYEDKTIID	AGVRLFPEAN	VENNEGRRSK	RGARRLKRR	60
IHLRDLRICKQ	LSEYNLVDLQ	NIPQSPSPYE	IRVKGLREEL	SKDELVIALL	HIAKRRGIHN	120
VEAVDETQDE	GNELSTKEQL	AKNNNNPLDK	YVCELLLERL	KDGKVGEKNN	RFKTTDIKE	180
VKQLLLETOKE	AHQLDDDFIN	RYIDLIELTR	EYFEGPGKGS	PFGWGGDLKK	WYETLMGHCT	240
YFPNELRSVK	YSYSADLFNA	LNDLNNLVIQ	REGNSKLEYH	EKYHIIENVF	KQKKKPTLKQ	300
IAANEIGVSPD	DIKGFRINKS	GKETTTEFQL	YHDLKKVLLID	QSILEVNQLL	DQIAIBILTY	360
QDKESIINEL	NQLEINNEQ	DKESENLSLG	YNGTHRSLK	CINLVIEELW	HTSRNQMEIF	420
TYLNIKPKKI	DLAKTNKIPK	NMIDEFLSP	VVKRTFGQAI	VNINKVIEKY	GVPEDIIEI	480
ARESNNSKDKQ	KFINSLOKKN	ETTRKRINEI	IGQYGNQNAK	RLVEKIRLHD	EQEGKCLYS	540
ESIPLLEDLIN	NPQYYEVDHI	IPIRSVSDNS	YQNKVLVKQ	ENSKKSNRTP	YQYFNSGETK	600
LSYNQFKQHV	SLNLSKSKDRI	SKKKKEYLE	ERDINKYEVQ	KEFINRNLDV	TRYATRELTN	660
YLKAYFSAND	MDVKVKTING	SFTDYLRKW	KFKKERNHG	KHHAEDALII	ANADFLFKEN	720
KKLKKKANAIL	EQPSLDNGKS	DATVENDNEY	WTFSIPKQV	NDIKEFRDFK	FSHRVDKKPN	780
RQLINDTLYS	TRKIEHNTFI	VSPITNIYSK	DNDELKKKFN	KNPEKFMLYQ	HDPKTFEKE	840
VIMKQYANEK	NPLAKYHEET	GEYLTKYSKK	NNGPIVKTIK	VLGDKVGKHL	DVTHKYKYSN	900
SKIVKKTINP	YRFDVYLTDK	GYKFITISQY	DVLKKDNYYY	ILKEKYEELK	IKKSISDTDQ	960
FIGSFYVNDL	IKINDQIFKV	VGVNNDLLNR	IELLDLDISY	KEYCKINNIK	TNRRIKGITK	1020
KITNIEKFST	DVLGNLYKAH	SNHPQLIFQ	RD			1052

SEQ ID NO: 60	moltype = AA	length = 1072
FEATURE	Location/Qualifiers	
REGION	1..1072	
	note = source = /note="Bacillus sp. APG08290.1"	
source	1..1072	
	mol_type = protein	
	organism = Bacillus sp.	

SEQUENCE: 60						
MSELDYRIGL	DIGTNSIGWG	VIELFWNKDR	ERYEKVRIVD	KGVRMFDKAE	IPNKGASLAE	60
PRIARISSRR	RLNRKSQRKK	EIRNLNVQHG	MITQEELDLL	YPLSKKSIDI	WDIRLDGLDR	120
LLNHLEWARL	LIHLAQRRGF	KSNRKSELKD	AETGKVLSI	QVNEKRLFLY	RTVGEWIKD	180
AERFSKYDRRR	NSPNEYVFSV	SRADALEKEIV	TLFEAQRKFQ	SSYASKNLQE	TYLQIWAHQL	240
PFASGNAILN	LKVGYCSLLKG	KERRIPKATY	TFQYFSALDQ	VNRTTRLGPDF	QPFTQECKE	300
IILDKMFQRTD	YKKKKTIP	SYDIRKWL	LDETIQFKGL	NYDPNEELKK	IEKKFIFINLK	360
AFYEIKKVV	NYAERTNEAF	STLDYDAIAY	ALTVYKTDK	IRSYLKKSN	LSKRCYDDQL	420
IEELFTLSYT	KFGHLSFKAI	NHVLPIMQEG	RTYQEAIHQ	GYDTTNLKKE	NRSMLPLIP	480
DEITNPIVKR	AITQARKVVA	AIIRRYGSPN	SVHIELAREL	SKSHDERKK	MTAHDENYKK	540
NKGAISLNE	NGILNPTGYD	IVRYKQHGE	GERCAYSKE	IPPDTFFNEL	KKERNGSPIL	600
EVDHILPYSQ	SFIDSYHNV	LVYSDENRNK	GNRIPYTYFL	ETNKDWAEFE	RYVRSNKLFS	660
KKKREYLLKK	TYLPRSELEI	KERHLNDTRY	ASTFLKNFIE	QNLQFKEVEV	NLRKKRVQTV	720
NGVITAHLRK	RWGLEKNRQE	TYLHHAMDAI	IVACTDHMV	TRITEYYQIK	ESNKSVKPY	780
FPPMPWEGFRD	ELLSHLASQP	IAKKISEELK	AGYQSSDYIF	VSRMPKRSVT	GAAHQOTIRR	840
KGGIDKKGKT	IIKRVRLKD	IKFDENGDFK	MVGKEQDLAT	YEAIKQRYLE	HRKNSKKAFE	900
TPLYKPSKKG	TGNLIKRVK	EGQTKAFVRE	VNGGVAQNSD	LVRVDFEKD	DKYYMVPIVY	960
PDTVCSELPK	KVVKSGKGYE	QWLTLDNSFT	FKSSLYPYD	VRLVKGNEDR	FLYFGTLID	1020
SDRLNFKDVN	KPSKQNEYRY	SLKTIENLEK	YEVGVLGDLR	LVKQETRRIF	NR	1072

SEQ ID NO: 61	moltype = AA	length = 1072
FEATURE	Location/Qualifiers	

-continued

---

```

REGION          1..1072
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polypeptide"
REGION          1..1072
                note = source = /note="nAPG08290.1"
source           1..1072
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 61
MSLEDYRIGL AIGTN SIGWG VIELFWN KDR ERYEK VRIVD KGVRM FD KAE IPNKG ASLAE 60
P RRIARSS RR LRLRK SQRK EIRNL LVQHG M I QEE LD L YPLSKKSIDI WDIRLDGLDR 120
L L NHLEW AR L IHLAQ ORRGK KSNRK SEELK AETGKV LSSI QVNEK RLFLY RTV GEMWIKD 180
A EFSKY DRRR N SPNEY VFSV SRAD LEK EIV TL FEA QRKFQ SSYASK NLQE TYLQI WAHQL 240
P FAS GNA ILN KVGYC SLLK KERR IWK L TFOYF SALDQ VNRTRL GPDF QPF TOE QKEI 300
I LD KMF QRTD YKKK T IPEV SYY DIRK WLE LDETIQ FKGL NYDPNE ELKK IEKKP FINLK 360
A FYE I KK VVA MYAERT NEAF STLD YD AIA Y ALTVY KTD K IRSYL KKSNN LSKRC YDD QL 420
I EELFT LS YT KFGHLS FKAI NHVL PIM QEG RTY QEA IHQ L GYDTT NLK KE NRS MFL PLIP 480
D EITNP IVK R AITQARK VVN AII RRYG SPN SVHIELAREL SKSHDER KKI MTAHDEN YKK 540
N KGAISIL E NGILN P TGYD IVRY K LWK EQ GERCAYSL KE IPPDTFF NEL KKERNG SPIL 600
EVDHIL PYSQ SFIDS YHN KV LVY SDEN RNK GN RIP YTF L ETNKD WE AFE RYVRSN KLF S 660
K KKREY LLKK TYLP RESE LI KERH LND TRY ASTFL KNFIE QNLQF KEVEV NLRKK R VQT V 720
N GVITA HLRK RW GLEK NR QE TYLHHAM DAI I VACTDH HMV TRITE YYQIK ESNK SVK PY 780
FPMPW EGFR D ELLS HLAS QP IAKK ISEELK AGYQSS DYIF VSRMP KRS VT GA AHQ DTI RR 840
KG GI DKK GKT II KRV RLKD IKFDENG DPK MVG KEQDL AT YEAI KQ RYLE HRKNS KKA FE 900
T PLY KPS KKG TG NLI KRV KI EG QT KAF VRE VNGG VQA QNS D LVR VDL F EKD DK YYM VP IY 960
PD TVC SEL PK KVVK SGK GYE QWL TLD NS PT FK S L YP DYL VRL VKG NED R FLY FG TL DID 1020
SD RL NF KDV N KPS KQNEY RY SLKT IEN LEK YEVGV LGDL R LVK QET R RIF NR 1072

SEQ ID NO: 62      moltype = DNA length = 65
FEATURE          Location/Qualifiers
misc_feature     1..65
                note = source = /note="Description of Artificial Sequence:
                                Synthetic coligonucleotide"
misc_feature     1..65
                note = source = /note="CF E60X nAPG06646 Target 1"
source            1..65
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 62
aatgagttta ggat ttttct ttgaaggccag ctatctatcc cattctctgc aaaagaataa 60
aaagt                      65

SEQ ID NO: 63      moltype = DNA length = 65
FEATURE          Location/Qualifiers
misc_feature     1..65
                note = source = /note="Description of Artificial Sequence:
                                Synthetic coligonucleotide"
misc_feature     1..65
                note = source = /note="CF E60X nAPG06646 Target 2"
source            1..65
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 63
attaatg agttt gat ttttca agc cagctatcta tcccattctc tgcaaaagaa 60
aaaaa                      65

SEQ ID NO: 64      moltype = DNA length = 65
FEATURE          Location/Qualifiers
misc_feature     1..65
                note = source = /note="Description of Artificial Sequence:
                                Synthetic coligonucleotide"
misc_feature     1..65
                note = source = /note="CF E60X nAPG06646 Target 3"
source            1..65
                mol_type = other DNA
                organism = synthetic construct
SEQUENCE: 64
gcattaaatg a gttttaggatt tttctt gaa gccagctatc tatccccatc tctgcaaaag 60
aaaaa                      65

SEQ ID NO: 65      moltype = DNA length = 65
FEATURE          Location/Qualifiers
misc_feature     1..65
                note = source = /note="Description of Artificial Sequence:
                                Synthetic coligonucleotide"
misc_feature     1..65

```

---

-continued

---

```

source          note = source = /note="CF E60X nAPG06646 Target 4"
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 65
aaggcatta atgagtttag gatTTTCTT tgaagccagc tatctatccc attctctgca 60
aaaga                                         65

SEQ ID NO: 66      moltype = DNA length = 65
FEATURE
misc_feature       Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature       1..65
note = source = /note="CF E60X nAPG06646 Target 5"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 66
gaagggcatt aatgagttta ggattttctt ttgaagccag ctatctatcc cattctctgc 60
aaaag                                         65

SEQ ID NO: 67      moltype = DNA length = 65
FEATURE
misc_feature       Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature       1..65
note = source = /note="CF E60X nAPG06646 Target 6"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 67
cgaagggcat taatgagttt aggatttttc tttgaagcca gctatctatc ccattctctg 60
caaaa                                         65

SEQ ID NO: 68      moltype = DNA length = 65
FEATURE
misc_feature       Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature       1..65
note = source = /note="CF E60X nAPG09882 Target 1"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 68
gagttagga tttttttttt aagccagcta tctatccat tctctgaaaa agaataaaaa 60
gtggg                                         65

SEQ ID NO: 69      moltype = DNA length = 65
FEATURE
misc_feature       Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature       1..65
note = source = /note="CF E60X nAPG09882 Target 2"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 69
tgagttagg atttttttttt gaaggcagct atctatccca ttctctgcaa aagaataaaa 60
agtgg                                         65

SEQ ID NO: 70      moltype = DNA length = 65
FEATURE
misc_feature       Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature       1..65
note = source = /note="CF E60X nAPG09882 Target 3"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 70
atgagtttag gatTTTCTT tgaagccago tatctatccc attctctgca aaagaataaa 60

```

---

-continued

---

```

aagtg                                         65

SEQ ID NO: 71      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF E60X nAPG09882 Target 4"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 71
agggcattaa tgagtttagg attttctt gaagccagct atctatccca ttctctgcaa  60
aagaa                                         65

SEQ ID NO: 72      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF E60X nAPG00969 Target 1"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 72
gtttaggatt tttcttgaa gccagctatc tatcccattc tctgcaaaa aataaaaagt  60
ggcac                                         65

SEQ ID NO: 73      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF E60X nAPG00969 Target 2"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 73
agtttaggat tttcttgaa agccagctat ctagccattt ctctgcaaaa gaataaaaag  60
tggga                                         65

SEQ ID NO: 74      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF E60X nAPG03850 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 74
ggattttct ttgaagccag ctagctatcc cattctctgc aaaagaataa aaagtggac  60

SEQ ID NO: 75      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF E60X nAPG03850 Target 2"
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 75
agtttaggat tttcttgaa agccagctat ctagccattt ctctgcaaaa gaataaaaag  60

SEQ ID NO: 76      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65

```

---

---

-continued

---

```

source          note = source = /note="CF E60X nAPG07433.1 Target 1"
               1..65
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 76
gaaggccatt aatgagttta ggattttctt tttgaagccag ctatctatcc cattctcgc  60
aaaag                                     65

SEQ ID NO: 77      moltype = DNA  length = 60
FEATURE          Location/Qualifiers
misc_feature     1..60
note = source = /note="Description of Artificial Sequence:
                  Synthetic oligonucleotide"
misc_feature     1..60
note = source = /note="CF E60X nAPG09748 Target 1"
source          1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 77
gtccccacttt ttattttttt gcagagaatg ggatagatag ctggcttcaa agaaaaatcc  60

SEQ ID NO: 78      moltype = DNA  length = 60
FEATURE          Location/Qualifiers
misc_feature     1..60
note = source = /note="Description of Artificial Sequence:
                  Synthetic oligonucleotide"
misc_feature     1..60
note = source = /note="CF E60X nAPG07553 Target 1"
source          1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 78
agttaggat ttttttttga agccagctat ctatcccatt ctctgaaaaa gaataaaaaag  60

SEQ ID NO: 79      moltype = DNA  length = 60
FEATURE          Location/Qualifiers
misc_feature     1..60
note = source = /note="Description of Artificial Sequence:
                  Synthetic oligonucleotide"
misc_feature     1..60
note = source = /note="CF E60X nAPG05586 Target 1"
source          1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 79
tttaggattt ttcttttgaag ccagctatct atcccaattctt ctgaaaaaaga ataaaaaagt  60

SEQ ID NO: 80      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = source = /note="Description of Artificial Sequence:
                  Synthetic oligonucleotide"
misc_feature     1..25
note = source = /note="CF E60X nAPG06646 Target 1"
source          1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 80
tttgaagccag ctatctatcc cattc                                         25

SEQ ID NO: 81      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25
note = source = /note="Description of Artificial Sequence:
                  Synthetic oligonucleotide"
misc_feature     1..25
note = source = /note="CF E60X nAPG06646 Target 2"
source          1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 81
tcttttgaagc cagctatcta tccca                                         25

SEQ ID NO: 82      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature     1..25

```

---

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG06646 Target 3"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 82
tttcttgaa gccagctatc tatcc                                25

SEQ ID NO: 83      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG06646 Target 4"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 83
gatttttctt tgaaggccagc tatct                                25

SEQ ID NO: 84      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG06646 Target 5"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 84
ggatttttctt ttgaaggccag ctata                                25

SEQ ID NO: 85      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG06646 Target 6"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 85
aggatttttc ttgttaaggcca gctat                                25

SEQ ID NO: 86      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG09882 Target 1"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 86
aagccagcta tcttatccccat tctct                                25

SEQ ID NO: 87      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF E60X nAPG09882 Target 2"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 87
gaagccagct atctatccccat ttctc                                25

SEQ ID NO: 88      moltype = DNA length = 25

```

---

-continued

---

```

FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF E60X nAPG09882 Target 3"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 88
tgaaggcagc tatctatccc attct          25

SEQ ID NO: 89                         moltype = DNA length = 25
FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF E60X nAPG09882 Target 4"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 89
attttttttt gaagccagct atctt          25

SEQ ID NO: 90                         moltype = DNA length = 25
FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF E60X nAPG00969 Target 1"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 90
gccagctatc tatccccatt tctgc          25

SEQ ID NO: 91                         moltype = DNA length = 25
FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF E60X nAPG00969 Target 2"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 91
agccagctat ctatcccatt ctctg          25

SEQ ID NO: 92                         moltype = DNA length = 20
FEATURE                                Location/Qualifiers
misc_feature                          1..20
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..20
                                         note = source = /note="CF E60X nAPG03850 Target 1"
source                               1..20
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 92
ctatctatcc cattctctgc                20

SEQ ID NO: 93                         moltype = DNA length = 20
FEATURE                                Location/Qualifiers
misc_feature                          1..20
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..20
                                         note = source = /note="CF E60X nAPG03850 Target 2"
source                               1..20
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 93
agccagctat ctatcccatt                20

```

---

-continued

---

```

SEQ ID NO: 94      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF E60X nAPG07433.1 Target 1"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 94
ggattttct ttgaagccag ctatc                                     25

SEQ ID NO: 95      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF E60X nAPG09748 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 95
gcagagaatg ggatagatag                                         20

SEQ ID NO: 96      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF E60X nAPG07553 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 96
agccagctat ctatcccatt                                         20

SEQ ID NO: 97      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF E60X nAPG05586 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 97
ccagctatct atccccattct                                         20

SEQ ID NO: 98      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..164
note = source = /note="CF E60X nAPG06646 Target 1 sgRNA"
1..164
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 98
ttgaagccag ctatctatcc cattcgccat aattcctctg taaaacttaa agaagggtta  60
tagagtttatt atggtaaggc aatatgccgt ggcgttgggg atcgctatg tccggttta 120
ccggatctcc ctaaaaggta ctaacttgg ttatgcacct tttt                164

SEQ ID NO: 99      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..164
note = source = /note="CF E60X nAPG06646 Target 2 sgRNA"
1..164

```

---

-continued

---

```

mol_type = other RNA
organism = synthetic construct

SEQUENCE: 99
tctttgaagc cagctatcta tcccagccat aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 100      moltype = RNA  length = 164
FEATURE           Location/Qualifiers
misc_feature      1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..164
note = source = /note="CF E60X nAPG06646 Target 3 sgRNA"
source            1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 100
tttcttgaa gccagctatc tatccgcat aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 101      moltype = RNA  length = 164
FEATURE           Location/Qualifiers
misc_feature      1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..164
note = source = /note="CF E60X nAPG06646 Target 4 sgRNA"
source            1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 101
gatttttctt tgaaggccagc tatctgccat aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 102      moltype = RNA  length = 164
FEATURE           Location/Qualifiers
misc_feature      1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..164
note = source = /note="CF E60X nAPG06646 Target 5 sgRNA"
source            1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 102
ggatttttctt ttgaaggccag ctatgccat aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 103      moltype = RNA  length = 164
FEATURE           Location/Qualifiers
misc_feature      1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..164
note = source = /note="CF E60X nAPG06646 Target 6 sgRNA"
source            1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 103
aggatttttc ttgaaggcca gctatgccat aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 104      moltype = RNA  length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..118
note = source = /note="CF E60X nAPG09882 Target 1 sgRNA"
source            1..118
mol_type = other RNA

```

---

---

-continued

---

```

SEQUENCE: 104          organism = synthetic construct
aagccagcta tctatccccat tctctgtttt tgtactctca ataaaaagg 60
tacaaaata aggcattttg ccgaatttac cggcctacat atgttagggcg gttttttt    118

SEQ ID NO: 105          moltype = RNA  length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF E60X nAPG09882 Target 2 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 105          moltype = RNA  length = 118
gaagccagct atctatccca ttctctgtttt tgtactctca ataaaaagg 60
tacaaaata aggcattttg ccgaatttac cggcctacat atgttagggcg gttttttt    118

SEQ ID NO: 106          moltype = RNA  length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF E60X nAPG09882 Target 3 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 106          moltype = RNA  length = 118
tgaagccagc tatcttatccc attctgtttt tgtactctca ataaaaagg 60
tacaaaata aggcattttg ccgaatttac cggcctacat atgttagggcg gttttttt    118

SEQ ID NO: 107          moltype = RNA  length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF E60X nAPG09882 Target 4 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 107          moltype = RNA  length = 118
atttttcctt gaagccagct atctatgtttt tgtactctca ataaaaagg 60
tacaaaata aggcattttg ccgaatttac cggcctacat atgttagggcg gttttttt    118

SEQ ID NO: 108          moltype = RNA  length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF E60X nAPG00969 Target 1 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 108          moltype = RNA  length = 118
gccagctatc tatcccatc tctcggtttt agtactctgt gaaaggcacag aatctactaa 60
aataaggcat aatgccgtat ttaatccccat cataattctg atgggatttt ttatattt    118

SEQ ID NO: 109          moltype = RNA  length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF E60X nAPG00969 Target 2 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 109          moltype = RNA  length = 118
agccagctat ctatcccatc ctctcggtttt agtactctgt gaaaggcacag aatctactaa 60
aataaggcat aatgccgtat ttaatccccat cataattctg atgggatttt ttatattt    118

SEQ ID NO: 110          moltype = RNA  length = 163

```

---

-continued

---

```

FEATURE          Location/Qualifiers
misc_feature    1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature    1..163
note = source = /note="CF E60X nAPG03850 Target 1 sgRNA"
source          1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 110
ctatctatcc cattatctgc gctatagttc cataagaaaa aagtttctta agttactata 60
gttaaggccaa tgaccctgtgg cgtttggga tcgccttata ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcctttat ttt                163

SEQ ID NO: 111      moltype = RNA length = 163
FEATURE          Location/Qualifiers
misc_feature    1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature    1..163
note = source = /note="CF E60X nAPG03850 Target 2 sgRNA"
source          1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 111
agccaggtat ctatcccatt gctatagttc cataagaaaa aagtttctta agttactata 60
gttaaggccaa tgaccctgtgg cgtttggga tcgccttata ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcctttat ttt                163

SEQ ID NO: 112      moltype = RNA length = 135
FEATURE          Location/Qualifiers
misc_feature    1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature    1..135
note = source = /note="CF E60X nAPG07433.1 Target 1 sgRNA"
source          1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 112
ggatttttct ttgaagccag ctatcgcat agttcattta aagccaaaag tggctttgat 60
gtttctatga taagggttcc gaccctgtgc gtcggggatc gcctgccccat tgaaatggc 120
ttctcccat ttatt                                135

SEQ ID NO: 113      moltype = RNA length = 135
FEATURE          Location/Qualifiers
misc_feature    1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature    1..135
note = source = /note="CF E60X nAPG09748Target 1 sgRNA"
source          1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 113
cgacgggttag aggccgtatg tcgatttgtt ttaatttcgtt gggtgtgcat tggcgcttc 60
cattacaggg cggcttaccac gaatagccac gaagttaaaag cttcgtggct agcacgcaga 120
gaatgggata gatag                                135

SEQ ID NO: 114      moltype = RNA length = 151
FEATURE          Location/Qualifiers
misc_feature    1..151
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature    1..151
note = source = /note="CF E60X nAPG07553 Target 1 sgRNA"
source          1..151
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 114
agccaggtat ctatcccatt gctatagttc cataagaaa cttaaatgttac tataatgtagg 60
gcaatgaccc gtggcgtttg gggatcgctt catccattac ggatattctc cccatgtgaa 120
aaggcacctaa gcataaggctt aagggtgcctt t                                151

SEQ ID NO: 115      moltype = RNA length = 110
FEATURE          Location/Qualifiers

```

---

---

-continued

---

```

misc_feature          1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..110
note = source = /note="CF E60X nAPG05586 Target 1 sgRNA"
source                1..110
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 115
cgagcttatct atccattatct gttattgtac tctcaataaaa aagtatttga gaatctacaa  60
taataaggca ttggccgaa ttaccgcct tacatatgtt gggcggttt           110

SEQ ID NO: 116      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF G542X nAPG06646 Target 1"
source                1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 116
cgttgacctc cactcagtgt gattccacct tctcaaagaa ctatattgtc tttctctgca  60
aactt                           65

SEQ ID NO: 117      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF G542X nAPG06646 Target 2"
source                1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 117
gacctccact cagtgtgatt ccaccccttc aaagaactat attgttttc tctgcaaact  60
tggag                           65

SEQ ID NO: 118      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF G542X nAPG06646 Target 3"
source                1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 118
cctccactca gtgtgattcc accttctcaa agaactataat tgtctttctc tgcaaacttg  60
gagat                           65

SEQ ID NO: 119      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF G542X nAPG06646 Target 4"
source                1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 119
caactcagtg tgattccacc ttctcaaaga actatattgt ctttctctgc aaacttggag  60
atgtc                           65

SEQ ID NO: 120      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF G542X nAPG09882 Target 1"
source                1..65

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 120
tcttgctcggtgacacctccac tcagttgtat tccacccctt caaagaacta tattgtcttt 60
ctctg                                         65

SEQ ID NO: 121      moltype = DNA length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF G542X nAPG09882 Target 2"
source
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 121
ttgctcggttgcacctccactc agtgtgattt cacccttctca aagaactata ttgtctttct 60
ctgca                                         65

SEQ ID NO: 122      moltype = DNA length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF G542X nAPG09882 Target 3"
source
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 122
cactcagtgt gattccacct tctcaaagaa ctatattgtc tttctctgca aactggaga 60
tgtcc                                         65

SEQ ID NO: 123      moltype = DNA length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF G542X nAPG03850 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 123
tgacacctccac tcagttgtat tccacccctt caaagaacta tattgtcttt ctctgcaaac 60

SEQ ID NO: 124      moltype = DNA length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF G542X nAPG03850 Target 2"
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 124
tcagttgtat tccacccctt caaagaacta tattgtcttt ctctgcaaac ttggagatgt 60

SEQ ID NO: 125      moltype = DNA length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF G542X nAPG09748 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 125
agagaaaagac aatatagttc tttgagaagg tggaatcaca ctgagtggag gtcaacggac 60

SEQ ID NO: 126      moltype = DNA length = 60
FEATURE
misc_feature
1..60

```

---

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF G542X nAPG07553 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 126
tcagtgtat tccaccttct caaagaacta tattgtcttt ctctgcaaac ttggagatgt 60

SEQ ID NO: 127      moltype = DNA length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF G542X nAPG05586 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 127
cgttgacctc cactcagtgt gattccacct tctcaaagaa cttatattgtc tttctctgca 60

SEQ ID NO: 128      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF G542X nAPG06646 Target 1"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 128
gattccacct tctcaaagaa ctata 25

SEQ ID NO: 129      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF G542X nAPG06646 Target 2"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 129
cacacccatc aaagaactat attgt 25

SEQ ID NO: 130      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF G542X nAPG06646 Target 3"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 130
accttctcaa agaactatata tgtct 25

SEQ ID NO: 131      moltype = DNA length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF G542X nAPG06646 Target 4"
source
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 131
ttctcaaaga actatattgt ctttc 25

SEQ ID NO: 132      moltype = DNA length = 25

```

---

-continued

---

```

FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF G542X nAPG09882 Target 1"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 132
tcagtgtat tccacaccttct caaag          25

SEQ ID NO: 133                      moltype = DNA length = 25
FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF G542X nAPG09882 Target 2"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 133
agtgtgattc caccttctca aagaa          25

SEQ ID NO: 134                      moltype = DNA length = 25
FEATURE                                Location/Qualifiers
misc_feature                          1..25
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..25
                                         note = source = /note="CF G542X nAPG09882 Target 3"
source                               1..25
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 134
ttctcaaagaa ctatattgtc tttct          25

SEQ ID NO: 135                      moltype = DNA length = 20
FEATURE                                Location/Qualifiers
misc_feature                          1..20
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..20
                                         note = source = /note="CF G542X nAPG03850 Target 1"
source                               1..20
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 135
tccacaccttct caaagaacta              20

SEQ ID NO: 136                      moltype = DNA length = 20
FEATURE                                Location/Qualifiers
misc_feature                          1..20
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..20
                                         note = source = /note="CF G542X nAPG03850 Target 2"
source                               1..20
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 136
caaagaacta tattgtcttt                20

SEQ ID NO: 137                      moltype = DNA length = 20
FEATURE                                Location/Qualifiers
misc_feature                          1..20
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic oligonucleotide"
misc_feature                          1..20
                                         note = source = /note="CF G542X nAPG09748 Target 1"
source                               1..20
                                         mol_type = other DNA
                                         organism = synthetic construct
SEQUENCE: 137
tttgagaagg tggaatcaca                20

```

---

---

-continued

---

```

SEQ ID NO: 138      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF G542X nAPG07553 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 138
caaagaacta tattgtcttt                                         20

SEQ ID NO: 139      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF G542X nAPG05586 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 139
gattccacca tctcaaagaa                                         20

SEQ ID NO: 140      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..164
note = source = /note="CF G542X nAPG06646 Target 1 sgRNA"
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 140
gattccacca tctcaaagaa ctatagccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgcctg ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaaggtaa ctaactttgg tttagtcacct tttt                164

SEQ ID NO: 141      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..164
note = source = /note="CF G542X nAPG06646 Target 2 sgRNA"
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 141
ccaccccttc aaagaactat attgtgccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgcctg ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaaggtaa ctaactttgg tttagtcacct tttt                164

SEQ ID NO: 142      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..164
note = source = /note="CF G542X nAPG06646 Target 3 sgRNA"
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 142
acccctctaa agaactatata ttgtgccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgcctg ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaaggtaa ctaactttgg tttagtcacct tttt                164

SEQ ID NO: 143      moltype = RNA   length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:

```

---

-continued

---

```

misc_feature      Synthetic polynucleotide"
1..164
source          note = source = /note="CF G542X nAPG06646 Target 4 sgRNA"
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 143
tttcaaaaga actatattgt ctttcgcatt aattcctctg taaaacttaa agaagggtta 60
tagagttatt atggtaaggc aatatgccgt ggcggtgggg atccgcctatg tccgggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 144      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
source          note = source = /note="CF G542X nAPG09882 Target 1 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 144
tcagtgat tccaccccttct caaagggttt tgtaactctca ataaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cgccctacat atgttagggcg gttttttt 118

SEQ ID NO: 145      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
source          note = source = /note="CF G542X nAPG09882 Target 2 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 145
agtgtgattc caccttctca aagaagttt tgtaactctca ataaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cgccctacat atgttagggcg gttttttt 118

SEQ ID NO: 146      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
source          note = source = /note="CF G542X nAPG09882 Target 3 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 146
tctcaaagaa ctatattgtc tttctgtttt tgtaactctca ataaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cgccctacat atgttagggcg gttttttt 118

SEQ ID NO: 147      moltype = RNA length = 163
FEATURE           Location/Qualifiers
misc_feature      1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..163
source          note = source = /note="CF G542X nAPG03850 Target 1 sgRNA"
1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 147
tccaccccttct caaagaacta gctatagttc cataagaaaa aagtttctta agttactata 60
gttaaaaaaa tgacccgtgg cggtggggat tcgccttata ctggatgga tattctcccc 120
atgtgaaaag cacctaagca tagcgctatg gtgttttat ttt 163

SEQ ID NO: 148      moltype = RNA length = 163
FEATURE           Location/Qualifiers
misc_feature      1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..163
source          note = source = /note="CF G542X nAPG03850 Target 2 sgRNA"
1..163

```

---

---

-continued

---

```

mol_type = other RNA
organism = synthetic construct

SEQUENCE: 148
caaaagaacta tattgtcttt gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaggccaa tgaccctgtgg cgtttgggaa tcgccttatac ctggtatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt 163

SEQ ID NO: 149      moltype = RNA length = 135
FEATURE             Location/Qualifiers
misc_feature        1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature        1..135
note = source = /note="CF G542X nAPG09748 Target 1 sgRNA"
source              1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 149
cgacggtagttag aggccgtatg tcgatttgc ttaatttcgt gcggtgtcat tgcgtcctc 60
cattacagg cggctaccac gaatagccac gaagtaaaag cttcgtggct agcactttga 120
gaaggtggaa tcaca 135

SEQ ID NO: 150      moltype = RNA length = 151
FEATURE             Location/Qualifiers
misc_feature        1..151
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature        1..151
note = source = /note="CF G542X nAPG07553 Target 1 sgRNA"
source              1..151
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 150
caaaagaacta tattgtcttt gctatagttc cataagaaag cttaagttac tatagtaagg 60
gcaatgacc gcggcggttg gggatcgctt catccattac ggatattctc cccatgtgaa 120
aaggcaccaa gcataaggct aagggtgttt t 151

SEQ ID NO: 151      moltype = RNA length = 110
FEATURE             Location/Qualifiers
misc_feature        1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature        1..110
note = source = /note="CF G542X nAPG05586 Target 1 sgRNA"
source              1..110
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 151
gattccacact tctcaaagaa gttattgtac tctcaataaa aagttattga gaatctacaa 60
taataaggca tcttgcgaa ttaccgccc tacatatgtt gggcggttt 110

SEQ ID NO: 152      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..65
note = source = /note="CF Q493X nAPG09882 Target 1"
source              1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 152
gatattttct ttaatgggtgc caggcataat ccagggaaac taagaacaga atgaaattct 60
tccac 65

SEQ ID NO: 153      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..65
note = source = /note="CF Q493X nAPG09882 Target 2"
source              1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 153

```

---

-continued

---

```

atattttctt taatggtgcc aggataatac cagaaaaact aagaacagaa tgaaattctt 60
ccact                                         65

SEQ ID NO: 154      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF Q493X nAPG09882 Target 3"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 154
tttctttaa tgggtgcagg cataatccag gaaaactaag aacagaatga aattttcca 60
ctgtg                                         65

SEQ ID NO: 155      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF Q493X nAPG09882 Target 4"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 155
tttcttaat ggtgccaggc ataatccagg aaaactaaga acagaatgaa attttccac 60
tgtgc                                         65

SEQ ID NO: 156      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF Q493X nAPG09882 Target 5"
source            1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 156
ttctttaatgt gccaggca taatccagga aaactaagaa cagaatgaaa ttcttcact 60
gtgtc                                         65

SEQ ID NO: 157      moltype = DNA  length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..60
note = source = /note="CF Q493X nAPG09748 Target 1"
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 157
taagcacagt ggaagaattt cattctgttc tttagtttcc tggattatgc ctggaccat 60

SEQ ID NO: 158      moltype = DNA  length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..60
note = source = /note="CF Q493X nAPG09748 Target 2"
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 158
aagcacagtg gaagaatttc attctgttct tagtttccct ggattatgcc tggcaccatt 60

SEQ ID NO: 159      moltype = DNA  length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

```

---

---

-continued

---

```

misc_feature          1..60
note = source = /note="CF Q493X nAPG09748 Target 3"
source               1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 159
acagtggaaag aatttcattc tgttcttagt tttcctggat tatgcctggc accattaaag  60

SEQ ID NO: 160      moltype = DNA length = 60
FEATURE             Location/Qualifiers
misc_feature        1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
source               1..60
note = source = /note="CF Q493X nAPG09748 Target 4"
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 160
ggaagaattt cattctgttc tttagtttcc tggattatgc ctggcaccat taaagaaaaat  60

SEQ ID NO: 161      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
source               1..65
note = source = /note="CF Q493X nAPG00969 Target 1"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 161
gatattttct ttaatggtgc caggcataat ccagggaaac taagaacaga atgaaaattct  60
tccac                           65

SEQ ID NO: 162      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
source               1..65
note = source = /note="CF Q493X nAPG00969 Target 2"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 162
ttctttaatg gtgccaggca taatccagga aaactaagaa cagaatgaaa ttcttccact  60
gtgct                           65

SEQ ID NO: 163      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
source               1..65
note = source = /note="CF Q493X nAPG06646 Target 1"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 163
tttaatggtg ccaggcataa tccaggaaaa ctaagaacag aatgaaaattc ttccactgtg  60
cttaa                           65

SEQ ID NO: 164      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature        1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
source               1..65
note = source = /note="CF Q493X nAPG06646 Target 2"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 164
aatggtgcca ggcataatcc aggaaaacta agaacagaat gaaattctc cactgtgctt  60
aattt                           65

```

---

-continued

---

```

SEQ ID NO: 165      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF Q493X nAPG01604 Target 1"
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 165
ttcttaatg gtgccaggca taatccagga aaactaagaa cagaatgaaa ttcttccact 60

SEQ ID NO: 166      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF Q493X nAPG01604 Target 2"
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 166
ttaatggc caggcataat ccaggaaaaac taagaacaga atgaaaattct tccactgtc 60

SEQ ID NO: 167      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF Q493X nAPG03850 Target 1"
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 167
cttaatggt gccaggcata atccaggaaa actaagaaca gaatgaaatt cttccactgt 60

SEQ ID NO: 168      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF Q493X nAPG07553 Target 1"
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 168
cttaatggt gccaggcata atccaggaaa actaagaaca gaatgaaatt cttccactgt 60

SEQ ID NO: 169      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF Q493X nAPG09882 Target 1"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 169
caggcataat ccaggaaaaac taaga                                25

SEQ ID NO: 170      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF Q493X nAPG09882 Target 2"
1..25
mol_type = other DNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 170
aggcataatc caggaaaact aagaa 25

SEQ ID NO: 171      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature    1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..25
note = source = /note="CF Q493X nAPG09882 Target 3"
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 171
cataatccag gaaaactaag aacag 25

SEQ ID NO: 172      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature    1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..25
note = source = /note="CF Q493X nAPG09882 Target 4"
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 172
ataatccagg aaaactaaga acaga 25

SEQ ID NO: 173      moltype = DNA  length = 25
FEATURE          Location/Qualifiers
misc_feature    1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..25
note = source = /note="CF Q493X nAPG09882 Target 5"
source          1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 173
taatccagga aaactaagaa cagaa 25

SEQ ID NO: 174      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature    1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..20
note = source = /note="CF Q493X nAPG09748 Target 1"
source          1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 174
cattctgttc ttagtttcc 20

SEQ ID NO: 175      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature    1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..20
note = source = /note="CF Q493X nAPG09748 Target 2"
source          1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 175
attctgttct tagtttccct 20

SEQ ID NO: 176      moltype = DNA  length = 20
FEATURE          Location/Qualifiers
misc_feature    1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature    1..20
note = source = /note="CF Q493X nAPG09748 Target 3"
source          1..20

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 176
tgttcttagt tttcttgat                                     20

SEQ ID NO: 177      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="CF Q493X nAPG09748 Target 4"
source            1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 177
ttagtttcc tggattatgc                                     20

SEQ ID NO: 178      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF Q493X nAPG00969 Target 1"
source            1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 178
caggcataat ccaggaaaaac taaga                                     25

SEQ ID NO: 179      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF Q493X nAPG00969 Target 2"
source            1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 179
taatccagga aaactaagaa cagaa                                     25

SEQ ID NO: 180      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF Q493X nAPG06646 Target 1"
source            1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 180
tccaggaaaa ctaagaacag aatga                                     25

SEQ ID NO: 181      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF Q493X nAPG06646 Target 2"
source            1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 181
aggaaaaacta agaacagaat gaaat                                     25

SEQ ID NO: 182      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..20

```

---

---

-continued

---

```

source          note = source = /note="CF Q493X nAPG01604 Target 1"
                1..20
                mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 182
taatccagga aaactaagaa                                         20

SEQ ID NO: 183      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                    Syntheticoligonucleotide"
1..20
note = source = /note="CF Q493X nAPG01604 Target 2"
1..20
source          mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 183
ccaggaaaaac taagaacaga                                         20

SEQ ID NO: 184      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                    Syntheticoligonucleotide"
1..20
note = source = /note="CF Q493X nAPG03850 Target 1"
1..20
source          mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 184
atccaggaaa actaagaaca                                         20

SEQ ID NO: 185      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                    Syntheticoligonucleotide"
1..20
note = source = /note="CF Q493X nAPG07553 Target 1"
1..20
source          mol_type = other DNA
                organism = synthetic construct

SEQUENCE: 185
atccaggaaa actaagaaca                                         20

SEQ ID NO: 186      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
                    Syntheticpolynucleotide"
1..118
note = source = /note="CF Q493X nAPG09882 Target 1 sgRNA"
1..118
source          mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 186
caggcataat ccaggaaaaac taagagttt tgtactctca ataaaaagtt attgagaatc  60
tacaaaaata aggcattttg ccgaatttac cgccctacat atgtagggcg gttttttt   118

SEQ ID NO: 187      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
                    Syntheticpolynucleotide"
1..118
note = source = /note="CF Q493X nAPG09882 Target 2 sgRNA"
1..118
source          mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 187
aggcataat caggaaaaact aagaagttt tgtactctca ataaaaagtt attgagaatc  60
tacaaaaata aggcattttg ccgaatttac cgccctacat atgtagggcg gttttttt   118

SEQ ID NO: 188      moltype = RNA length = 118
FEATURE           Location/Qualifiers

```

---

-continued

---

```

misc_feature          1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..118
note = source = /note="CF Q493X nAPG09882 Target 3 sgRNA"
source                1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 188
cataatccag qaaaactaaag aacagggttt tgcgtctca ataaaaagg attgagaatc 60
tacaaaaata aggcattttg ccgaatttc cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 189      moltype = RNA length = 118
FEATURE
misc_feature          1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..118
note = source = /note="CF Q493X nAPG09882 Target 4 sgRNA"
source                1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 189
taaatccagg aaaactaaaga acagaggttt tgcgtctca ataaaaagg attgagaatc 60
tacaaaaata aggcattttg ccgaatttc cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 190      moltype = RNA length = 118
FEATURE
misc_feature          1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..118
note = source = /note="CF Q493X nAPG09882 Target 5 sgRNA"
source                1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 190
taatccagga aaactaagaa cagaaggttt tgcgtctca ataaaaagg attgagaatc 60
tacaaaaata aggcattttg ccgaatttc cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 191      moltype = RNA length = 135
FEATURE
misc_feature          1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..135
note = source = /note="CF Q493X nAPG09748 Target 1 sgRNA"
source                1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 191
cgacggtag aggccgtatg tcgatttgc ttaatttcgat gcgtgtcat tgcgtcctc 60
cattacaggg cggctaccac gaatagccac gaagtaaaag ctgcgtggct agcaccatc 120
tgttcttagt ttccc                                135

SEQ ID NO: 192      moltype = RNA length = 135
FEATURE
misc_feature          1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..135
note = source = /note="CF Q493X nAPG09748 Target 2 sgRNA"
source                1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 192
cgacggtag aggccgtatg tcgatttgc ttaatttcgat gcgtgtcat tgcgtcctc 60
cattacaggg cggctaccac gaatagccac gaagtaaaag ctgcgtggct agcaccatc 120
tgttcttagt ttccc                                135

SEQ ID NO: 193      moltype = RNA length = 135
FEATURE
misc_feature          1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..135

```

---

---

-continued

---

```

source          note = source = /note="CF Q493X nAPG09748 Target 3 sgRNA"
                1..135
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 193
cgacggtag aggccgtatg tcgatttgc ttaattcgt gcgtgtcat tgctgtc 60
cattacaggg cggctaccac gaatagccac gaagtaaaag ctctgtggct agca 120
ttatcc tggat                                135

SEQ ID NO: 194      moltype = RNA length = 135
FEATURE          Location/Qualifiers
misc_feature     1..135
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
                1..135
                note = source = /note="CF Q493X nAPG09748 Target 4 sgRNA"
                1..135
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 194
cgacggtag aggccgtatg tcgatttgc ttaattcgt gcgtgtcat tgctgtc 60
cattacaggg cggctaccac gaatagccac gaagtaaaag ctctgtggct agca 120
tttcctggat tatgc                                135

SEQ ID NO: 195      moltype = RNA length = 118
FEATURE          Location/Qualifiers
misc_feature     1..118
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
                1..118
                note = source = /note="CF Q493X nAPG00969 Target 1 sgRNA"
                1..118
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 195
caggcataat ccagggaaac taagagttt agtactctgt gaaagcacag aatctactaa 60
aataaggcat aatgccgtat ttaatccat cataattctg atgggattt ttatattt    118

SEQ ID NO: 196      moltype = RNA length = 118
FEATURE          Location/Qualifiers
misc_feature     1..118
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
                1..118
                note = source = /note="CF Q493X nAPG00969 Target 2 sgRNA"
                1..118
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 196
taatccagga aaactaagaa cagaagttt agtactctgt gaaagcacag aatctactaa 60
aataaggcat aatgccgtat ttaatccat cataattctg atgggattt ttatattt    118

SEQ ID NO: 197      moltype = RNA length = 164
FEATURE          Location/Qualifiers
misc_feature     1..164
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
                1..164
                note = source = /note="CF Q493X nAPG06646 Target 1 sgRNA"
                1..164
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 197
tccaggaaaa ctaagaacag aatggccat aattccctcg taaaactaa agaaggtaa 60
tagagttatt atggtaaggc aatatgcgt ggcgttgggg atcgcctatg tccggttta 120
ccggatctcc ctaaagggtga ttatcttgg ttatcacct tttt                164

SEQ ID NO: 198      moltype = RNA length = 164
FEATURE          Location/Qualifiers
misc_feature     1..164
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polynucleotide"
                1..164
                note = source = /note="CF Q493X nAPG06646 Target 2 sgRNA"
                1..164
                mol_type = other RNA

```

---

---

-continued

---

```

SEQUENCE: 198          organism = synthetic construct
aggaaaaacta agaacagaat gaaatgccat aattcctctg taaaacttaa agaaggtta  60
tagagtatt atggtaaggc aatatgcgtt ggcgttgggg atcgctatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg ttagtcacct tttt                164

SEQ ID NO: 199          moltype = RNA  length = 105
FEATURE
misc_feature           Location/Qualifiers
1..105
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature           1..105
note = source = /note="CF Q493X nAPG01604 Target 1 sgRNA"
source                 1..105
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 199
taatccagga aactaagaa gtttagtac tctgtaaaaa gttacagaat ctactaaaac  60
aaggcaaat gccgtgttta tctcgtcaac ttgttggcga gattt                  105

SEQ ID NO: 200          moltype = RNA  length = 105
FEATURE
misc_feature           Location/Qualifiers
1..105
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature           1..105
note = source = /note="CF Q493X nAPG01604 Target 2 sgRNA"
source                 1..105
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 200
cgaggaaaac taagaacaga gtttagtac tctgtaaaaa gttacagaat ctactaaaac  60
aaggcaaat gccgtgttta tctcgtcaac ttgttggcga gattt                  105

SEQ ID NO: 201          moltype = RNA  length = 163
FEATURE
misc_feature           Location/Qualifiers
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature           1..163
note = source = /note="CF Q493X nAPG03850 Target 1 sgRNA"
source                 1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 201
atccaggaaa actaagaaca gctatagttc cataagaaaa aagtttctta agttactata  60
gttaaggccaa tgaccctgtgg cggttggga tcgccttata ctggatgtga tattctccc 120
atgtgaaaag cacctaagca tagcgtatg gtgttttat ttt                163

SEQ ID NO: 202          moltype = RNA  length = 151
FEATURE
misc_feature           Location/Qualifiers
1..151
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature           1..151
note = source = /note="CF Q493X nAPG07553 Target 1 sgRNA"
source                 1..151
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 202
atccaggaaa actaagaaca gctatagttc cataagaaa cttaagttac tatagtaagg  60
gcaatgaccc gtggcgttt gggatcgcct catccattac ggatattctc cccatgtgaa 120
aagcacctaa gcataaggct aagtgcttt t                151

SEQ ID NO: 203          moltype = DNA  length = 65
FEATURE
misc_feature           Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature           1..65
note = source = /note="CF R553X nAPG06646 Target 1"
source                 1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 203
ccaataatta gttattcacc ttgctaaaga aattcttgct cattgacctc cactcagtgt  60

```

---

-continued

---

```

gattc                                65

SEQ ID NO: 204      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG06646 Target 2"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 204
caataatttag ttattcacct tggtaaaagaa attcttgctc attgacacctc actcagtgtg 60
attcc                                         65

SEQ ID NO: 205      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG06646 Target 3"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 205
ataaatttagtt attcaccttg ctaaaagaaat tcttgctcat tgacacctcc acgtgtgat 60
tccac                                         65

SEQ ID NO: 206      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG06646 Target 4"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 206
aatttagttt tcacaccttgct aaagaaaattc ttgctcattg acctccactc agtgtgattc 60
cacct                                         65

SEQ ID NO: 207      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG06646 Target 5"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 207
tcacaccttgct aaagaaaattc ttgctcattg acctccactc agtgtgattc cacttctcc 60
aagaa                                         65

SEQ ID NO: 208      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG06646 Target 6"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 208
caccttgcta aagaaaattct tgctcattga cctccactca gtgtgattcc accttctcca 60
agaac                                         65

SEQ ID NO: 209      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:

```

---

---

-continued

---

```

misc_feature      Synthetic oligonucleotide"
source          1..65
                note = source = /note="CF R553X nAPG06646 Target 7"
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 209
ccttgctaaa gaaattcttg ctcattgacc tccactcagt gtgattccac cttctccaag  60
aacta                               65

SEQ ID NO: 210      moltype = DNA length = 65
FEATURE
misc_feature      Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF R553X nAPG07433.1 Target 1"
source          1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 210
ccaaataatta gttattcacc ttgctaaaga aattcttgct cattgacctc cactcagtgt  60
gattc                               65

SEQ ID NO: 211      moltype = DNA length = 65
FEATURE
misc_feature      Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF R553X nAPG07433.1 Target 2"
source          1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 211
tcacccttgct aaagaaattc ttgctcattt acctccactc agtgtgattc caccttctcc  60
aagaa                               65

SEQ ID NO: 212      moltype = DNA length = 65
FEATURE
misc_feature      Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF R553X nAPG07433.1 Target 3"
source          1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 212
ccttgctaaa gaaattcttg ctcattgacc tccactcagt gtgattccac cttctccaag  60
aacta                               65

SEQ ID NO: 213      moltype = DNA length = 65
FEATURE
misc_feature      Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF R553X nAPG09882 Target 1"
source          1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 213
aataattagt tattcacctt gctaaagaaa ttcttgctca ttgacctcca ctcagtgtga  60
ttcca                               65

SEQ ID NO: 214      moltype = DNA length = 65
FEATURE
misc_feature      Location/Qualifiers
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF R553X nAPG09882 Target 2"
source          1..65
mol_type = other DNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 214
attagttatt caccttgcta aagaaattct tgctcattga cctccactca gtgtgatcc 60
acctt                                         65

SEQ ID NO: 215      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF R553X nAPG09882 Target 3"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 215
tattcacctt gctaaagaaa ttcttgctca ttgacacctca ctcagtgtga ttccaccc 60
tccaa                                         65

SEQ ID NO: 216      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF R553X nAPG03850 Target 1"
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 216
tattcacctt gctaaagaaa ttcttgctca ttgacacctca ctcagtgtga ttccaccc 60

SEQ ID NO: 217      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF R553X nAPG03850 Target 2"
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 217
ttcaccttgc taaagaaatt ctgtgttgcatt gacccactt ctagtgttgc 60

SEQ ID NO: 218      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..60
note = source = /note="CF R553X nAPG03850 Target 3"
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 218
caccttgcta aagaaattct tgctcattga cctccactca gtgtgattcc accttctcca 60

SEQ ID NO: 219      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25
note = source = /note="CF R553X nAPG06646 Target 1"
source
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 219
ttgctaaaga aattcttgct cattg                                         25

SEQ ID NO: 220      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..25

```

---

-continued

---

```

source          note = source = /note="CF R553X nAPG06646 Target 2"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 220
tgctaaagaa attcttgctc attga                                25

SEQ ID NO: 221      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
source          note = source = /note="CF R553X nAPG06646 Target 3"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 221
ctaaagaaat tcttgctcat tgacc                                25

SEQ ID NO: 222      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG06646 Target 4"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 222
aaagaaaattc ttgctcattt acctc                                25

SEQ ID NO: 223      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG06646 Target 5"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 223
ttgctcattt acctccactc agtgt                                25

SEQ ID NO: 224      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG06646 Target 6"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 224
tgctcattta cctccactca gtgtg                                25

SEQ ID NO: 225      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG06646 Target 7"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 225
ctcattgacc tccactcagt gtgtat                                25

SEQ ID NO: 226      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:

```

---

-continued

---

```

misc_feature      Synthetic oligonucleotide"
1..25
source          note = source = /note="CF R553X nAPG07433.1 Target 1"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 226
ttgctaaaga aattcttgct cattg                                25

SEQ ID NO: 227      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
source          note = source = /note="CF R553X nAPG07433.1 Target 2"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 227
ttgctcattg acctccactc agtgt                                25

SEQ ID NO: 228      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG07433.1 Target 3"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 228
ctcattgacc tccactcagt gtgat                                25

SEQ ID NO: 229      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG09882 Target 1"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 229
gtttaagaaa ttcttgctca ttgac                                25

SEQ ID NO: 230      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG09882 Target 2"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 230
aagaaaattct tgctcattga cctcc                                25

SEQ ID NO: 231      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R553X nAPG09882 Target 3"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 231
ttcttgctca ttgacacctca ctcag                                25

SEQ ID NO: 232      moltype = DNA length = 20
FEATURE          Location/Qualifiers

```

---

-continued

---

```

misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="CF R553X nAPG03850 Target 1"
source                1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 232
tttttgtcata ttgacactca                                         20

SEQ ID NO: 233      moltype = DNA length = 20
FEATURE             Location/Qualifiers
misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="CF R553X nAPG03850 Target 2"
source                1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 233
cttgctcatt gacccactca                                         20

SEQ ID NO: 234      moltype = DNA length = 20
FEATURE             Location/Qualifiers
misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="CF R553X nAPG03850 Target 3"
source                1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 234
tgctcattga cctccactca                                         20

SEQ ID NO: 235      moltype = RNA length = 164
FEATURE             Location/Qualifiers
misc_feature          1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..164
note = source = /note="CF R553X nAPG06646 Target 1 sgRNA"
source                1..164
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 235
ttgtctaaaga aattcttgc tattggccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt              164

SEQ ID NO: 236      moltype = RNA length = 164
FEATURE             Location/Qualifiers
misc_feature          1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..164
note = source = /note="CF R553X nAPG06646 Target 2 sgRNA"
source                1..164
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 236
tgctctaaaga attcttgc tattggccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt              164

SEQ ID NO: 237      moltype = RNA length = 164
FEATURE             Location/Qualifiers
misc_feature          1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..164
note = source = /note="CF R553X nAPG06646 Target 3 sgRNA"
source                1..164
mol_type = other RNA

```

---

---

-continued

---

```

SEQUENCE: 237          organism = synthetic construct
ctaaagaat tcttgctcat tgaccgccat aattcctctg taaaacttaa agaaggtta  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt             164

SEQ ID NO: 238          moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R553X nAPG06646 Target 4 sgRNA"
source
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 238          organism = synthetic construct
aaagaaaattc ttgctcattt acctcgccat aattcctctg taaaacttaa agaaggtta  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt             164

SEQ ID NO: 239          moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R553X nAPG06646 Target 5 sgRNA"
source
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 239          organism = synthetic construct
ttgctcattt acctccactc agtgtgcccattt aattcctctg taaaacttaa agaaggtta  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt             164

SEQ ID NO: 240          moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R553X nAPG06646 Target 6 sgRNA"
source
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 240          organism = synthetic construct
tgctcatttga cttccactca gtgtggccat aattcctctg taaaacttaa agaaggtta  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt             164

SEQ ID NO: 241          moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R553X nAPG06646 Target 7 sgRNA"
source
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 241          organism = synthetic construct
ctcatttggacc tccacttgcgt gtgtggccat aattcctctg taaaacttaa agaaggtta  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgccatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt             164

SEQ ID NO: 242          moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..135
note = source = /note="CF R553X nAPG07433.1 Target 1 sgRNA"
source
mol_type = other RNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 242
ttgctaaaga aattcttgct cattggcat agtccattt aagccaaaag tggctttat 60
gtttctatga taagggttgc gaccctggc gtcgggatc gcctgcccatt gaaatggc 120
ttctccccat ttatt 135

SEQ ID NO: 243      moltype = RNA length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="CF R553X nAPG07433.1 Target 2 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 243
ttgctcattt acctccactc agtgtgtcat agtccattt aagccaaaag tggctttat 60
gtttctatga taagggttgc gaccctggc gtcgggatc gcctgcccatt gaaatggc 120
ttctccccat ttatt 135

SEQ ID NO: 244      moltype = RNA length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="CF R553X nAPG07433.1 Target 3 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 244
cttatttacc tccactttagt gtgtatgtcat agtccattt aagccaaaag tggctttat 60
gtttctatga taagggttgc gaccctggc gtcgggatc gcctgcccatt gaaatggc 120
ttctccccat ttatt 135

SEQ ID NO: 245      moltype = RNA length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF R553X nAPG09882 Target 1 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 245
gctaaagaaa ttctttgtca ttgacgtttt tgtactctca ataaaaaggattt attgagaatc 60
tacaaaaata aggcattttt ccgaattttac cgccctacat atgttagggcg gttttttt 118

SEQ ID NO: 246      moltype = RNA length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF R553X nAPG09882 Target 2 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 246
aaaaaaatttct tgctcatttga cctccgtttt tgtactctca ataaaaaggattt attgagaatc 60
tacaaaaata aggcattttt ccgaattttac cgccctacat atgttagggcg gttttttt 118

SEQ ID NO: 247      moltype = RNA length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF R553X nAPG09882 Target 3 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 247
ttctttgtca ttgacgtttt ctcagggtttt tgtactctca ataaaaaggattt attgagaatc 60
tacaaaaata aggcattttt ccgaattttac cgccctacat atgttagggcg gttttttt 118

```

---

-continued

---

```

SEQ ID NO: 248      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..163
note = source = /note="CF R553X nAPG03850 Target 1 sgRNA"
1..163
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 248
ttcttgcgtca ttgacctcca gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaaggcaa tgaccctgtgg cgtttggga tcgccttata ctggtatgga tattctcccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt 163

SEQ ID NO: 249      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..163
note = source = /note="CF R553X nAPG03850 Target 2 sgRNA"
1..163
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 249
cttgctcatt gacctccact gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaaggcaa tgaccctgtgg cgtttggga tcgccttata ctggtatgga tattctcccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt 163

SEQ ID NO: 250      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..163
note = source = /note="CF R553X nAPG03850 Target 3 sgRNA"
1..163
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 250
tgctcattgtca cctccactca gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaaggcaa tgaccctgtgg cgtttggga tcgccttata ctggtatgga tattctcccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt 163

SEQ ID NO: 251      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic coligonucleotide"
1..65
note = source = /note="CF R1162X nAPG09882 Target 1"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 251
ggtttacctt ctgttggcat gtcaatgaac ttaaagactc agctcacaga tcgcacatctga 60
aataaa 65

SEQ ID NO: 252      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic coligonucleotide"
1..65
note = source = /note="CF R1162X nAPG09882 Target 2"
1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 252
accttctgtt ggcatgtcaa tgaacttaaa gactcagctc acagatcgca tctgaaataa 60
aaataa 65

SEQ ID NO: 253      moltype = DNA  length = 65
FEATURE

```

---

-continued

---

```

misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF R1162X nAPG09882 Target 3"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 253
ctgttggcat gtcaatgaac ttaagactc agtcacaga tcgcatctga aataaaataa  60
acaac                           65

SEQ ID NO: 254      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF R1162X nAPG09882 Target 4"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 254
tgttggcatg tcaatgaact taaagactca gctcacagat cgcatctgaa ataaaaataa  60
caaca                           65

SEQ ID NO: 255      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF R1162X nAPG09882 Target 5"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 255
gttggcatgt caatgaactt aaagactcg ctcacagatc gcatctgaaa taaaataaac  60
aacat                           65

SEQ ID NO: 256      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF R1162X nAPG06646 Target 1"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 256
tttacacctt gttggcatgt caatgaactt aaagactcg ctcacagatc gcatctgaaa  60
taaaa                           65

SEQ ID NO: 257      moltype = DNA length = 65
FEATURE
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF R1162X nAPG06646 Target 2"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 257
tacccctgtg tggcatgtca atgaacttaa agactcgat cacagatcgc atctgaaata  60
aaaat                           65

SEQ ID NO: 258      moltype = DNA length = 60
FEATURE
misc_feature          1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..60
note = source = /note="CF R1162X nAPG06646 Target 3"
source               1..60

```

---

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 258
tggcatgtca atgaacttaa agactcagct cacagatcgc atctgaaata aaaataacaa 60

SEQ ID NO: 259      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60
                  note = source = /note="CF R1162X nAPG03850 Target 1"
source            1..60
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 259
tacccattctgt tggcatgtca atgaacttaa agactcagct cacagatcgc atctgaaata 60

SEQ ID NO: 260      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60
                  note = source = /note="CF R1162X nAPG03850 Target 2"
source            1..60
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 260
ttctgttggc atgtcaatga acttaaagac tcagctcaca gatcgcatct gaaataaaaa 60

SEQ ID NO: 261      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60
                  note = source = /note="CF R1162X nAPG03850 Target 3"
source            1..60
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 261
tggcatgtca atgaacttaa agactcagct cacagatcgc atctgaaata aaaataacaa 60

SEQ ID NO: 262      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60
                  note = source = /note="CF R1162X nAPG05586 Target 1"
source            1..60
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 262
ttacccattctgt ttggcatgtc aatgaactta aagactcagc tcacagatcg catctgaaat 60

SEQ ID NO: 263      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60
                  note = source = /note="CF R1162X nAPG05586 Target 2"
source            1..60
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 263
ctgttggcat gtcaatgaac ttaaagactc agctcacaga tcgcacatcg aataaaaata 60

SEQ ID NO: 264      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..60

```

---

---

-continued

---

```

source          note = source = /note="CF R1162X nAPG05586 Target 3"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 264
tgtcaatgaa cttaaagact cagtcacag atcgcatctg aaataaaaat aacaacattt 60

SEQ ID NO: 265      moltype = DNA length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..65
note = source = /note="CF R1162X nAPG00969 Target 1"
1..65
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 265
ggtttacatt ctgttggcat gtcaatgaac ttaaagactc agtcacaga tcgcacatctga 60
aataa                      65

SEQ ID NO: 266      moltype = DNA length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..65
note = source = /note="CF R1162X nAPG00969 Target 2"
1..65
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 266
gtggcatgt caatgaactt aaagactcg ctcacagatc gcatctgaaa taaaaataac 60
aacat                      65

SEQ ID NO: 267      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF R1162X nAPG07553 Target 1"
1..60
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 267
tggcatgtca atgaacttaa agactcgat cacagatcgc atctgaaaataaaaacaa 60

SEQ ID NO: 268      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..60
note = source = /note="CF R1162X nAPG01604 Target 1"
1..60
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 268
gcatgtcaat gaacttaaag actcgactca cagatcgat ctgaaaataaaa aataacaaca 60

SEQ ID NO: 269      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF R1162X nAPG09882 Target 1"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 269
gtcaatgaac ttaaagactc agctc

SEQ ID NO: 270      moltype = DNA length = 25
FEATURE           Location/Qualifiers

```

---

-continued

---

```

misc_feature          1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..25
note = source = /note="CF R1162X nAPG09882 Target 2"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 270
tgaacttaaa gactcagtc acaga                                25

SEQ ID NO: 271      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature        1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="CF R1162X nAPG09882 Target 3"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 271
ttaaagactc agctcacaga tcgca                                25

SEQ ID NO: 272      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature        1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="CF R1162X nAPG09882 Target 4"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 272
taaagactca gtcacagat cgcata                               25

SEQ ID NO: 273      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature        1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="CF R1162X nAPG09882 Target 5"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 273
aaagactcg ctcacagat gcata                                25

SEQ ID NO: 274      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature        1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="CF R1162X nAPG06646 Target 1"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 274
caatgaacct aaagactcg ctcac                                25

SEQ ID NO: 275      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature        1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="CF R1162X nAPG06646 Target 2"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 275
atgaacttaa agactcagct cacag                                25

```

---

-continued

---

```

SEQ ID NO: 276      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..25
                  note = source = /note="CF R1162X nAPG06646 Target 3"
source            1..25
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 276
aacttaaaga ctcagtcac agatc                                25

SEQ ID NO: 277      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..20
                  note = source = /note="CF R1162X nAPG03850 Target 1"
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 277
atgaacttaa agactcagct                                20

SEQ ID NO: 278      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..20
                  note = source = /note="CF R1162X nAPG03850 Target 2"
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 278
acttcaaagac tcagtcacac                                20

SEQ ID NO: 279      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..20
                  note = source = /note="CF R1162X nAPG03850 Target 3"
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 279
agactcagct cacagatcgc                                20

SEQ ID NO: 280      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..20
                  note = source = /note="CF R1162X nAPG05586 Target 1"
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 280
aatgaactta aagactcagc                                20

SEQ ID NO: 281      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic oligonucleotide"
misc_feature      1..20
                  note = source = /note="CF R1162X nAPG05586 Target 2"
source            1..20
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 281

```

---

-continued

---

ttaaagactc agctcacaga	20
SEQ ID NO: 282	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = source = /note="Description of Artificial Sequence: Syntheticoligonucleotide"
misc_feature	1..20
	note = source = /note="CF R1162X nAPG05586 Target 3"
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 282	
cagctcacag atcgcatctg	20
SEQ ID NO: 283	moltype = DNA length = 25
FEATURE	Location/Qualifiers
misc_feature	1..25
	note = source = /note="Description of Artificial Sequence: Syntheticoligonucleotide"
misc_feature	1..25
	note = source = /note="CF R1162X nAPG00969 Target 1"
source	1..25
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 283	
gtcaatgaac ttaaagactc agctc	25
SEQ ID NO: 284	moltype = DNA length = 25
FEATURE	Location/Qualifiers
misc_feature	1..25
	note = source = /note="Description of Artificial Sequence: Syntheticoligonucleotide"
misc_feature	1..25
	note = source = /note="CF R1162X nAPG00969 Target 2"
source	1..25
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 284	
aaagacttag ctcacagatc gcata	25
SEQ ID NO: 285	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = source = /note="Description of Artificial Sequence: Syntheticoligonucleotide"
misc_feature	1..20
	note = source = /note="CF R1162X nAPG07553 Target 1"
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 285	
agactcaact cacagatcgc	20
SEQ ID NO: 286	moltype = DNA length = 20
FEATURE	Location/Qualifiers
misc_feature	1..20
	note = source = /note="Description of Artificial Sequence: Syntheticoligonucleotide"
misc_feature	1..20
	note = source = /note="CF R1162X nAPG01604 Target 1"
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 286	
actcagtc a cagatcgcat	20
SEQ ID NO: 287	moltype = RNA length = 118
FEATURE	Location/Qualifiers
misc_feature	1..118
	note = source = /note="Description of Artificial Sequence: Syntheticpolynucleotide"
misc_feature	1..118
	note = source = /note="CF R1162X nAPG09882 Target 1 sgRNA"
source	1..118
	mol_type = other RNA

---

-continued

---

```

SEQUENCE: 287          organism = synthetic construct
gtcaatgaac ttaaagactc agctcgaaaa ttgtactctca ataaaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cggccatcat atgttagggcg gttttttt    118

SEQ ID NO: 288          moltype = RNA  length = 118
FEATURE                  Location/Qualifiers
misc_feature             1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature             1..118
note = source = /note="CF R1162X nAPG09882 Target 2 sgRNA"
source                   1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 288          moltype = RNA  length = 118
tgaacttaaa gactcagctc acagatttt ttgtactctca ataaaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cggccatcat atgttagggcg gttttttt    118

SEQ ID NO: 289          moltype = RNA  length = 118
FEATURE                  Location/Qualifiers
misc_feature             1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature             1..118
note = source = /note="CF R1162X nAPG09882 Target 3 sgRNA"
source                   1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 289          moltype = RNA  length = 118
ttaaagactc agctcacaga tcgcagaaaa ttgtactctca ataaaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cggccatcat atgttagggcg gttttttt    118

SEQ ID NO: 290          moltype = RNA  length = 118
FEATURE                  Location/Qualifiers
misc_feature             1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature             1..118
note = source = /note="CF R1162X nAPG09882 Target 4 sgRNA"
source                   1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 290          moltype = RNA  length = 118
taaaagactca gotcacagat cgcatgtttt ttgtactctca ataaaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cggccatcat atgttagggcg gttttttt    118

SEQ ID NO: 291          moltype = RNA  length = 118
FEATURE                  Location/Qualifiers
misc_feature             1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature             1..118
note = source = /note="CF R1162X nAPG09882 Target 5 sgRNA"
source                   1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 291          moltype = RNA  length = 118
aaagactctag ctcacagatc gcacgtttt ttgtactctca ataaaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaatttac cggccatcat atgttagggcg gttttttt    118

SEQ ID NO: 292          moltype = RNA  length = 164
FEATURE                  Location/Qualifiers
misc_feature             1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature             1..164
note = source = /note="CF R1162X nAPG06646 Target 1 sgRNA"
source                   1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 292          moltype = RNA  length = 164
caatgaacctt aaagactctag ctcacgcccattt aattcctctg taaaacttaa agaaggttta 60
tagagtattt atggtaaggc aatatggcg ggcgttgggg atcgccatgt tccgggtttta 120
ccggatctcc ctaaagggtga ctaactttgg ttgtcacct tttt                164

```

-continued

---

```

SEQ ID NO: 293      moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R1162X nAPG06646 Target 2 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 293
atgaacttaa agactcagct cacaggccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgctatg tccggttta 120
cccgatctcc ctaaaaggta ctaactttgg tttagtcacct tttt                164

SEQ ID NO: 294      moltype = RNA  length = 164
FEATURE
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF R1162X nAPG06646 Target 3 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 294
aacttaaaga ctcagotcac agatggccat aattcctctg taaaacttaa agaaggtaa  60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atcgctatg tccggttta 120
cccgatctcc ctaaaaggta ctaactttgg tttagtcacct tttt                164

SEQ ID NO: 295      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..163
note = source = /note="CF R1162X nAPG03850 Target 1 sgRNA"
source
1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 295
atgaacttaa agactcagct gctatagttc cataagaaaa aagtttctta agttactata  60
gttaaggccaa tgacccgtgg cgtttgggg tcgccttattc ctggtatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcttttat ttt                  163

SEQ ID NO: 296      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..163
note = source = /note="CF R1162X nAPG03850 Target 2 sgRNA"
source
1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 296
acttaaagac tcagctcaca gctatagttc cataagaaaa aagtttctta agttactata  60
gttaaggccaa tgacccgtgg cgtttgggg tcgccttattc ctggtatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcttttat ttt                  163

SEQ ID NO: 297      moltype = RNA  length = 163
FEATURE
misc_feature
1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..163
note = source = /note="CF R1162X nAPG03850 Target 3 sgRNA"
source
1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 297
agactcagct cacagatcgc gctatagttc cataagaaaa aagtttctta agttactata  60
gttaaggccaa tgacccgtgg cgtttgggg tcgccttattc ctggtatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcttttat ttt                  163

SEQ ID NO: 298      moltype = RNA  length = 110

```

-continued

---

```

FEATURE                               Location/Qualifiers
misc_feature                         1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..110
note = source = /note="CF R1162X nAPG05586 Target 1 sgRNA"
source                                1..110
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 298
atagaactta aagactcagc gttattgtac tctcaataaa aagttattga gaatctacaa  60
taataaggca ttttgcggaa ttaccgccc tacatatgtat gggcggttt           110

SEQ ID NO: 299      moltype = RNA  length = 110
FEATURE                               Location/Qualifiers
misc_feature                         1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..110
note = source = /note="CF R1162X nAPG05586 Target 2 sgRNA"
source                                1..110
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 299
ttaaagactc agctcacaga gttattgtac tctcaataaa aagttattga gaatctacaa  60
taataaggca ttttgcggaa ttaccgccc tacatatgtat gggcggttt           110

SEQ ID NO: 300      moltype = RNA  length = 110
FEATURE                               Location/Qualifiers
misc_feature                         1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..110
note = source = /note="CF R1162X nAPG05586 Target 3 sgRNA"
source                                1..110
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 300
caagtcacac atcgcatctg gttattgtac tctcaataaa aagttattga gaatctacaa  60
taataaggca ttttgcggaa ttaccgccc tacatatgtat gggcggttt           110

SEQ ID NO: 301      moltype = RNA  length = 118
FEATURE                               Location/Qualifiers
misc_feature                         1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..118
note = source = /note="CF R1162X nAPG00969 Target 1 sgRNA"
source                                1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 301
gtcaatgaac ttaaagactc agctcggttt agtactctgt gaaagcacag aatctactaa  60
aataaggcat aatgccgtat ttaatcccat cataattctg atgggatttt ttatattt     118

SEQ ID NO: 302      moltype = RNA  length = 118
FEATURE                               Location/Qualifiers
misc_feature                         1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..118
note = source = /note="CF R1162X nAPG00969 Target 2 sgRNA"
source                                1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 302
aaagactcag ctcacagatc gcacatgttt agtactctgt gaaagcacag aatctactaa  60
aataaggcat aatgccgtat ttaatcccat cataattctg atgggatttt ttatattt     118

SEQ ID NO: 303      moltype = RNA  length = 151
FEATURE                               Location/Qualifiers
misc_feature                         1..151
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature                         1..151
note = source = /note="CF R1162X nAPG07553 Target 1 sgRNA"

```

---

-continued

---

```

source          1..151
               mol_type = other RNA
               organism = synthetic construct

SEQUENCE: 303
actctagct cacagatcgc gctatagttc cataaggaaag cttaagttac tatagttaagg 60
gcaatgaccc gtggcgttt gggatcgctt catccattac ggatattctc cccatgtgaa 120
aaggcacctaa gcataaggct aaggtgctt t 151

SEQ ID NO: 304      moltype = RNA  length = 105
FEATURE           Location/Qualifiers
misc_feature      1..105
note = source = /note="Description of Artificial Sequence:
                     Synthetic polynucleotide"
misc_feature      1..105
note = source = /note="CF R1162X nAPG01604 Target 1 sgRNA"
source            1..105
               mol_type = other RNA
               organism = synthetic construct

SEQUENCE: 304
actcagctca cagatcgcac gtttagtac tctgtaaaaa gttacagaat ctactaaaac 60
aaggcaaat gccgtgttta tctcgtcaac ttgttggcga gattt 105

SEQ ID NO: 305      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF W1282X nAPG09882 Target 1"
source            1..65
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 305
gtgtgtcttg ggattcaata actttgcaac agtgaaggaa agcctttgga gtgataaccac 60
agggtg 65

SEQ ID NO: 306      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF W1282X nAPG09882 Target 2"
source            1..65
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 306
gtcttggat tcaataactt tgcaacagtg aaggaaagcc tttggagtga taccacaggt 60
gagca 65

SEQ ID NO: 307      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF W1282X nAPG09882 Target 3"
source            1..65
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 307
cttgggattc aataactttg caaacgtgaa gggaaagcctt tggagtgata ccacagggtg 60
gcaaa 65

SEQ ID NO: 308      moltype = DNA  length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..65
note = source = /note="CF W1282X nAPG09882 Target 4"
source            1..65
               mol_type = other DNA
               organism = synthetic construct

SEQUENCE: 308
gggattcaat aactttgcaa cagtgaagga aaggcctttg agtgataccca caggtgagca 60

```

---

-continued

---

```

aaagg                                         65

SEQ ID NO: 309      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF W1282X nAPG09882 Target 5"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 309
gattcaataa ctttgcaaca gtgaggaaa gccttggag tgataccaca ggtgagcaa  60
aggac                                         65

SEQ ID NO: 310      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF W1282X nAPG06646 Target 1"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 310
tcgtatgggt gtcttggat tcaataactt tgcaacagtg aaggaaagcc tttggagtga  60
tacca                                         65

SEQ ID NO: 311      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF W1282X nAPG06646 Target 2"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 311
ttgggattca ataactttgc aacagtgaag gaaagcctt ggagtgatac cacaggttag  60
caaaa                                         65

SEQ ID NO: 312      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF W1282X nAPG06646 Target 3"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 312
tgggattcaa taactttgca acagtgaagg aaagccttg gagtgatacc acaggttagc  60
aaaaag                                         65

SEQ ID NO: 313      moltype = DNA  length = 65
FEATURE
misc_feature
1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature
1..65
note = source = /note="CF W1282X nAPG06646 Target 4"
source
1..65
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 313
ggattcaata actttgcaac agtgaaggaa agccttggaa gtgataccac aggtgagcaa  60
aagga                                         65

SEQ ID NO: 314      moltype = DNA  length = 60
FEATURE
misc_feature
1..60
note = source = /note="Description of Artificial Sequence:

```

---

-continued

---

```

misc_feature      Synthetic oligonucleotide"
1..60
source          note = source = /note="CF W1282X nAPG03850 Target 1"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 314
tgtcttggaa ttcaataact ttgcaacagt gaaggaaagc ctttggagtg ataccacagg 60

SEQ ID NO: 315      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
1..60
source          note = source = /note="CF W1282X nAPG03850 Target 2"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 315
gtcttggat tcaataactt tgcaacagtg aaggaaagcc tttggagtgta taccacaggt 60

SEQ ID NO: 316      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
1..60
note = source = /note="CF W1282X nAPG03850 Target 3"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 316
cttgggattc aataactttg caacagtgaa gcaaagcctt tggagtgata ccacaggtga 60

SEQ ID NO: 317      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
1..60
note = source = /note="CF W1282X nAPG03850 Target 4"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 317
tgggattcaa taactttgca acagtgaagg aaagccttg gagtgataacc acaggtgagc 60

SEQ ID NO: 318      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
1..60
note = source = /note="CF W1282X nAPG07553 Target 1"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 318
cttgggattc aataactttg caacagtgaa gcaaagcctt tggagtgata ccacaggtga 60

SEQ ID NO: 319      moltype = DNA length = 60
FEATURE           Location/Qualifiers
misc_feature      1..60
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
1..60
note = source = /note="CF W1282X nAPG07553 Target 2"
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 319
tgggattcaa taactttgca acagtgaagg aaagccttg gagtgataacc acaggtgagc 60

SEQ ID NO: 320      moltype = DNA length = 60
FEATURE           Location/Qualifiers

```

---

-continued

---

```

misc_feature          1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..60
note = source = /note="CF W1282X nAPG01604 Target 1"
source               1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 320
tcttggattt caataacttt gcaacagtga aggaaagcct ttggagtgtat accacaggta 60

SEQ ID NO: 321      moltype = DNA length = 60
FEATURE             Location/Qualifiers
misc_feature          1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..60
note = source = /note="CF W1282X nAPG01604 Target 2"
source               1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 321
cttggggatttca aataacttttgc caacagtgaag gaaagcctt tggagtgtata ccacaggta 60

SEQ ID NO: 322      moltype = DNA length = 65
FEATURE             Location/Qualifiers
misc_feature          1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..65
note = source = /note="CF W1282X nAPG07433.1 Target 1"
source               1..65
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 322
ttggggatttca ataacttttgc aacagtgaag gaaagcctt ggagtgtata cacaggtag 60
caaaaa              65

SEQ ID NO: 323      moltype = DNA length = 60
FEATURE             Location/Qualifiers
misc_feature          1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..60
note = source = /note="CF W1282X nAPG09748 Target 1"
source               1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 323
gtatcactcc aaaggctttc cttcactgtt gcaaaggat tgaatccaa gacacaccat 60

SEQ ID NO: 324      moltype = DNA length = 60
FEATURE             Location/Qualifiers
misc_feature          1..60
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..60
note = source = /note="CF W1282X nAPG05586 Target 1"
source               1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 324
gattcaataa ctttgcaaca gtgaaggaaa gccttggag tgataccaca ggtgagcaaa 60

SEQ ID NO: 325      moltype = DNA length = 25
FEATURE             Location/Qualifiers
misc_feature          1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..25
note = source = /note="CF W1282X nAPG09882 Target 1"
source               1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 325
actttgcaac agtgaaggaa agcct

```

---

-continued

---

```

SEQ ID NO: 326      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG09882 Target 2"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 326
tgcaacagtg aaggaaagcc tttgg                                25

SEQ ID NO: 327      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG09882 Target 3"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 327
caacagtgaa ggaaaggcctt tggag                                25

SEQ ID NO: 328      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG09882 Target 4"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 328
cagtgaagga aaggccttgg agtga                                25

SEQ ID NO: 329      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG09882 Target 5"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 329
gtgaaggaaa gcctttggag tgata                                25

SEQ ID NO: 330      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG06646 Target 1"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 330
tcaataactt tgcaacagtg aagga                                25

SEQ ID NO: 331      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG06646 Target 2"
1..25
mol_type = other DNA
organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 331
aacagtgaag gaaaggcctt ggagt 25

SEQ ID NO: 332      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF W1282X nAPG06646 Target 3"
source            1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 332
acagtgaagg aaaggccttg gagtg 25

SEQ ID NO: 333      moltype = DNA  length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="CF W1282X nAPG06646 Target 4"
source            1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 333
agtgaaggaa agcctttgga gtgtat 25

SEQ ID NO: 334      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="CF W1282X nAPG03850 Target 1"
source            1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 334
ttgcaacagt gaaggaaagc 20

SEQ ID NO: 335      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="CF W1282X nAPG03850 Target 2"
source            1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 335
tgcaacacgtg aaggaaagcc 20

SEQ ID NO: 336      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="CF W1282X nAPG03850 Target 3"
source            1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 336
caacagtcaa ggaaaggcctt 20

SEQ ID NO: 337      moltype = DNA  length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="CF W1282X nAPG03850 Target 4"
source            1..20

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 337
acagtgaagg aaagccttgc                                     20

SEQ ID NO: 338      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF W1282X nAPG07553 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 338
caaacagtgaa ggaaaggccttgc                                     20

SEQ ID NO: 339      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF W1282X nAPG07553 Target 2"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 339
acagtgaagg aaagccttgc                                     20

SEQ ID NO: 340      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF W1282X nAPG01604 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 340
gcaacagtga aggaaaggccttgc                                     20

SEQ ID NO: 341      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
note = source = /note="CF W1282X nAPG01604 Target 2"
1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 341
caaacagtgaa ggaaaggccttgc                                     20

SEQ ID NO: 342      moltype = DNA  length = 25
FEATURE
misc_feature
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="CF W1282X nAPG07433.1 Target 1"
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 342
aacagtgaag gaaaggccttt ggagt                                     25

SEQ ID NO: 343      moltype = DNA  length = 20
FEATURE
misc_feature
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20

```

---

-continued

---

```

source          note = source = /note="CF W1282X nAPG09748 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 343
cttcactgtt gcaaaggat 20

SEQ ID NO: 344      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..20
source          note = source = /note="CF W1282X nAPG05586 Target 1"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 344
gtgaaggaaa gcctttggag 20

SEQ ID NO: 345      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
source          note = source = /note="CF W1282X nAPG09882 Target 1 sgRNA"
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 345
actttgcAAC agtgaaggaa agcctgttt tgtaCTCTCA ataaaaAGTT attgagaATC 60
tacaaaATA aggcattttc ccgaatttac cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 346      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF W1282X nAPG09882 Target 2 sgRNA"
1..118
source          mol_type = other RNA
organism = synthetic construct

SEQUENCE: 346
tgcaaacagtg aaggaaAGCC tttgggttt tgtaCTCTCA ataaaaAGTT attgagaATC 60
tacaaaATA aggcattttc ccgaatttac cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 347      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF W1282X nAPG09882 Target 3 sgRNA"
1..118
source          mol_type = other RNA
organism = synthetic construct

SEQUENCE: 347
caacagtgaa ggaaAGCCTT tggaggttt tgtaCTCTCA ataaaaAGTT attgagaATC 60
tacaaaATA aggcattttc ccgaatttac cggcctacat atgttagggcg gttttttt 118

SEQ ID NO: 348      moltype = RNA length = 118
FEATURE           Location/Qualifiers
misc_feature      1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..118
note = source = /note="CF W1282X nAPG09882 Target 4 sgRNA"
1..118
source          mol_type = other RNA
organism = synthetic construct

SEQUENCE: 348
cagtgaaggaa aagcctttgg agtgaggttt tgtaCTCTCA ataaaaAGTT attgagaATC 60
tacaaaATA aggcattttc ccgaatttac cggcctacat atgttagggcg gttttttt 118

```

---

-continued

---

```

SEQ ID NO: 349      moltype = RNA length = 118
FEATURE
misc_feature
1..118
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..118
note = source = /note="CF W1282X nAPG09882 Target 5 sgRNA"
source
1..118
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 349
gtgaaggaaa gcctttggag ttagatgttt tgactctca ataaaaagtt attgagaatc 60
tacaaaaata aggcattttg ccgaattac cgcctacat atgttagggcg gttttttt 118

SEQ ID NO: 350      moltype = RNA length = 164
FEATURE
Location/Qualifiers
1..164
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF W1282X nAPG06646 Target 1 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 350
tcataatact tgcaaacagt aaggagccat aattcctctg taaaacttaa agaaggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atgcctatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 351      moltype = RNA length = 164
FEATURE
Location/Qualifiers
1..164
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF W1282X nAPG06646 Target 2 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 351
aaacagtgaag gaaaggcctt ggagtggcat aattcctctg taaaacttaa agaaggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atgcctatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 352      moltype = RNA length = 164
FEATURE
Location/Qualifiers
1..164
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF W1282X nAPG06646 Target 3 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 352
acagtgaagg aaaggccttgg gagggtggcat aattcctctg taaaacttaa agaaggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atgcctatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 353      moltype = RNA length = 164
FEATURE
Location/Qualifiers
1..164
misc_feature
1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..164
note = source = /note="CF W1282X nAPG06646 Target 4 sgRNA"
source
1..164
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 353
agtgaaggaa agcctttggatgtggcat aattcctctg taaaacttaa agaaggtta 60
tagagttatt atggtaaggc aatatgccgt ggcgttgggg atgcctatg tccggttta 120
ccggatctcc ctaaagggtga ctaactttgg tttagtcacct tttt 164

SEQ ID NO: 354      moltype = RNA length = 163
FEATURE
Location/Qualifiers

```

---

---

-continued

---

```

misc_feature          1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..163
note = source = /note="CF W1282X nAPG03850 Target 1 sgRNA"
source                1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 354
ttgcaacagt gaaaggaaagc gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaggggcaa tgaccctgtgg cgtttgggaa tcgccttatac ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt                  163

SEQ ID NO: 355      moltype = RNA length = 163
FEATURE              Location/Qualifiers
misc_feature          1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..163
note = source = /note="CF W1282X nAPG03850 Target 2 sgRNA"
source                1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 355
tgcaacagtg aaggaaagcc gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaggggcaa tgaccctgtgg cgtttgggaa tcgccttatac ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt                  163

SEQ ID NO: 356      moltype = RNA length = 163
FEATURE              Location/Qualifiers
misc_feature          1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..163
note = source = /note="CF W1282X nAPG03850 Target 3 sgRNA"
source                1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 356
caacagtgaa ggaaaggcctt gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaggggcaa tgaccctgtgg cgtttgggaa tcgccttatac ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt                  163

SEQ ID NO: 357      moltype = RNA length = 163
FEATURE              Location/Qualifiers
misc_feature          1..163
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..163
note = source = /note="CF W1282X nAPG03850 Target 4 sgRNA"
source                1..163
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 357
acagtgaagg aaaggcctt gctatagttc cataagaaaa aagtttctta agttactata 60
gtaaggggcaa tgaccctgtgg cgtttgggaa tcgccttatac ctggatgga tattctccc 120
atgtgaaaag cacctaagca tagcgctatg gtgcgtttat ttt                  163

SEQ ID NO: 358      moltype = RNA length = 151
FEATURE              Location/Qualifiers
misc_feature          1..151
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..151
note = source = /note="CF W1282X nAPG07553 Target 1 sgRNA"
source                1..151
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 358
caacagtgaa ggaaaggcctt gctatagttc cataagaaaa cttaaatgttac tatagtaagg 60
gcaatgaccc gtggcggtt gggatcgctt catccattac ggatattctc cccatgtgaa 120
aagcacctaa gcataaggct aagggtgc ttt                                151

SEQ ID NO: 359      moltype = RNA length = 151
FEATURE              Location/Qualifiers
misc_feature          1..151

```

---

---

-continued

---

```

        note = source = /note="Description of Artificial Sequence:
        Synthetic polynucleotide"
misc_feature
1..151
note = source = /note="CF W1282X nAPG07553 Target 2 sgRNA"
source
1..151
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 359
acagtgaagg aaagcctttg gctatagttc cataagaaa cttaaggtag tatagttggaa 60
gcaatgaccq gtggcggttg gggatcgctt catccattac ggatattctc cccatgtgaa 120
aagcacactaa gcataaggctt aagggtgttt t 151

SEQ ID NO: 360      moltype = RNA length = 105
FEATURE
misc_feature
1..105
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..105
note = source = /note="CF W1282X nAPG01604 Target 1 sgRNA"
source
1..105
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 360
gcaacagtga aggaaaggcctt gtttttagtac tctgtaaaaa gttacagaat ctactaaaac 60
aaggcaaat gcccgtttt aac 105

SEQ ID NO: 361      moltype = RNA length = 105
FEATURE
misc_feature
1..105
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..105
note = source = /note="CF W1282X nAPG01604 Target 2 sgRNA"
source
1..105
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 361
caacagtgaa ggaaaggcctt gtttttagtac tctgtaaaaa gttacagaat ctactaaaac 60
aaggcaaat gcccgtttt aac 105

SEQ ID NO: 362      moltype = RNA length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..135
note = source = /note="CF W1282X nAPG07433.1 Target 1 sgRNA"
source
1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 362
aacagtgaag gaaaggcctt ggagtgtcat agttccattaa aagccaaaag tggctttgat 60
gtttctatga taagggtttc gaccctggc gtcggggatc gcctgccccat tgaaatggc 120
ttctccccat ttatt 135

SEQ ID NO: 363      moltype = RNA length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..135
note = source = /note="CF W1282X nAPG09748 Target 1 sgRNA"
source
1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 363
cgacgggttag aggccgtatg tcgatttgc ttaatttcgtt gctgtgtcat tgcgtcctc 60
cattacaggg cggctaccac gaatagccac gaatgtttttt 120
ctgttgcaaa gttat 135

SEQ ID NO: 364      moltype = RNA length = 110
FEATURE
misc_feature
1..110
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..110

```

---

-continued

---

```

source          note = source = /note="CF W1282X nAPG05586 Target 1 sgRNA"
                1..110
                mol_type = other RNA
                organism = synthetic construct

SEQUENCE: 364
gtgaaggaaa gcctttggag gttattgtac tctcaataaa aagtatttga gaatctacaa  60
taataaggca ttcttgccgaa ttaccgccc tacatatgtt gggcggttt           110

SEQ ID NO: 365      moltype = AA length = 8
FEATURE          Location/Qualifiers
REGION           1..8
                note = source = /note="Description of Artificial Sequence:
                                Synthetic peptide"
REGION           1..8
                note = source = /note="APG07433.1 deleted motif"
source            1..8
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 365
LKKERNGA                                                 8

SEQ ID NO: 366      moltype = AA length = 1063
FEATURE          Location/Qualifiers
REGION           1..1063
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polypeptide"
REGION           1..1063
                note = source = /note="APG07433.1 engineered deletion"
source            1..1063
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 366
MRELDYRIGL DIGTNSIGWG VIELSWNKDR ERYEKVRIVD QGVRMFDRAE MPKTGASLAE  60
PERRIARSSRR RLNRKSRKQK NIRNLNVQHG VITQEELDSL YPLSKKSMIDI WGIRLDGLDR 120
LLNHFEWARL LIHHLAQRGRG KSNRKSELKD TETGKVLSLY RTVGEMWMKD 180
PDPSKYDRKR NSPNEYVFSV SRAELEKIV TLFAAQRRFQ SPYASKDLQE TYLQIWTHQL 240
PFASGNAILN KVGYCSLLKG KERRIPKATY TFQYFSALDQ VNRTRLGPDF QPFTKEQREI 300
ILNNMFQRTD YYKKKTipev TYYDINKWLE LDFTIQFKGL NYDPNEELKK IEKKPFINLK 360
AFYEINKVVA NYSSERTNETF STLDYDGIGY ALTVYKTDKD IRSYLKSSHN LPKRCYDQL 420
IEELLSLSYT KFGHSLKAI NHVLSIMQKG NTYKEAVDQL GYDTSGLKKE KRSKFLPPIS 480
DEITNPIVKR ALTQARKVVN AIIRRHSPEH SVHIELAREL SKNHDERTKI VSAQDENYKK 540
NIKGAISILSE HGILNPTGYD IVRYKLWKEQ GERCAYSLSKE IPADTFNNEP ILEVHDILPY 600
SQSFIDSYHN KVLVYSDENR KKGKNRIPVTY FLETNNDWEA FERYVRSNKF FSKKKREYLL 660
KRAYLPRESE LIKERHLNDT RYASTFLKNF IEQNLQFKEA EDNPRKRRVQ TVNGVITAHF 720
RKRWGLEKDR QETYLHHAMD AIIVACTDH MVTTRVTEYYQ IKESNKSVKK PYFPMPWEGF 780
RDELLSHLAS QPIAKKISEE LKAGYQSLDY IFVSRMPKRS ITGAAHQQT MRKGIDKKG 840
KTIIEERLHL KDIKFDENGD FKVMGKEQDM ATYEAIKQRY LEHGKNSKKA FETPLYKPSK 900
KGTCGNLKR VKEGQAQKSFV REVNGGVAQN GDLVRVDSLFE KDDKYYMVPY YVPDTVCSEL 960
PKKVVASSKG YEQWLTLDNS FTFKFSLYPY DLVRLVKGDE DRFLYFGTLD IDSDRLNFKD 1020
VNKPSKKNEY RYSLKTIEDL EKYEVGVLD LRLVRKETTR NFH                   1063

SEQ ID NO: 367      moltype = AA length = 8
FEATURE          Location/Qualifiers
REGION           1..8
                note = source = /note="Description of Artificial Sequence:
                                Synthetic peptide"
REGION           1..8
                note = source = /note="APG08290.1 deleted motif"
source            1..8
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 367
LKKERNGS                                                 8

SEQ ID NO: 368      moltype = AA length = 1064
FEATURE          Location/Qualifiers
REGION           1..1064
                note = source = /note="Description of Artificial Sequence:
                                Synthetic polypeptide"
REGION           1..1064
                note = source = /note="APG08290.1 engineered deletion"
source            1..1064
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 368
MSELDYRIGL DIGTNSIGWG VIELFWNKDR ERYEKVRIVD KGVRMFDKAE IPNKGASLAE  60

```

---

-continued

---

PRRIARSSRR	RLNRKSQRKK	EIRNLVQHG	MITQEELDLL	YPLSKKSIDI	WDIRLDGLDR	120
LINHLEWARL	LIHLAQRRGF	KSNRKSELKD	AETGKVLSII	QVNEKRLFLY	RTVGBMWIKD	180
AEFSKYDRRR	NSPNEYVFSV	SRADLEKEIV	TLFEAQRKFQ	SSYASKNLQE	TYLQIWAHQL	240
PFASGNAILN	KVGYSLLKG	KERRIPKATY	TFQYFSALDQ	VNRTRLGPDF	QPFTQEKEI	300
ILDKMFQRTD	YYKKKTIPEV	SYYDIRKWLE	LDETIQFKGL	NYDPNEELKK	IEKKPFINLK	360
AFYEIKKVA	NYAERTNEAP	STLDYDATAV	ALTVYKTDKF	IRSYLKKSNH	LSKRCYDDQL	420
IEEELFTLSYT	KFGHLSFKAI	NHVLPIMQEG	RTYQEAIHQI	GYDTTNLKE	NRSMFLPLIP	480
DEITNPPIVKR	AITQARKVVM	AIIRRYGSPN	SVHIELAREL	SKSHDERKKI	MTAHDENYKK	540
NKGAISILIE	NGILNPNTGYD	IVRYKLWKEQ	GERCAYSLSKE	IPPDTFFNEP	ILEVDHILPY	600
SQSFIDSYHN	KVLVSDENK	NKGNRIPYTY	FLETNKDWEA	FERVYRSNKKI	FSKKKREYLL	660
KKTYLPRESE	LIKERHLNDT	RYASTFLKNF	IEQNLQFKEV	EVNLRKRVQ	TVNGVITAHL	720
RKRWGLEKNR	QBTYLNHAMD	AIIAVCTDH	MVTRITEYYQ	IKESNKSVKI	PYFPMPWEGF	780
RDELLSHLAS	QPIAKKISEE	LKAGYQSSDY	IFVSRMPKRS	VTGAAHQDTI	RRKGIDKKG	840
KTIIIKRVR	KDIKFDENGI	FKMVGKEQDL	ATYEAIAKORY	LEHRKNSKKA	FETPLYKPSK	900
KGTGNLKLKV	KIEGQTKAFV	REVNGVVAQN	SDLVRVDSLGE	KDDKYYMVP	YVPDTVCSEL	960
PKKVKSGKG	YEQWLTLDNS	FTFKSSLYPY	DLVRLVKCNE	DRFLYFGTLD	IDSDRLNFKD	1020
VNKPSKQNEY	RYSLKTIENL	EKYEVGVLD	LRLVKQETR	IFNR		1064

---

SEQ ID NO: 369	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = source = /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
misc_feature	1..20	
	note = source = /note="SGN000139 target sequence"	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 369		
aggttttaat ggccccagcct		20
SEQ ID NO: 370	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = source = /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
misc_feature	1..20	
	note = source = /note="SGN000143 target sequence"	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 370		
catggcagta cattagagca		20
SEQ ID NO: 371	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = source = /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
misc_feature	1..20	
	note = source = /note="SGN000169 target sequence"	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 371		
cacatctcgaa gcaagacgtt		20
SEQ ID NO: 372	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = source = /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
misc_feature	1..20	
	note = source = /note="SGN000173 target sequence"	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 372		
cttctatacgcc ctcccttcccc		20
SEQ ID NO: 373	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = source = /note="Description of Artificial Sequence: Synthetic oligonucleotide"	
misc_feature	1..20	

---

---

-continued

---

```

source          note = source = /note="SGN000186 target sequence"
1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 373
ggacagtgcg catccctcg                                         20

SEQ ID NO: 374      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..20
note = source = /note="SGN000194 target sequence"
1..20
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 374
gccgcacagc attcaggctcg                                         20

SEQ ID NO: 375      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="SGN000926 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 375
agagccatca ccatcacatc cctaa                                         25

SEQ ID NO: 376      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="SGN000927 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 376
ggccaaaatc cagctgcctt ccttg                                         25

SEQ ID NO: 377      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="SGN000928 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 377
gtttctactc ttggcttaca accca                                         25

SEQ ID NO: 378      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature      1..25
note = source = /note="SGN000929 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 378
atctggaggg aacttacagc atatg                                         25

SEQ ID NO: 379      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:

```

---

-continued

---

```

misc_feature      Syntheticoligonucleotide"
1..25
source          note = source = /note="SGN000930 target sequence"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 379
gaacaactca aatggaaatg aatat                                25

SEQ ID NO: 380      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Syntheticoligonucleotide"
1..25
source          note = source = /note="SGN000931 target sequence"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 380
tcctgttcca tcaccatcaa aaaaa                                25

SEQ ID NO: 381      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Syntheticoligonucleotide"
1..25
note = source = /note="SGN000935 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 381
tgttggttac ctccctgccca ccacc                                25

SEQ ID NO: 382      moltype = DNA length = 25
FEATURE          Location/Qualifiers
misc_feature      1..25
note = source = /note="Description of Artificial Sequence:
Syntheticoligonucleotide"
1..25
note = source = /note="SGN001101 target sequence"
1..25
source          mol_type = other DNA
organism = synthetic construct

SEQUENCE: 382
atattttctt taatggtgcc aggca                                25

SEQ ID NO: 383      moltype = RNA length = 130
FEATURE          Location/Qualifiers
misc_feature      1..130
note = source = /note="Description of Artificial Sequence:
Syntheticpolynucleotide"
1..130
note = source = /note="SGN000139 sgRNA"
1..130
source          mol_type = other RNA
organism = synthetic construct

SEQUENCE: 383
aggtttaat ggcccagcct gtcatagttc cattaaagcc aaaagtggct ttgatgttcc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctg cccattgaaa tgggcttctc 120
cccatttatt                                              130

SEQ ID NO: 384      moltype = RNA length = 130
FEATURE          Location/Qualifiers
misc_feature      1..130
note = source = /note="Description of Artificial Sequence:
Syntheticpolynucleotide"
1..130
note = source = /note="SGN000143 sgRNA"
1..130
source          mol_type = other RNA
organism = synthetic construct

SEQUENCE: 384
catggcagta cattagagca gtcatagttc cattaaagcc aaaagtggct ttgatgttcc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctg cccattgaaa tgggcttctc 120

```

---

-continued

---

```

cccatttattt                               130

SEQ ID NO: 385      moltype = RNA  length = 130
FEATURE
misc_feature
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..130
note = source = /note="SGN000169 sgRNA"
source
1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 385
cacatctcg acaagacgtt gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgtcgg ggatcgcctg cccattgaaa tgggcttctc 120
cccatttattt                               130

SEQ ID NO: 386      moltype = RNA  length = 130
FEATURE
misc_feature
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..130
note = source = /note="SGN000173 sgRNA"
source
1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 386
cttctatagc ctccccc gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgtcgg ggatcgcctg cccattgaaa tgggcttctc 120
cccatttattt                               130

SEQ ID NO: 387      moltype = RNA  length = 130
FEATURE
misc_feature
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..130
note = source = /note="SGN000186 sgRNA"
source
1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 387
ggacagtgcg catctccctg gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgtcgg ggatcgcctg cccattgaaa tgggcttctc 120
cccatttattt                               130

SEQ ID NO: 388      moltype = RNA  length = 130
FEATURE
misc_feature
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..130
note = source = /note="SGN000194 sgRNA"
source
1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 388
ggcgcacagc attcaggatcg gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgtcgg ggatcgcctg cccattgaaa tgggcttctc 120
cccatttattt                               130

SEQ ID NO: 389      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..135
note = source = /note="SGN000926 sgRNA"
source
1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 389
agagccatca ccatcacatc cctaagtcat agttccatga aagccaaaag tggctttgat 60
gtttctatga taagggtttc ggcccgatgc gtcggggatc gcctgccccat tccgatggc 120
ttctcccat ttattt                               135

```

---

-continued

---

```

SEQ ID NO: 390      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="SGN000927 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 390
ggccaaaatc cagctgcctt ccttgtcat agttccatga aagccaaaag tggcttgat  60
gtttctatga taagggttcc ggcccggtgc gtcgggatc gcctgcccattccgatggc 120
ttctccccat ttatt                                135

SEQ ID NO: 391      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="SGN000928 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 391
gtttctactc ttggcttaca acccagtcat agttccatga aagccaaaag tggcttgat  60
gtttctatga taagggttcc ggcccggtgc gtcgggatc gcctgcccattccgatggc 120
ttctccccat ttatt                                135

SEQ ID NO: 392      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="SGN000929 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 392
atctggaggc aacttacagc atatggcat agttccatga aagccaaaag tggcttgat  60
gtttctatga taagggttcc ggcccggtgc gtcgggatc gcctgcccattccgatggc 120
ttctccccat ttatt                                135

SEQ ID NO: 393      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="SGN000930 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 393
gaacaactca aatggaaatg aatatgtcat agttccatga aagccaaaag tggcttgat  60
gtttctatga taagggttcc ggcccggtgc gtcgggatc gcctgcccattccgatggc 120
ttctccccat ttatt                                135

SEQ ID NO: 394      moltype = RNA  length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..135
note = source = /note="SGN000931 sgRNA"
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 394
tcctgttcca tcaccatcaa aaaaagtcat agttccatga aagccaaaag tggcttgat  60
gtttctatga taagggttcc ggcccggtgc gtcgggatc gcctgcccattccgatggc 120
ttctccccat ttatt                                135

```

-continued

---

```

SEQ ID NO: 395      moltype = RNA length = 135
FEATURE
misc_feature
1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..135
note = source = /note="SGN000935 sgRNA"
source
1..135
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 395
tgttggttac ctccctgcca ccaccgtcat agttccatga aagccaaaag tggctttgat 60
gtttctatga taagggttgc ggcccggtgc gtcggggatc gcctgccccat tccgatggc 120
tttccccat ttatt 135

SEQ ID NO: 396      moltype = RNA length = 130
FEATURE
Location/Qualifiers
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..130
note = source = /note="SGN001101 sgRNA"
source
1..130
mol_type = other RNA
organism = synthetic construct
SEQUENCE: 396
atattttttttaatggtgcc aggcaagtcat agttccattaa aagccaaaag tggctttgat 60
gtttctatga taagggttgc gaccgggtgc gtcggggatc gcctgccccat tgaaatggc 120
tttccccat 130

SEQ ID NO: 397      moltype = AA length = 1063
FEATURE
REGION
1..1063
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION
1..1063
note = source = /note="nAPG07433.1-del"
source
1..1063
mol_type = protein
organism = synthetic construct
SEQUENCE: 397
MRELDYRIGL AIGTNSIGWG VIELSWNKDR ERYEKVRIVD QGVRMFDRAE MPKTGASLAE 60
PERRIARSSRR RLNRKSQRKK NIRNLVQHG VITQEELDSL YPLSKKSMIDI WGIRLDGLDR 120
LLNHFEWARL LIHLAQRRGF KSNRKSELKD TETGKVLSI QLNKEKRLSLY RTVGEMWMKD 180
PDFSKYDRKR NSPNEYVFSV SRAELEKEIV TLFAAQRRFQ SPYASKDLQE TYLQIWTHQL 240
PFASGNAILN KVGYCSLLKG KERRIPKATY TFQYFSALDQ VNRTRLGPDF QPFTKEQREI 300
ILNNMfqRTLD EYKKTQFQKLY LDETIQFKGL NYDPNEELKK IEKKPFINLK 360
AFYEINKVVA NYSERTNETF STLDYDGIGY ALTVYKTDKD IRSYLKSSHN LPKRCYDDQL 420
IEELLSSLYT KFGHLSLKA NHVLSIMQKG NTYKEAVDQL GYDTSGLKKE KRSPKLPPIS 480
DEITNPPIVKR ALTQARKVNV AIIRRHSHPH SVHIELARE SKNHDERTKI VSAQDENYKK 540
NIKGAISLSE HGILNPTGYD IVRYKLWKBQ GERCAYSLKE IPADTFNNEP ILEVHDILPY 600
SQSFIDSYHN KVLYVSDENR KKGNRIPYTY FLETNKDWEA FERYVRSNKF FSKKKREYLL 660
KRAYLPRESE LIKERHLNDT RYASTFLKNF IEONLQFKEA EDNPNRKRKVQ TVNGVITAHF 720
RKRWGLEKDR QETYIHHAMD AIIVACTDHH MVTRVTEYYQ IKESNKSVKK PYFPMPWEGF 780
RDELLSHLAS QPIAKKISEER LKAGYQSLDY IFVSRMPKRS ITGAAHQQT MRKGIDKKG 840
KTIIEERLHL KDIKFDENGD FKMGVQKQDM ATYEAIKQRY LEHGKNSKKA FETPLYKPSK 900
KGIGGNLLIKRV KVEGQAKSFV REVNGVVAQN GDLVRLVDLFE KDDKYYMVPY YVPDTVCSEL 960
PKVVVASSKG YEQWLTLDSN FTFKFSLYPY DLVRLVKGD DRFLYFGTLD IDSDRLNFKD 1020
VNKPSSKNEY RYSLKTIEDL EKYEVGVLD LRLVRKETR NFH 1063

SEQ ID NO: 398      moltype = AA length = 1064
FEATURE
REGION
1..1064
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION
1..1064
note = source = /note="nAPG08290.1-del"
source
1..1064
mol_type = protein
organism = synthetic construct
SEQUENCE: 398
MSELDYRIGL AIGTNSIGWG VIELFWNKDR ERYEKVRIVD KGVRMFDKAE IPNKGASLAE 60
PERRIARSSRR RLNRKSQRKK EIRNLVQHG MITQEELLLL YPLSKKSIDI WDIRLDGLDR 120
LLNHLEWARL LIHLAQRRGF KSNRKSELKD AETGKVLSI QVNEKRLFLY RTVGEMWIKD 180
AEFSKYDRRR NSPNEYVFSV SRADLEKEIV TLFEAQRKFQ SSYASKNLQE TYLQIWAHQL 240
PFASGNAILN KVGYCSLLKG KERRIPKATY TFQYFSALDQ VNRTRLGPDF QPFTQEOKEI 300

```

---

-continued

---

ILDKMFQRTD YYKKKTIPEV SYYDIRKWLE LDETIQFKGL NYDPNEELKK IEKKPFINLK	360
AFYEIKKVA NYAERTNEAF STLDYDATAV ALTVYKTDKD IRSYLUKSNN LSKRCYDDQL	420
IEEFLTSYT KFGHLSFKAI NHVLPIMQEG RTYQEAIHQI GYDTTNLKE NRSMFLPLIP	480
DEITNPIVKR AITQARKVVA AIIRRYGSPN SVHIELAREL SKSHDERKKI MTAHDENYKK	540
NKGAISILIE NGILNPTGYD IVRYKLWKEQ GERCAYSLSKE IPPDTFFNEP ILEVDHILPY	600
SQSFIDSYHH KVLYSDENE NKGNRIPYTY FLETNKDWEA FERYVRSNKK FSKKKREYLL	660
KKTYLPRESE LIKERHLNDT RYASTFLKNF IEQNLQFKEV EVNLRKKRVQ TVNGVITAHL	720
RKRWGLEKRN QETYLHAMD AIIVACTDHH MVTRITEYYQ IKESNKSVKK PYFPMPWEGF	780
RDELLSHLAS QPIAKKISEE LKAGYQSSDY IFVSRMPKRS VTGAAHDQTI RRKGIDKKG	840
KTIIKRVRL KDIKFDENGE FKVMGKEQDL ATYEAIKORY LEHRKNSKKA FETPLYKPSK	900
KGTGNLKR V KIEGQTAKAF REVNGVQAQN SDLVRVDLFE KDDKYYMVPY YVPDTVSEL	960
PKKVKSGKG YEQWLTLDS TTFKSSLYPY DLVRLVKGNE DRFLYGTLD IDSDRNLNFKD	1020
VNKPSKQNEY RYSLKTIENTL EKYEVGVLD LRLVKQETRR IFNR	1064

SEQ ID NO: 399	moltype = AA length = 169
FEATURE	Location/Qualifiers
REGION	1..169
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..169
	note = source = /note="LPG50140 protein sequence"
source	1..169
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 399	
MSDLELNHEY WMRHALQAK RARDEGEVPV GAVLVLNNOV IGEGLWNRAIG LHDPTAHAEI	60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNSK RGAAGSLMN	120
LNYPGMNHQV QIIDGVLAPE CGSLLCDFYR MPRQVFNQQK AESTSINGD	169

SEQ ID NO: 400	moltype = AA length = 164
FEATURE	Location/Qualifiers
REGION	1..164
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..164
	note = source = /note="LPG50141 protein sequence"
source	1..164
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 400	
MSNPELTHEW WMRYALTTLAK RAREEGEVPV GAVLVLNNOV IGEGLWNRAIG LHDPTAHAEI	60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLMN	120
LNYPGMNHRI EFTEGVLRDE CAAMLCDFYR QPRQVFNALK TGNA	164

SEQ ID NO: 401	moltype = AA length = 169
FEATURE	Location/Qualifiers
REGION	1..169
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..169
	note = source = /note="LPG50142 protein sequence"
source	1..169
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 401	
MSIPELNHDW WMRHALTLAK RAREEGEVPV GAVLVLNQV IGEGLWNRAIG LHDPTAHAEI	60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLINV	120
LNYPGMNHRV AITEGVLRREE CAAMLCDFYR QPRQVFNALK KPAGDINAF	169

SEQ ID NO: 402	moltype = AA length = 172
FEATURE	Location/Qualifiers
REGION	1..172
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..172
	note = source = /note="LPG50143 protein sequence"
source	1..172
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 402	
MSNPELNHEY WMRYALTTLAK RARDEGEVPV GAVLVLDQV IGEGLWNRAIG LHDPTAHAEI	60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG RLVFGVRNSK RGAAGSLNV	120
LNYPGMNHHI EMEEGVLRDE CAAMLCDFYR QPRQVFNALK KSPPDINNLQ AR	172

SEQ ID NO: 403	moltype = AA length = 169
FEATURE	Location/Qualifiers

---

-continued

---

```

REGION          1..169
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION          1..169
note = source = /note="LPG50144 protein sequence"
source          1..169
mol_type = protein
organism = synthetic construct
SEQUENCE: 403
MSNPELTHDH WMRHALTLAQ RARNEGEVPV GAVLVLNGQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTVL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLINV 120
LNYPGMNHVR EIIEGVLRDE CAAMLCDFYR HPRQVFNALK KNAGTINTQ 169

SEQ ID NO: 404      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50145 protein sequence"
source           1..166
mol_type = protein
organism = synthetic construct
SEQUENCE: 404
MSDTELNHEY WMRHALMLAK RARDEGEVPV GAVLVLKNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG NLVFGVRNSK RGAAGSLINV 120
LNYPGMNHVR EIAEGVLADE CSAMLCDFYR HPRQVFNALK QAAKHI 166

SEQ ID NO: 405      moltype = AA length = 171
FEATURE          Location/Qualifiers
REGION           1..171
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..171
note = source = /note="LPG50146 protein sequence"
source           1..171
mol_type = protein
organism = synthetic construct
SEQUENCE: 405
MSDIELNHEY WMRHALMLAK RAREEGEVPV GAVLVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG HLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRI EFTEGVLADE CSGMLCDFYR YPRQVFNTLK QAAKAINPAA Q 171

SEQ ID NO: 406      moltype = AA length = 173
FEATURE          Location/Qualifiers
REGION           1..173
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..173
note = source = /note="LPG50147 protein sequence"
source           1..173
mol_type = protein
organism = synthetic construct
SEQUENCE: 406
MSIPELNHDW WMRHALTLAK RAREEGEVPV GAVLVLNGQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLMN 120
LNYPGMNHVR EITEGVLRDE CAAMLCDFYR QPRQVFNALK KPAGDINALQ NNR 173

SEQ ID NO: 407      moltype = AA length = 168
FEATURE          Location/Qualifiers
REGION           1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..168
note = source = /note="LPG50148 protein sequence"
source           1..168
mol_type = protein
organism = synthetic construct
SEQUENCE: 407
MSNPEFTHEY WMRHALTLAR RARDEGEVPV GAVLVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNSK RGAAGSLMN 120
LNYPGMNHQV KTIGGVLAPE CSGLLCDFYR MPRQVFNQQK AELKSIND 168

SEQ ID NO: 408      moltype = AA length = 167
FEATURE          Location/Qualifiers
REGION           1..167

```

---

---

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..167
note = source = /note="LPG50149 protein sequence"
1..167
mol_type = protein
organism = synthetic construct
SEQUENCE: 408
MSDAELTHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG RLIFGVRNSK RGAAGSLINV 120
LNYPGMNHVRV EVVEGILRDE CAGMLCDFYR QPRAVKNALK KGATDIN 167

SEQ ID NO: 409      moltype = AA length = 167
FEATURE          Location/Qualifiers
REGION           1..167
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..167
note = source = /note="LPG50150 protein sequence"
1..167
mol_type = protein
organism = synthetic construct
SEQUENCE: 409
MSDAELTHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG RLIFGVRNSK RGAAGSLINV 120
LNYPGMNHVRV EVVEGILRDE CAGMLCAFYR QPRAVKNALK KGATDVL 167

SEQ ID NO: 410      moltype = AA length = 169
FEATURE          Location/Qualifiers
REGION           1..169
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..169
note = source = /note="LPG50151 protein sequence"
1..169
mol_type = protein
organism = synthetic construct
SEQUENCE: 410
MSDLELNHEY WMRHALQLAQ RARDEGEVPV GAVLVYNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLDDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNEK RGAAGSLMV 120
LRYPGMNHQV QIIDGVLAPE CSGLLCDFYR MPRQQKNQQK AESTSSPGD 169

SEQ ID NO: 411      moltype = AA length = 167
FEATURE          Location/Qualifiers
REGION           1..167
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..167
note = source = /note="LPG50152 protein sequence"
1..167
mol_type = protein
organism = synthetic construct
SEQUENCE: 411
MSDNELNHEY WMRHALGLAK RAREEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLDDTTL YVTFEPCVMC AGAMVHSRIG TLVFGVRNSK RGAAGSLMV 120
LNYPGMNHVRV EIVEGILSES CAAMLCDFYR QPRAVKNALK KAADPA 167

SEQ ID NO: 412      moltype = AA length = 164
FEATURE          Location/Qualifiers
REGION           1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..164
note = source = /note="LPG50153 protein sequence"
1..164
mol_type = protein
organism = synthetic construct
SEQUENCE: 412
MSDTEFTHEH WMRHALTLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVQ QNYRLDDTTL YVTFEPCVMC AGAMVHSRIG HLVFGVRNSK RGAIGSLMV 120
LGYPGMNHQV QVSEGVLATE CSAMLCDFYR APRLVKNALK EKAR 164

SEQ ID NO: 413      moltype = AA length = 171
FEATURE          Location/Qualifiers
REGION           1..171
note = source = /note="Description of Artificial Sequence:

```

---

-continued

---

```

REGION           Synthetic polypeptide"
1..171
source          note = source = /note="LPG50154 protein sequence"
1..171
mol_type = protein
organism = synthetic construct

SEQUENCE: 413
MSSESEFTHEH WMRHALTLAR RAREEGEVPV GAVLVVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDSTL YVTFEPCVMC AGAMVHGRIG NLVFGVRNSK RGAIGSLMVN 120
VGYPGMNHQI NVIEGVLAEE CSAMLCDFYR APRLVKNALK EKARNGNNPN K 171

SEQ ID NO: 414      moltype = AA length = 164
FEATURE          Location/Qualifiers
REGION           1..164
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..164
note = source = /note="LPG50155 protein sequence"
1..164
mol_type = protein
organism = synthetic construct

SEQUENCE: 414
MSNPELTTHEH WMRYALTAK RAREEGEVPV GAVLVVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLMVN 120
LNYPGMNHRI EFTEGVLRDE CAAMLCDFYR QPRLVKNALK TGNA 164

SEQ ID NO: 415      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50156 protein sequence"
1..166
mol_type = protein
organism = synthetic construct

SEQUENCE: 415
MSDPELNHEH WMRHALQLAK RAREEGEVPV GAVLVVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDTTL YVTFEPCVMC SGAMIHSRIG TVVFGVRNEK RGAAGSLNV 120
LYRPGMNHQV NVLGGVLAPA CSEMLCEFYR MPRQQKNRQK AESKLS 166

SEQ ID NO: 416      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50157 protein sequence"
1..166
mol_type = protein
organism = synthetic construct

SEQUENCE: 416
MSDNELNHEH WMRHALTLAQ RAREEGEVPV GAVLVLQNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGMVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRV EITEGVLAAD CSSMLCDFYR HPREQKNALK RAAHSN 166

SEQ ID NO: 417      moltype = AA length = 168
FEATURE          Location/Qualifiers
REGION           1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..168
note = source = /note="LPG50158 protein sequence"
1..168
mol_type = protein
organism = synthetic construct

SEQUENCE: 417
MSNPEHNHEH WMRHALTLAQ RARDEGEVPV GAVLVYNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNEK RGAAGSLMVN 120
LGYPGMNHQV QTIGGVLAPE CSGLLCDFYR MPRQQKNQQK AELNQPGD 168

SEQ ID NO: 418      moltype = AA length = 168
FEATURE          Location/Qualifiers
REGION           1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"

```

---

-continued

---

```

REGION          1..168
source          note = source = /note="LPG50159 protein sequence"
                1..168
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 418
MSDLELNHEY WMRHALSLAK RARDEGEVPV GAVLVLNNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDDTL YVTFEPCVMC SGAMVHSRIG TLVYGVRSNEK RGAAGSLMN 120
LNYPGMNHQV QIIGGVLAPD CSGLLCDFYR MPRQQKNQQK AELKSSGD 168

SEQ ID NO: 419      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50160 protein sequence"
source          1..166
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 419
MSDHEFNDYE WMRHALTLAK RAREEGEVPV GAVLVLNNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDDTL YVTFEPCVMC AGAMVHSRIS RLVFGVRNSK RGAAGSLIN 120
LNYPGMNHRV EITEGILAES CSAMLCDFYR WPREVKNALK KARQEE 166

SEQ ID NO: 420      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50161 protein sequence"
source          1..166
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 420
MSQTELTHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDDTL YVTFEPCVMC AGAMVHGRIG TLVFGVRNSK RGAVGSLMNI 120
TGYPGMNHQV QVIEGILATE CSAMLCAFYR QPRLVKNALK EAAKTA 166

SEQ ID NO: 421      moltype = AA length = 167
FEATURE          Location/Qualifiers
REGION           1..167
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..167
note = source = /note="LPG50162 protein sequence"
source          1..167
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 421
MSNPELNHDY WMRHALSLAK RAREEGEVPV GAVLVRNNEV IGEGNRAIG LHDPTAHAEI 60
MALRQGGMVL QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLMN 120
LNYPGMNHRV EIVEGVLRDE CAGMLCDFYR QPRLVKNAQK KGAEPLI 167

SEQ ID NO: 422      moltype = AA length = 172
FEATURE          Location/Qualifiers
REGION           1..172
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..172
note = source = /note="LPG50163 protein sequence"
source          1..172
                mol_type = protein
                organism = synthetic construct
SEQUENCE: 422
MSNPELNHEY WMRYALTAK RARDEGEVPV GAVLVYNDQV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG RLVFGVRNSK RGAAGSLLN 120
LNYPGMNHHI EMEEGVLRDE CAAMLCDFYR QPBMVKNALK KSPPDSPNLQ AR 172

SEQ ID NO: 423      moltype = AA length = 168
FEATURE          Location/Qualifiers
REGION           1..168
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..168

```

---

---

-continued

---

```

source          note = source = /note="LPG50164 protein sequence"
               1..168
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 423
MSNPEFTHEY WMRHALTLAR RARDEGEVPV GAVLVLNNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNEK RGAAGSLMNV 120
LGYPGMNHQV KTIGGVLAPE CSGLLCDFYR MPRQQKNQOK AELKSSGD           168

SEQ ID NO: 424      moltype = AA length = 165
FEATURE          Location/Qualifiers
REGION           1..165
               note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
               1..165
               note = source = /note="LPG50165 protein sequence"
source            1..165
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 424
MSDNEFNHEY WMRHALTLAQ RARDEGEVPV GAVLVLDNQV IGEGNRAIG LHDPTAHAEI 60
MALRQGGMVL QNYRLINATL YVTFEPCVMC AGAMVHSRIG HVVFGVRNSK RGAAGSLMNV 120
LNYPGMNHRV EVTEGVLRQ CAGMLCDFYR EPREQFNALR KAQKA             165

SEQ ID NO: 425      moltype = AA length = 170
FEATURE          Location/Qualifiers
REGION           1..170
               note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
               1..170
               note = source = /note="LPG50166 protein sequence"
source            1..170
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 425
MSDNEFNHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGMVL QNYRLIDATL YVTFEPCIMC AGAMVHSRIG QVVFGVRNSK RGAAGSLINI 120
LNYPGMNHRV DVTEGVLSER CANMLCDFYR EPRLQFNAAQ KAEKAGNAAA          170

SEQ ID NO: 426      moltype = AA length = 169
FEATURE          Location/Qualifiers
REGION           1..169
               note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
               1..169
               note = source = /note="LPG50167 protein sequence"
source            1..169
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 426
MSNPELTHDH WMRHALTLAQ RARNEGEVPV GAVLVLNQV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDTVL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRV EIIEGVLRDE CAAMLCDFYR HPRLVKNALK KNAGTSPQ           169

SEQ ID NO: 427      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
               note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
               1..166
               note = source = /note="LPG50168 protein sequence"
source            1..166
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 427
MSDTELNHEY WMRHALMLAK RARDEGEVPV GAVLVLKNOV IGEGNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG NLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRV EIAEGVLADE CSAMLCDFYR HPRQQQNALK QAAKHD           166

SEQ ID NO: 428      moltype = AA length = 171
FEATURE          Location/Qualifiers
REGION           1..171
               note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
               1..171
               note = source = /note="LPG50169 protein sequence"

```

---

---

-continued

---

```

source          1..171
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 428
MSDIELNHEY WMRHALMLAK RAREEGEVPV GAVLVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG HLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRI EFTEGVLADE CSGMLCDFYR YPRQQQNTLK QAAKANPPAA Q      171

SEQ ID NO: 429      moltype = AA length = 165
FEATURE          Location/Qualifiers
REGION           1..165
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..165
note = source = /note="LPG50170 protein sequence"
source            1..165
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 429
MSDNELNHER WMRHALTLAQ RARDEGEVPV GAVLVYQNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRV AITEGVLAES CSAMLCDFYR HPREQKNALR RAAQS             165

SEQ ID NO: 430      moltype = AA length = 166
FEATURE          Location/Qualifiers
REGION           1..166
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..166
note = source = /note="LPG50171 protein sequence"
source            1..166
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 430
MSDLELNDEY WMRHALTLAK RAREEGEVPV GAVLVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLIDATL YVTFEPCVMC AGAMVHSRIA RLVFGVRNSK RGAAGSLMVN 120
LNYPGMNHRV EISEGVLAES CSAMLCDFYR WPREVKNALK KAREQN             166

SEQ ID NO: 431      moltype = AA length = 169
FEATURE          Location/Qualifiers
REGION           1..169
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..169
note = source = /note="LPG50172 protein sequence"
source            1..169
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 431
MSDLELDHEY WMRHALLLAK RARDEGEVPV GAVLVLNNQV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLLDTTL YVTFEPCVMC SGAMVHSRIG TLVYGVRENK RGAAGSLMVN 120
LGYPGMNHQV QVIDGVLAPE CSGLLCDFYR MPRQQKNQQK AESTSSRGD             169

SEQ ID NO: 432      moltype = AA length = 162
FEATURE          Location/Qualifiers
REGION           1..162
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..162
note = source = /note="LPG50173 protein sequence"
source            1..162
               mol_type = protein
               organism = synthetic construct
SEQUENCE: 432
MSDTELTHEY WMRHALMLAQ RARDEGEVPV GAVLVLNNRV IGEAWNRAIG LHDPTAHAEI 60
MALRQGLLVL QNYRLLDTTL YVTFEPCVMC AGAMVHGRIG TLVFGVRNLK RGAAGSLMVN 120
LNYPGMNHRV EIVEGTLSDL CSGMLCEFYR QPRLAFNAQK QA      162

SEQ ID NO: 433      moltype = AA length = 173
FEATURE          Location/Qualifiers
REGION           1..173
note = source = /note="Description of Artificial Sequence:
                  Synthetic polypeptide"
REGION           1..173
note = source = /note="LPG50174 protein sequence"
source            1..173

```

---

---

-continued

---

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 433
MSIPELNHDV WMRHALTLAK RAREEGEVPV GAVLVLNGQV IGEGNRAIG LHDPTAHEAI 60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG QLVFGVRNSK RGAAGSLMN 120
LNYPGMNHRV EITEGVLRDE CAAMLCDFYR QPRLVKNALQ KPAGDPSALQ NNR 173

SEQ ID NO: 434      moltype = AA length = 166
FEATURE           Location/Qualifiers
REGION            1..166
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..166
note = source = /note="LPG50175 protein sequence"
1..166
mol_type = protein
organism = synthetic construct

SEQUENCE: 434
MSDIEQNHEY WMRHALVLAK RAREEGEVPV GAVLVLNNQV IGEGNRAIG LHDPTAHEAI 60
MALRQGGLVL QNYRLIDATL YVTFEPCVMC AGAMVHSRIA RLVFGVRNSK RGAAGSLMN 120
LNYPGMNHRV EISEGVLAGS CSAMLCDFYR WPREVKNALK KAREQN 166

SEQ ID NO: 435      moltype = AA length = 153
FEATURE           Location/Qualifiers
REGION            1..153
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..153
note = source = /note="LPG50176 protein sequence"
1..153
mol_type = protein
organism = synthetic construct

SEQUENCE: 435
MSDIEQNHEY WMRHALVLAK RAREEGEVPV GAVLVLNNQV IGEGNRAIG LHDPTAHEAI 60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHGRIG SLVFGVRNSK RGAAGSLINV 120
LNYPGMNHRV EMTEGVLADE CSAMLCDFYR HPR 153

SEQ ID NO: 436      moltype = AA length = 168
FEATURE           Location/Qualifiers
REGION            1..168
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..168
note = source = /note="LPG50177 protein sequence"
1..168
mol_type = protein
organism = synthetic construct

SEQUENCE: 436
MCNPDRHEY WMRHALTLAQ RARDEGEVPV GAVLVLNNQV IGEGNRAIG LHDPTAHEAI 60
MALRQGGMVL QNYRLLDTTL YVTFEPCVMC SGAMVHSRIG TLVFGVRNEK RGAAGSLNV 120
LGYPGMNHQV KTIGGVLAPA CSALLCDFYR MPRQQKNQQK AELKLSND 168

SEQ ID NO: 437      moltype = AA length = 162
FEATURE           Location/Qualifiers
REGION            1..162
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..162
note = source = /note="LPG50178 protein sequence"
1..162
mol_type = protein
organism = synthetic construct

SEQUENCE: 437
MSAIELNHEY WMRHALGLAQ RARDEGEVPV GAVLVYQNOV IGEGNRAIG LHDPTAHEAI 60
MALRQGGLVL QNYRLIDTTL YVTFEPCVMC AGAMVHSRIG RVVFGVRNSK RGAAGSLMN 120
LNYPGMNHRV EVTEGVLAGE CSAMLCDFYR APRAQFNAQK RP 162

SEQ ID NO: 438      moltype = AA length = 169
FEATURE           Location/Qualifiers
REGION            1..169
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..169
note = source = /note="LPG50179 protein sequence"
1..169
mol_type = protein

```

---

---

-continued

---

```

SEQUENCE: 438          organism = synthetic construct
MSNPELNHEY WMRYALTTLAK RAREEGEVPV GAVLVLNTERV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGMVL QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG HLVFGVRNSK RGAAGSLMN 120
LNYPGMNHVR AITEGVLRDE CAAMLCDFYR QPRQVKNALK KTLSDSSEQ 169

SEQ ID NO: 439          moltype = AA length = 168
FEATURE
REGION
1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..168
note = source = /note="LPG50180 protein sequence"
1..168
mol_type = protein
organism = synthetic construct
SEQUENCE: 439
MSNPEHDHEY WMRHALNLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDTTL YVTFEPCVMC SGAMIHSRIG TLVYGVRENK RGAAGSLMN 120
LGYPGMNHQV NVIGGVLAQD CSARLCDFYR MPRQQKNQQR AELKAQGD 168

SEQ ID NO: 440          moltype = AA length = 168
FEATURE
REGION
1..168
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..168
note = source = /note="LPG50181 protein sequence"
1..168
mol_type = protein
organism = synthetic construct
SEQUENCE: 440
MSNPELNHEY WMRHALQLAQ RARDEGEVPV GAVLVLNNQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLLDTTL YVTFEPCVMC SGAMIHSRIG TVVYGVRENK RGAAGSLMN 120
LSYPGMNHQV KVIGEVVLAPA CSAMLCDFYR MPRQQKNQQR AEWKLSGE 168

SEQ ID NO: 441          moltype = AA length = 171
FEATURE
REGION
1..171
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..171
note = source = /note="LPG50182 protein sequence"
1..171
mol_type = protein
organism = synthetic construct
SEQUENCE: 441
MSNPELNHEY WMRYALTTLAK RARDEGEVPV GAVLVYHDQV IGEGLWNRAIG LHDPTAHAEI 60
MALRQGGLVL QNYRLIDDTL YVTFEPCVMC AGAMVHSRIG RLVFGVRNSK RGAAGSLMN 120
LNYPGMNHQI DMEEGVLRDE CAAMLCDFYR LPRIVKNALQ QSPPDSTNLH A 171

SEQ ID NO: 442          moltype = AA length = 32
FEATURE
REGION
1..32
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
1..32
note = source = /note="L32 Linker sequence"
1..32
mol_type = protein
organism = synthetic construct
SEQUENCE: 442
SGSSSSGGSSG SETPGTSESA TPESGGSSG GS 32

SEQ ID NO: 443          moltype = DNA length = 507
FEATURE
misc_feature
1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..507
note = source = /note="Mammalian codon optimized LPG50140"
1..507
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 443
atgtctgatc tggactgaa tcacgagtagc tggatgcggc acggccctgca actggccaag 60

```

---

-continued

---

cgggccagag atgaggcgaa	ggtgccagtgc ggccgcgtgc	tggtgtgaa caaccaggtc	120
atcgagaag gttggaaacag	ggccatcgcc ctgcattgacc	ccacagccca tgccgaaatc	180
atggccctga gacaggccgg	actgggtctg cagaactata	gggttggatgaa caccacccgt	240
tacggtacct tcgagccttg	tgtgtgtgc tccggcgcta	tggtcacag cagaatccgc	300
acactggctt ttggcgtag	aaacagcaag cgccggagctg	ctggcagctt gatgaatgtg	360
ctgaactacc cccggcatgaa	ccaccagggtg caaatcatcg	acggcgctgt ccggccctgaa	420
tgccggcggac tgctgtcgaa	cttctaccgg atgcctagac	agggttcaa ccagcagaaa	480
gccggagagca cctctatcaa	cgccgac		507

SEQ ID NO: 444	moltype = DNA length = 492		
FEATURE	Location/Qualifiers		
misc_feature	1..492		
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"		
misc_feature	1..492		
	note = source = /note="Mammalian codon optimized LPG50141"		
source	1..492		
	mol_type = other DNA		
	organism = synthetic construct		
SEQUENCE: 444			
atgtccaacc ccgagctgac	ccacgagcac tggatgagat	acggccctgac actggccaaag	60
cggggccagag aggaaggcgaa	agtggccagtgc ggccgcgtgc	tggtgtgaa caaccagggtg	120
atcgagaagat gctggaaatag	ggccatggaa ctgcattgacc	ctacagccca cgccgaaatc	180
atggccctgaa gacaggccgg	cctgggtctg cagaactata	gactgtatcgaa caccacccgt	240
tacggtacat tcgagccttg	tgtgtgtgc gccggcgccca	tggtcacag cagaatccgc	300
caagctggctt ttggcgtagc	aaacagcaaa cccggccgtgt	caggctctt gatgaatgtg	360
ctcaactacc cccggcatgaa	ccacacatcg ggttccaccc	aggggagtctt gcggggacgag	420
tgccgtcgta tgctgtcgaa	cttctaccgg cagcctagac	agggttcaa ccggccctgaaag	480
accggcaacg cc			492

SEQ ID NO: 445	moltype = DNA length = 507		
FEATURE	Location/Qualifiers		
misc_feature	1..507		
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"		
misc_feature	1..507		
	note = source = /note="Mammalian codon optimized LPG50142"		
source	1..507		
	mol_type = other DNA		
	organism = synthetic construct		
SEQUENCE: 445			
atgagcatcc ccgagctgaa	ccacgacgtg tggatgcggc	acgctttac actggccaaag	60
cggggccagag aagaaggcgaa	agtggccagtgc ggccgcgtgc	tggtgtgaa cggccagggtg	120
atcgccgagg gctggacacag	ggccatggaa ctgcattgacc	ctaccgcaca cgccgagatc	180
atggccctgaa gacaggccgg	actgggtctg cagaactata	ggctgtatcgaa caccacccgt	240
tacggtacat tcgagccttg	cgtgtgtgc gccggcgccca	tggtcacag cagaatccgc	300
caagctgggtt tcggcgtagc	aaacagcaaa aggcccgtgt	cggatctt gatcaacgtg	360
ctgaattacc cccggcatgaa	ccatagatcg gccatcaccc	aggggagtgt cagagaggaa	420
tgtgcccoca tgctgtcgaa	cttctacaga caacctagac	aggctttaa ccggccctgaaag	480
aaacactgtcg gcgatatacaa	tgccttc		507

SEQ ID NO: 446	moltype = DNA length = 516		
FEATURE	Location/Qualifiers		
misc_feature	1..516		
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"		
misc_feature	1..516		
	note = source = /note="Mammalian codon optimized LPG50143"		
source	1..516		
	mol_type = other DNA		
	organism = synthetic construct		
SEQUENCE: 446			
atgagcaacc ccgagctgaa	tcacgagttac tggatgcggt	acggccctgac actggccaaag	60
cggggccagag atgaaaggcgaa	agtggctgtg ggccgcgtgc	tggtgtgaa cgaccagggtg	120
atcgggagaat gctggaaatag	ggccatggaa ctgcattgacc	ccacacccca cgccgagatc	180
atggccctgaa gacaggccgg	actgggtctg cagaactata	ggctgtatcgaa caccacccgt	240
tacggtacat tcgagccttg	tgtgtgtgc gccggcgccca	tggtcacatc tagaaatccgc	300
agactgggtt tcggcgtagc	aaacagcaag aggcccgtgt	cggcagctt gctgaacgtg	360
ctcaattacc ctggaaatgaa	ccacacatcg ggttccaccc	aggggcgctgt gcggggacgag	420
tgccgtcgta tgctgtcgaa	cttctacaga cagcctagac	aggctttaa ccggccctgaaag	480
aaatccccac ctgatatacaa	caacctgcaatcgaa	gttaga	516

SEQ ID NO: 447	moltype = DNA length = 507
FEATURE	Location/Qualifiers
misc_feature	1..507
	note = source = /note="Description of Artificial Sequence"

-continued

---

```

misc_feature      Synthetic polynucleotide"
1..507
note = source = /note="Mammalian codon optimized LPG50144"
source           1..507
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 447
atgagcaacc ctgagctgac acacgaccac tggatgagac acgcctgac cctggccag 60
cgggccagaa acgaggcgaa agtgcctgtg ggccgtgtc ttgtgtgaa tggccaagtg 120
atcggagaag gctggaaacg agccatcgcc ctgcatgacc caacagccca cgccgagatc 180
atggccctga ggcaggcgcc actggcttc cagaactata ggctgtatcga caccgtgtc 240
tacgatgttgc tccagcccttg tttgtatgtc gcggccgtca ttgtgtactc tagaatcgga 300
cagctgttct ttggcggtcg gaatagcaag cggggcgccg ctggctccct gatcaacgtg 360
cttaattacc cccggcatgaa ccacagatgtc gaaattatcg agggcgatct gagagatgag 420
tgcgcagcta tgctgtgcga cttctacaga catcctagac aagtgttcaa cgcctgaaa 480
aaacaacgccc gaaccatcaa caccccg 507

SEQ ID NO: 448      moltype = DNA length = 498
FEATURE             Location/Qualifiers
misc_feature        1..498
note = source = /note="Description of Artificial Sequence:
                     Synthetic polynucleotide"
1..498
note = source = /note="Mammalian codon optimized LPG50145"
source           1..498
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 448
atgagcgaca ccgagctgaa ccacgagtac tggatgcggc acgcctgtat gctggctaa 60
cggggccagag atgaggcgaa agtgcctgtg ggccgtgtc ttgtgttcaa aaaccaggatg 120
atcggagaatg gctggaaatg agccatcgcc ctgcatgacc ccacccgtca cgctgaaatc 180
atggccctga gacagggggg cctgtgtc cagaactata gactgtatcga taccacactg 240
tacgatgtatcc tccagcccttg tttgtatgtc gcggccgtca ttgtgtactc tagaatcggtc 300
aacctgttct ttggcggtcg gaacagcaag agggggcgctg ctggcgatct gatcaacgtg 360
ctgaattacc cccggcatgaa ccacagatgtc gaaatcgccg agggcgatct ggccqacgag 420
tgcctccgca tgctgtgcga cttctacccg catcctagac aagtgttcaa cgcctgaaa 480
caggccgcca agcacatc 498

SEQ ID NO: 449      moltype = DNA length = 513
FEATURE             Location/Qualifiers
misc_feature        1..513
note = source = /note="Description of Artificial Sequence:
                     Synthetic polynucleotide"
1..513
note = source = /note="Mammalian codon optimized LPG50146"
source           1..513
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 449
atgagcgaca tcgagctgaa tcacgagtac tggatgcggc acgcctgtat gctggccaaa 60
agagccaaag aggaaaggaga agtgcctgtg ggccgtgtc ttgtgtgaa caaccaggatg 120
atcggcgaaacg gctggaaaccc ggccatggc ctgcatgacc ctaccgtccca cgccgagatc 180
atggccctga gacagggggg actggcttc cagaactata gactgtatcga cacaacactg 240
tacgatgttgc tcgagcccttg tttgtatgtc gcggccgtca ttgtgtactc tagaatcggtc 300
cacctgttct ttggcgatccaa aactctaa cggggcgatct ctggctccct gatcaatgtg 360
ctgaactacc cccggcatgaa ccacccgtca gaattcaccgg agggcgatgtc ggctgtatgaa 420
tgcagccggca tgctgtgcga cttctacaga taccctagac aagtgttcaa caccctgaaa 480
caggccgcca aggccatcaa cccggccgtca cag 513

SEQ ID NO: 450      moltype = DNA length = 519
FEATURE             Location/Qualifiers
misc_feature        1..519
note = source = /note="Description of Artificial Sequence:
                     Synthetic polynucleotide"
1..519
note = source = /note="Mammalian codon optimized LPG50147"
source           1..519
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 450
atgagcatcc ctgagctgaa tcacgatgtc tggatgcggc acgcctgtat actggctaa 60
agagccaggaa aagaggcgaa agtgcctgtg ggccgtgtc ttgtgtgaa cggccaggatg 120
atcggagaag gctggaaaccc ggccatggc ctgcatgacc ccacccgtccca cgccgagatc 180
atggccctga gacagggggg actggcttc caaaattata gactgtatcga caccacccgt 240
tacgatgtatcc tcgagcccttg tttgtatgtc gcggccgtca ttgtgtactc tagaatcggtc 300
cagctgttct tgccgtgtcgca caacagcaag cggggcgatct ctggctccct gatcaacgtg 360

```

---

-continued

---

```

ctgaactacc ccggcatgaa tcatagatgt gaaatcaccc agggcggttct cagagatgag 420
tgccgcgata tgctgtgcga ctcttacccgg cagcttagac aggtcttaa cgccctgaaag 480
aaacctgcgcg ggcacatcaa cgccctgcag aacaacaga 519

SEQ ID NO: 451      moltype = DNA length = 504
FEATURE
misc_feature
1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..504
note = source = /note="Mammalian codon optimized LPG50148"
source
1..504
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 451
atagcaacc ccgagttcac acacgagttac tggatgcggc acgcccgtac actggccgc 60
agagccagag atgaggggca agtgcctgtg ggccgcgtgc tggctctgaa caaccagggt 120
atcgccgaaag gtggaaaccc ggccatttgcg ctgcgttgcacc ccacccgcga cgccgaaatc 180
atggccctgcg gacaggccgg actgtgtgtc cagaacttacggcgttgcg caccacccgt 240
tacgtgacat tcgagccatg tggatgtgtt agccggcgttgc tggcttcattt tagaatccgc 300
accctggttt tcggcggtcg gaaacgcaag agaggaggtt ctggcagctt gatgaacgtg 360
ctgaattttt ctggatgaa tcaccagggtt aagaccatcg gccggcggtt cggccctgaa 420
tgcggccggc tgctgtgcga ctcttacacaa atgccttagac aagtgtttaa ccagcagaaaa 480
ggcgagctga agtccatcaa cgac 504

SEQ ID NO: 452      moltype = DNA length = 501
FEATURE
misc_feature
1..501
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..501
note = source = /note="Mammalian codon optimized LPG50149"
source
1..501
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 452
atggcgacg ccgagctgac ccacgagttac tggatggagac acgcccgtac actggcccg 60
cgccgcagag atgaggggaga agtgcctgtg ggccgcgtgc tggctctgaa caaccagggt 120
atcgccgagg gtggaaatag agccatcgcc ctgcgttgcacc ccacccgcga tgctgaaatc 180
atggccctgcg gacaggccgg actgtgtgtc cagaacttacggcgttgc tggcttcattt cggatccgt 240
tacgtgacat tcgagccatg tggatgtgtt agccggcgttgc tggcttcattt cggatccgt 300
agactgtatct tcggcggtcg gaaacgcaag cggggcgcaag ctggatctt gattaacgtg 360
ctgaattttt ctggatgaa ccacagggtt gaagtgggtt aaggcatctt gagatgttgg 420
tgcggccggc tgctgtgcga ctcttacccgg caacatcgac aagtctttaa cgccctcaag 480
aaaggcgcca ccgacatcaa c 501

SEQ ID NO: 453      moltype = DNA length = 501
FEATURE
misc_feature
1..501
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..501
note = source = /note="Mammalian codon optimized LPG50150"
source
1..501
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 453
atggcgacg ccgagctgac ccacgagttac tggatggagac acgcccgtac actggcccg 60
agagctaggatggggaga agtccccgttggatgttgc tggctctgaa caaccagggt 120
atcgccgagg gtggaaatag agccatcgcc ctgcgttgcacc ccacccgcga cgccgaaatc 180
atggccctgcg gacaggccgg actgtgtgtc cagaacttacggcgttgc tggcttcattt cggatccgt 240
tacgtgacat tcgagccatg tggatgtgtt agccggcgttgc tggcttcattt cggatccgt 300
agactgtatct tcggcggtcg gaaacgcaag cggggcgccgg ctggatctt gattaacgtg 360
ctgaattttt ctggatgaa ccacagggtt gaagtgggtt aaggcatctt gagatgttgg 420
tgcggccggc tgctgtgcga ctcttacccgg caacatcgac aagtctttaa cgccctcaag 480
aaaggcgcca ccgacatcaa g 501

SEQ ID NO: 454      moltype = DNA length = 507
FEATURE
misc_feature
1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..507
note = source = /note="Mammalian codon optimized LPG50151"
source
1..507
mol_type = other DNA

```

---

---

-continued

---

```

SEQUENCE: 454
organism = synthetic construct
atgagcacc tggaaactgaa ccacgagtagc tggatgagac acgcctcgca actggccag 60
agggccagag atgaggaga agtgcgcgtg ggccgcgtgc tggctcacaa caaccaggtt 120
atcggagaag gctgaatag agccattggc ctgcgtaccc caaccgccta tgctgaatc 180
atggccctgc ggcaggccgg actgggtgtc cagaactacc ggctgtgga caccacctg 240
tatgtgacct ttgagccttg tggatgtgc tccggcgcca tggtcacag cagaatcgga 300
acactggtgt tggcggtgcg gaacgagaag cggggcgctg ctggcgact gatgaacgtg 360
ctgagatacc cccgcatgaa tcaccagtg caaatcatcg acggcggtgt ggcccctgaa 420
tgcagcgccg tgcgtgtgcg ctctacacaa atgcctagac aacgaaaaaa ccagcaaaag 480
gccgagtcta catctagccc tggat 507

SEQ ID NO: 455      moltype = DNA length = 501
FEATURE           Location/Qualifiers
misc_feature      1..501
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..501
note = source = /note="Mammalian codon optimized LPG50152"
source            1..501
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 455
atgagcaca acgagactgaa ccacgagtagc tggatgcggc acgcctggg cctcgccaaa 60
agagccagag aggaaggcga ggtgcgcgtg ggccgcgttc tggctctgaa caaccaggt 120
atcggagaacag gctggaaacag ggcattggc ctgcgtaccc caacagccca cgccgagatc 180
atggctgtga gacaggccgg cctgggtgtc cagaactata gactgacaga taccacctg 240
tacgtgacat ttgagccttg tggatgtgc gccggagcaa tggtcacag cagaatcgga 300
acccctgggt tggcggtgcg gaacgcaag cggggcgccg cgggcctct gatgaacgtg 360
ctgaattacc cccgcatgaa tcatacgatgt gaaattgtgg aaggatctt gaggcgatcc 420
tgcagcgccg tgcgtgtgcg ctctacacaa aaccttagag cctgtgaagaa cgccctgaaag 480
aaggccgtc atccgtccgc t 501

SEQ ID NO: 456      moltype = DNA length = 492
FEATURE           Location/Qualifiers
misc_feature      1..492
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..492
note = source = /note="Mammalian codon optimized LPG50153"
source            1..492
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 456
atgagcata cagaattcac ccacgagcac tggatgagac acgcctgtac actggctaa 60
agagcccggg acgaggccga agtgcgcgtg ggccgcgtgc tggatgtgaa caaccaggtt 120
atcggagaatg gctggaaatag agccattggc ctgcgtaccc caaccgccta cgccgagatc 180
atggccctga gacaggccgg cctggctgtc cagaactata ggctgtgga caccacctg 240
taacgtgacat ttgagccttg tggatgtgc gccggcgca tggatctt gatgaacgtg 300
catctggtgt tggcggtgcg gaacgcaag cggggcgccg tggatctt gatgaacgtg 360
ctggggctacc cccgcatgaa tcaccagtg cagggtgtccg aaggcggtgt ggccaccgaa 420
tgcagcgctc tgcgtgtgcg ctctacacaa aaccttagac tggtaaaaaa cgccctgaaag 480
gaaaaggcaca ga 492

SEQ ID NO: 457      moltype = DNA length = 513
FEATURE           Location/Qualifiers
misc_feature      1..513
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..513
note = source = /note="Mammalian codon optimized LPG50154"
source            1..513
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 457
atgagcagt ccgagttcac ccacgagcac tggatgcggc acgcctgtac actggccaga 60
agagccagag aggaaggcga ggtgcgcgtg ggccgcgtgc tggatgtgaa caaccaggt 120
atcggagaag gctggaaacag agccattggc ctgcgtaccc caaccgccta cgccgagatc 180
atggccctga gacaggccgg cctggctgtc cagaactata ggctgtgga caccacctg 240
tatgtgacat ttgagccttg tggatgtgc gccggcgctt tggatctt gatgaatgt 300
aatctggtgt tggcggtgcg gaacgcaag cggggcgccg tggatctt gatgaatgt 360
gtggggctacc cccgcatgaa ccaccaaata aacgtgtatcg aaggcggtct tgcagaagaa 420
tgcagcgccg tgcgtgtgcg ctctacacaa aaccttagac tggtaaaaaa cgccctgaaag 480
gaaaaggcaca gaaacggcaa caatccta aac aag 513

SEQ ID NO: 458      moltype = DNA length = 492

```

---

-continued

---

```

FEATURE                               Location/Qualifiers
misc_feature                         1..492
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic polynucleotide"
misc_feature                         1..492
                                         note = source = /note="Mammalian codon optimized LPG50155"
source                                1..492
                                         mol_type = other DNA
                                         organism = synthetic construct

SEQUENCE: 458
atgagcaacc ccgagctgac acacgagcac tggatgagat acgcctgac actggccaag 60
cggggccagag aggaaggcga agtgcctgtc ggccgcgtc tggatcgaa caaccagggt 120
atccggcgaag gctggAACAG agccatcgcc ctgcattgtc ctaccgccta cgccgaaatc 180
atggccctgc gacaggggcg  actgtgtgc cagaactata gactgtcgca caccacctg 240
tacgtgaccc tcgagcctt tggatgtgc gccggagact tggatgtc cagaatttggc 300
caactgttgtt tggatgtgcg gaacacgaa aggcccgtcg ctggctctt gatgttgt 360
ctgaattacc ccggcatgaa ccacagaatc gagtttacag gagggagtgt gcgggacgag 420
tgcggccgtt tgctgtgcga ctgttaccgg caacatcgac tggatcgaa cgccctgaaa 480
accggcaacg cc                                         492

SEQ ID NO: 459      moltype = DNA length = 498
FEATURE                               Location/Qualifiers
misc_feature                         1..498
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic polynucleotide"
misc_feature                         1..498
                                         note = source = /note="Mammalian codon optimized LPG50156"
source                                1..498
                                         mol_type = other DNA
                                         organism = synthetic construct

SEQUENCE: 459
atgagcgttc ctgagctgaa tcatgtat tggatgagac acgcctgca gctggctaaa 60
agagccagag aggaaggcga agtgcctgtc ggccgcgtc tggatcgaa caaccagggt 120
atccggcgggg gctggAACAG agccatcgcc ctgcattgtc ccaccgccta cgccgaaatc 180
atggccctgc gacaggggcg  actgtgtgc cagaactata ggtgtgtcgca caccacactg 240
tacgtgaccc tcgagcctt tggatgtgc gccggagact tggatgtc ctagaatcg 300
caactgttgtt tggatgtgcg gaacacgaa aggcccgtcg ccggcgtcgt gatgttgt 360
ctgagatacc ccggcatgaa ccacagggtg aacgtgtcg gcggcgtcgt gggccctgt 420
tggatcgaga tgctgtgcga attctacaga atgcctacac agcagaagaa ccggcaaaag 480
gccgagagca agtgcgtc                                         498

SEQ ID NO: 460      moltype = DNA length = 498
FEATURE                               Location/Qualifiers
misc_feature                         1..498
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic polynucleotide"
misc_feature                         1..498
                                         note = source = /note="Mammalian codon optimized LPG50157"
source                                1..498
                                         mol_type = other DNA
                                         organism = synthetic construct

SEQUENCE: 460
atgagcgttc acgacgtgaa tcacgagcac tggatgcggc acgcctgac actggccag 60
cgcggccagag aggaaggcga ggtgcctgtc ggccgcgtc tggatcgaa aaaccagggt 120
atccggcgggg gctggAAATAG agccatcgcc ctgcattgtc ccaccgccta tgccgagatc 180
atggccctgc gacaggggcg  actgtgtgc cagaactata ggtgtgtcgca caccacactg 240
tacgtgaccc tcgagcctt tggatgtgc gccggcgtca tggatgtc ctagaatcg 300
caactgttgtt tggatgtgcg gaacacgaa aggcccgtcg ctggcgtcgt gatgttgt 360
ctgagatacc ctggatcgaa ccacagggtg aacatcgac agggagtgt gtgtgtgt 420
tgcggcgtt tgctgtgcga ctgttaccgg caacatcgac aacagaagaa cgccctcaaa 480
agagccgttc actccaaac                                         498

SEQ ID NO: 461      moltype = DNA length = 504
FEATURE                               Location/Qualifiers
misc_feature                         1..504
                                         note = source = /note="Description of Artificial Sequence:
                                         Synthetic polynucleotide"
misc_feature                         1..504
                                         note = source = /note="Mammalian codon optimized LPG50158"
source                                1..504
                                         mol_type = other DNA
                                         organism = synthetic construct

SEQUENCE: 461
atgagcaacc ccgagcacaa ccacgagttac tggatgcggc acgcctgac cctggccag 60
aggggccagag atgagggaga agtgcctgtc ggccgcgtc tggatcgaa caaccagggt 120
atccggcgtt gctggAACAG agccatcgcc ctgcattgtc ccaccgccta cgccgaaatc 180

```

---

-continued

---

```

atggccctga gacaggcg  cctgggtctg cagaactacc ggctgttgc  cacaaccctg  240
tatgtacctt ttggatctt tggatgtt agcggcgctt tggtgactc tagaatccg  300
acactgggtt tcggcgctg caacgagaag cggggcgccg ctggcagct gatgaacgt  360
ctgggcattt cggcatgaa tcaccaggta caaacatcg gcccgttgc  420
tgctccggcc tgctgtgcg ctttctacaga atgccttagac aacagaaaa ccagcagaag  480
ggcaactga atcaaccctgg cgac                                504

SEQ ID NO: 462      moltype = DNA  length = 504
FEATURE           Location/Qualifiers
misc_feature      1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..504
note = source = /note="Mammalian codon optimized LPG50159"
source            1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 462
atgagcggacc tggaaactgaa tcacgagttac tggatggagac acggccctgag cctggctaa  60
agagccagag atgaggcgaa agtgcggctg ggcggcgctg tggtgttgc  caaccagggt  120
atcgccgagg gatggaaaccg ggccatttgcg ctgcatttgc  ccaccggccca cgctgaaatc 180
atggccctga ggcaggcg  actgggtctt cagaactacc gactgttgc  caccaccctg  240
tacgtgacat tcggatccatg tggatgttgc tctggcgctt tggtgcattc tagaatccg  300
acactggctt acggcgctg  gaacgagaag cggggcgccg cggcgacgt gatgaatgt  360
ctgggcattt ctggatgaa ccaccaggta caaatcatcg gcccgttgc  420
tgcaagcgcc tgctgtgcg ctttctacgc atgccttagac aacagaaaa ccagcagaag  480
ggcgagctga agtcccgagg agat                                504

SEQ ID NO: 463      moltype = DNA  length = 498
FEATURE           Location/Qualifiers
misc_feature      1..498
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..498
note = source = /note="Mammalian codon optimized LPG50160"
source            1..498
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 463
atgtctgtatc acgaggtaa cgtatgttac tggatggcg  acggccctgac cctggctaaa  60
agagccaggaa aagaggcgaa ggtgcctgtt ggcggcgctg tggtgttgc  caaccagggt  120
atcgggaaatag gatggaaatag agccatggc  ctgcatttgc  ccaccggccca tgctgaaatc 180
atggccctga gacaaggagg cctggccctt cagaactacc gcctgttgc  cgccacactg  240
tacgtgacat ttggatccctt tggatgttgc gccggcgccca tggtgcacag cagaatccg  300
cggtcggttt tcggcgctg  gaacgagaag cggggcgctt ctggcgacgt gattaacgt  360
ctgaattacc ccggcatgaa ccacaggta gaaatccatcg agggcatctt ggccgagtc  420
tgcaagcgcca tgctgtgcg ctttctacaga tggcttagag aggtgaagaa cggccctgaa  480
aaggccagac aggaggaa                                498

SEQ ID NO: 464      moltype = DNA  length = 498
FEATURE           Location/Qualifiers
misc_feature      1..498
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..498
note = source = /note="Mammalian codon optimized LPG50161"
source            1..498
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 464
atgagccaga ccgaaactgac ccacgaggat tggatggcg  acggccctgac actggcccaa  60
agagccagag acgaggcgaa agtgcctgtt ggcggcgctg tggtgttgc  caaccagggt  120
atcgccgaaatag gctggaaatag ggccatttgcg ctgcatttgc  ctaccggccca cgccgagatc 180
atggccctga gacaggcg  cctggccctt cagaactacc ggctgttgc  caccaccctg  240
tacgtgacat tcggatccctt tggatgttgc gccggcgacgt tggtgcacag cagaatccg  300
acactgggtt tcggcgctg  gaacgaaa agggcgctt tggtgcattt gatgaatatc  360
acaggctacc ccggcatgaa ccacaggta caagtgttgc agggcatctt ggctacagag  420
tgctccggcc tgctgtgcg ttttaccgc cagccttagac tggtaagaa cggccctgaa  480
gaagccgcca agaccgca                                498

SEQ ID NO: 465      moltype = DNA  length = 501
FEATURE           Location/Qualifiers
misc_feature      1..501
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..501

```

---

---

-continued

---

```

source          note = source = /note="Mammalian codon optimized LPG50162"
1..501
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 465
atgagcaacc ccgagctgaa ccatgattac tggatgcggc acgcctgag cctggccaa 60
cgggccagag aggaaggcga agtgcctgtg ggcccgtgc tggtggaa caacgagg 120
atccggcagg gatggAACAG agccatcgcc ctgcattacc ctacagccca cgccqagatc 180
atggccctga gacaggcgg catggcttc cagaactata gactgtatca caccaccctg 240
tacgtgaccc tccgagccttg tggatgtgc gccggcgc 300
cagctggatcc ttggcggttg aaattctaa cgccggatgt ctgggttccct gatgaacgtg 360
ctgaattacc ccggcatgaa ccacagatgt gaaaatcggtt aaggcgtgt gcgggacgag 420
tgcgcggaa tgctgtgcga ctctacagg caacctagac tggtaagaa cgcccagaaa 480
aaggcgctg aacacctgtat t 501

SEQ_ID NO: 466      moltype = DNA length = 516
FEATURE           Location/Qualifiers
misc_feature      1..516
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..516
note = source = /note="Mammalian codon optimized LPG50163"
source            1..516
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 466
atgagcaacc ccgagctgaa tcacggtac tggatgagat acgcctgac cctggccaa 60
agagccagag acaaggaga ggtgcctgtg ggccctgttc tggtgcataa cgaccagg 120
atccggcagg gctggAACCG ggccatcgcc ctgcattacc ctacagccca cgccqagatt 180
atggccctgc gccaggcggg cctgggtgtg cagaactacc ggctgtatca cacaaccctg 240
tacgtgaccc ttggcggttg ctgtgtgtgc gccggagcaa tggtgcacag cagaatcg 300
aagactggtgt tccggcggtcg gaaacgaa cggggcgctg ctgggttccct gatgaacgtg 360
ctgaattacc ctggaaatgaa ccacatcgatc gagatggaa aaggcgtgt gagatgg 420
tgcgcggcca tgctgtgtga ttctacaga caacctagaa tggtaagaa cgcccattaa 480
aagtccccac ctgacagcccc taatctgcag gccaga 516

SEQ_ID NO: 467      moltype = DNA length = 504
FEATURE           Location/Qualifiers
misc_feature      1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..504
note = source = /note="Mammalian codon optimized LPG50164"
source            1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 467
atgagcaacc ccgaaattcac ccacggtac tggatgagac acgcctgac cctggctaga 60
cgcggccggg acgaggcggg ggtgcctgtg ggccctgtgc tggtcttcaa caaccagg 120
atccggcagg gctggAAATAG agccatcgcc ctgcattacc ctacagccca cgctgaaatc 180
atggccctga gacaggcggg cctgggtgtg cagaactacc ggctgtgtga caccaccctg 240
tacgtgaccc ttggcggttg tggatgtgt agccggccca tggtgcactc tagaatcg 300
aactgttgt tccggcggtcg gaaacgaa cggggcgccg cggcgcgtt gatgaatgtg 360
ctggatatac ccggcatgaa ccacatcgatc aagaccatcg gaggcgtgt gcggccctgaa 420
tgcagcggac tgctgtgcga ctctacaga atgcctagac agcaaaagaa ccagcagaaa 480
cccgagctg agtcccgccg cgtat 504

SEQ_ID NO: 468      moltype = DNA length = 495
FEATURE           Location/Qualifiers
misc_feature      1..495
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..495
note = source = /note="Mammalian codon optimized LPG50165"
source            1..495
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 468
atgagcggaca acgaggtaaa ccacggtac tggatgagac acgcctgac cctggccag 60
cgcggccagg atgaggcggg ggtgcctgtg ggccctgtgc tggtcttcaa taaccagg 120
atccggaaag gctggAAATAG agccatcgcc ctgcattacc ctacagccca cgccqagatc 180
atggccctga ggcaggcggg catggcttc cagaactata gactgtatca cgctactgt 240
tacgtgaccc tccgagccttg ctgtgtgtgc gccggcgctt tggttccatc tagaatcg 300
cactgttgt tccggcggtcg gaaacgaa cggggcgccg ctggcgcgtt gatgaacgtg 360
ctgaactacc ccggcatgaa tcacagatgt gaagtgcacag aaggcgtgt gcgggaacag 420
tgcgcggcca tgctgtgcga ctctacaga aaccaaaagaa aacaattaa cgccctgaga 480

```

---

-continued

---

aaggctcaga aagcc	495
SEQ ID NO: 469	moltype = DNA length = 510
FEATURE	Location/Qualifiers
misc_feature	1..510
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..510
	note = source = /note="Mammalian codon optimized LPG50166"
source	1..510
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 469	
atgagcaca acgagctaa tcacgatcg tggatgagac acggccctgac cctggccag 60	
cgggccagag atgaggaga agtgcctgtg ggcgcgtgc tggatgtgaa caaccagg 120	
atcgccgaag gctggatag agccatcgcc ctgcgtatgc ctaccgcca cgctgaaatc 180	
atggccctga gacaggccgg aatggctcg cagaactata gactgtatcgca cgccacactg 240	
tacgtgacat tccggccatg tatcatgtgc gccggccca tggtgactc tagaatccgc 300	
cagggttgtt tccggcgtgcg caaacgaaag cggggcgctg cgggtccct gattaacatc 360	
ctgaactacc ctggcatgaa ccacagatgt gacgtgaccc agggcgctgt gagcgacgg 420	
tgcgccaaca tgctgtgcga ttcttacccg gaaacctagac tgcaatttaa cgccagaga 480	
aaggccgaga aagccggaaa tgccgctgt 510	
SEQ ID NO: 470	moltype = DNA length = 507
FEATURE	Location/Qualifiers
misc_feature	1..507
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..507
	note = source = /note="Mammalian codon optimized LPG50167"
source	1..507
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 470	
atgagcaacc ccgagctgac ccacgaccac tggatgcggc acggccctgac cctggccag 60	
agagccagaa acgaggaga agtgcctgtg ggcgcgttc tggatgtgaa cggccaaatg 120	
atcgccgaag gctggacacag agccatcgcc ctgcgtatgc ctaccgcca cgccgagatc 180	
atggccctgc ggcaggccgg actggctctc cagaactacc gggtatcgca caccgtctg 240	
tacgtgacat ttggacttgc tggatgtgc gctggccca tggtccatc tagaatccgc 300	
cagctgggtt tccggcgtgcg caatagaaag cggggcgctg cgggtccct gattaacatc 360	
ctgaactata ctggcatgaa ccacagatgt gaaatcatcg agggcgctgt gagatgatg 420	
tgcgcaacta tgctgtgcga ttcttacaga caccacgac tggatgtgaa cgccctgaaa 480	
aagaatccgc gaaacatcccc aacacag 507	
SEQ ID NO: 471	moltype = DNA length = 498
FEATURE	Location/Qualifiers
misc_feature	1..498
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..498
	note = source = /note="Mammalian codon optimized LPG50168"
source	1..498
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 471	
atgagcaca cagagctaa ccacgatcg tggatgcggc acggccctgat gctggctaaa 60	
cgcgccagag atgaggaga agtgcctgtg ggcgcgtgc tggatgtgaa gaaccagg 120	
atcgccgaag gctggacacag agccatcgga ctgcgtatgc ctaccgcca cgctgaaatc 180	
atggccctga gacaggccgg cctggctctc cagaactata gactgtatcgca caccacctg 240	
tacgtgacat ttggacttgc tggatgtgc gccggcgctg tggtccatc tagaatccgc 300	
aatctggttt tccggcgtgcg gaacagaaag cggggcgccg ctgggtccct gattaacatc 360	
ctgaattacc ctggcatgaa ccacagatgt gaaatcgccc agggcgctgt ggccgacgaa 420	
tgcagcgcaca tgctgtgcga ttcttacccg catccatagac agcagcaaaa cgccctgaaag 480	
caggccgcca agcagatc 498	
SEQ ID NO: 472	moltype = DNA length = 513
FEATURE	Location/Qualifiers
misc_feature	1..513
	note = source = /note="Description of Artificial Sequence: Synthetic polynucleotide"
misc_feature	1..513
	note = source = /note="Mammalian codon optimized LPG50169"
source	1..513
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 472	

---

-continued

---

```

atgagcgaca tcgagctgaa tcacgagta tggatgagac acggccgtat gctggccaag 60
agagccagag aggaaggcga achtgcgtg ggccgcgtgc tggatgtgaa caaccagggtg 120
atcgaggaaag gatggaaaccc ggcatecgcc ctgcgtatgatc ctacagccca cgccgagatc 180
atggccctga ggcaggccgg actggccctc cagaactaca gactgtatcga caccacccctg 240
tacgtgaccc ttgagccatg tggatgtgc gccggcgccca tggatgtgacag cagaatccggc 300
cacctgggtt tggatgtgcg gaacagcaag cggggcgctg ctggatgtccct gattaacgtg 360
ctgaactatc ctggcatgaa ccacagaatc gaatttacccgg aggccgtgtt ggctgtatgag 420
tgatctggca tgctgtgcga cttttacaga taccctagac agcagaaaa tacactgaaag 480
caggccgcta aagccaaaccc cccggccgccc cag 513

```

```

SEQ ID NO: 473 moltype = DNA length = 495
FEATURE Location/Qualifiers
misc_feature 1..495
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..495
note = source = /note="Mammalian codon optimized LPG50170"
source 1..495
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 473
atgagcgaca accgagctgaa ccacgagaga tggatgcggc atggccgtat cctggctcaa 60
agagccagag atggggcgca ggtggatgtgc ggccgcgtgc tggatgtccctt gattaacgtg 120
atcgaggaaag gatggaaaccc ggcatecgcc ctgcgtatgatc ctacagccca cgccgagatc 180
atggccctga gacaggccgg actggatgtgc cagaattacc gactgtatcga cacaacccctg 240
tacgtgaccc ttgagccatg tggatgtgc gccggcgccca tggatgtgactc tagaatcgga 300
cacctgggtt tggatgtgcg gaacagcaag cggggcgctg ctggatgtccct gattaacgtg 360
ctgaattatc ctggcatgaa ccacagaatc gaatttacccgg aggccgtgtt ggccgaatcc 420
tgagcgccca tgctgtgcga cttttacaga cacccttagag aacagaagaa cgccctgagg 480
caggccgctc agagc 495

```

```

SEQ ID NO: 474 moltype = DNA length = 498
FEATURE Location/Qualifiers
misc_feature 1..498
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..498
note = source = /note="Mammalian codon optimized LPG50171"
source 1..498
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 474
atgagcgata tggatgtgcg acggccgtat cctggccaaag 60
cggggccagag aagaggagaga agtccccctgtt ggccgcgtgc tggatgtccctt gattaacgtg 120
atcgaggaaag gatggaaaccc ggcatecgcc ctgcgtatgatc ctacagccca cgccgagatc 180
atggccctga gacaggccgg cctggatgtgc taaaattata gactgtatcga cggccacctg 240
tacgtgacat ttgagccatg tggatgtgc gccggcgccca tggatgtgactc tagaatcgcc 300
aggctgggtt tggatgtgcg gaacagcaag cggggcgctg ctggatgtccct gattaacgtg 360
ctgaattacc caggcatgaa ccacagaatc gaatttacccgg aggccgtgtt ggccgaatcc 420
tgagcgccca tgctgtgcga cttttacaga tggatgtgactc tagaatcgga 480
aaggccccggg aacagaac 498

```

```

SEQ ID NO: 475 moltype = DNA length = 507
FEATURE Location/Qualifiers
misc_feature 1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..507
note = source = /note="Mammalian codon optimized LPG50172"
source 1..507
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 475
atgagcgacc tggatgtgcg acggccgtat gctggctaaa 60
agagccagat agaggccgca agtccccctgtt ggccgcgtgc tggatgtccctt gattaacgtc 120
atcgaggaaag gatggaaatag ggcatecgcc ctgcgtatgatc ctacagccca tgccgaatc 180
atggccctga ggcaggccgg cctggatgtgc taaaattata gactgtatcga caccacactg 240
tatgtgaccc ttgagccatg tggatgtgc tggatgtccctt gattaacgtg 300
accctggatcc acggccgtgcg gaacagcaag cggggcgccca tggatgtccct gattaacgtg 360
ctggggatcc cgggtatgaa tggatgtgcg caatgtatc acggccgtgtt ggccctgaa 420
tgagcgccca tgctgtgcga cttttacaga tggatgtgactc tagaatcgga 480
ccggagagca ccacgac 507

```

```

SEQ ID NO: 476 moltype = DNA length = 486
FEATURE Location/Qualifiers
misc_feature 1..486

```

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..486
note = source = /note="Mammalian codon optimized LPG50173"
source 1..486
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 476
atgagcaca ccgagctgac ccacgagta tggatgcggc acgcccgtat gctggctaa 60
aggccatagg atgaaggcga agtgcacgtg ggagccgtgc tggatgtcaa caaccgggtg 120
atcgccgagg gtcggacacg agtcatcgaa ctgcgtatgtc ctacagccca cggcggatc 180
atggccctga gacaggccgg cctgtgtctg cagaactacc gcgtgtgtt caccacccctg 240
tacgttacat ttgagccttg tggatgtgc gcccggcgtt cggatccctt gatgaatgtg 300
acactgttgt tcggcgatcgaa gaaccttgg cggggccggc cggatccctt gatgaatgtg 360
ctgaattatc ctggcatgaa ccacagatgtt aaggaaacctt ctccggatc 420
tgcagcggca tgctgtgcgaa gtttacatcgaa cggccggatc tggatgttccaa cggccggatc 480
caggcc 486

SEQ ID NO: 477      moltype = DNA length = 519
FEATURE
misc_feature 1..519
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..519
note = source = /note="Mammalian codon optimized LPG50174"
source 1..519
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 477
atgtcttatcc ccgagctgac tcacgatgtg tggatgttggac acgtgtgttgc actggccaa 60
aggccatagg aggaaggcga agtgcacgtg ggagccgtgc tggatgttgcgaa cggccaaatgt 120
atcgccgagg gtcggacacg agccatcgaa ctgcgtatgtc ccaccggccca cggcggatc 180
atggccctgc gacaggccgg cctgtgtctg cagaactaca gactgtatcgaa cacaacccctg 240
tacgtgaccc ttcggcgatcgaa gaaccttgg cggggccggc cggatccctt gatgaacatgt 300
caactgttgt tcggcgatcgaa gaaccttgg cggggccggc cggatccctt gatgaacatgt 360
cttaattatc ctggcatgaa ccacatcgaa agggatgtgtt gagatgttgg 420
tgcagcggca tgctgtgcgaa ctttacatcgaa cggccggatc tggatgttccaa cggccggatc 480
aaacctgtccg cggccatcgaa aacaacccgg 519

SEQ ID NO: 478      moltype = DNA length = 498
FEATURE
misc_feature 1..498
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..498
note = source = /note="Mammalian codon optimized LPG50175"
source 1..498
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 478
atgagcata tggatgttggac tggatgttggc acgcccgtat gctggccaa 60
aggccatagg aagaggccgg agtgcacgtg ggagccgtgc tggatgttgcgaa caaccagggtg 120
atcgccgagg gtcggacacg ggcacatcgaa ctgcgtatgtc ccaccggccca tggccggatc 180
atggccctga gacaggccgg cctgtgtctg caaaattatc gcgtgtatcgaa cggccacccctg 240
tacgtgaccc ttcggcgatcgaa gtcgtgttgcgaa gcccggatcgaa tggatgttgcgaa cacaatcgcc 300
aggctgttgt tcggcgatcgaa gaaccttgg cggggccggc cggatccctt gatgaacatgt 360
ctgaactacc caggcatgaa ccacagatgtt gaaatcggatc agggatgtgtt cggccggatc 420
tgcagcggca tgctgtgcgaa ctttacatcgaa tggatgttccaa cggccggatc 480
aaagccatcgaa aacatcgaa 498

SEQ ID NO: 479      moltype = DNA length = 459
FEATURE
misc_feature 1..459
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature 1..459
note = source = /note="Mammalian codon optimized LPG50176"
source 1..459
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 479
atgagcaca tcgagcagaa ccacgagta tggatgcggc acgcccgtgt tctggccaa 60
cggccatagg aggaaggcga agtgcacgtg ggagccgtgc tggatgttgcgaa caaccagggtg 120
atcgccgagg gtcggacacg ggcacatcgaa ctgcgtatgtc ccaccggccca cggccggatc 180
atggccctga gacaggccgg actgtgtctg caaaattatc gcgtgtatcgaa cacaacccctg 240
tacgtgaccc ttcggcgatcgaa gtcgtgttgcgaa gcccggatcgaa tggatgttgcgaa cacaatcgcc 300

```

---

-continued

---

```

agectggct ttggcgtgcg gaacagcaag agaggcgccg ctggctctct gattaacgtg 360
ctgaattatc ctggaaatgaa ccacagatgt gaaaatgaccc agggcgctgt ggctgtatgaa 420
tgcagcgcca tgctgtgcga cttctacaga cacccccaga 459

SEQ ID NO: 480      moltype = DNA length = 504
FEATURE
misc_feature
1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..504
note = source = /note="Mammalian codon optimized LPG50177"
source
1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 480
atgtcaacc ctgagagaga tcacaggtac tggatgcggc acgcccgtac actggccag 60
cggggccagag atgaggcgca agtgcctgtg ggccgcgtgc tggatgtgaa caaccagggtt 120
atcgccgaaatg gatggaaatgg agccatcgcc ctgcgtatgacc ccacccgcata tgccgaaatc 180
atggccctgtg gacaggcgccg catggatgtg cagaactaca gactgtgtgaa caccacccgt 240
taacgtgacat ttgagccctgtg cgtatgtgtg tccggcgccata tggtccactc tagaatcggt 300
acactgggtt tccggcgccg gaacgagaag cggggcgctgt ctggcagctt gctgaatgtg 360
ctggggatata ctggcatgaa ccaccagggtg aqaccatcg gaggcgctgt cgcccccagct 420
tgcagcgccc tgctgtgcga cttctaccgc atgccttagac aacagaaaaa ccagcagaag 480
gccgagctga agctgagcaa cgac 504

SEQ ID NO: 481      moltype = DNA length = 486
FEATURE
misc_feature
1..486
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..486
note = source = /note="Mammalian codon optimized LPG50178"
source
1..486
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 481
atgagcgcca tcgagatgtgaa ccacaggtac tggatgcggc acgcccgtgg cctggcttag 60
cgcgcttagag atgaggcgca ggtcccccgtg ggccgcgtgc tggatgttacca gaaccagggt 120
atcgccgaaatg gatggaaatgg gcgcattgtgc ctgcgtatgacc ccacccgcata cgccgaaatc 180
atggccctgtg gacaggcgccg actggatgtg cagaattacc ggtgtatgtca caccacccgt 240
taacgtgacat ttgagccatgtg tggatgtgtg gccggcgctgt tggtccactc tagaatcggt 300
agagtgggtt tccggcgccatgg aaacagcaag cggggcgccg cggccagctt gatgaacgtg 360
ctcaattatc ctggcatgaa ccatagagggtg gaagtgcaccg agggcgctgt ggccggagaa 420
tgctccgcata tgctgtgcga cttctacaga gcccctagggt ctcaatttaa cgcccgaaag 480
agacctt 486

SEQ ID NO: 482      moltype = DNA length = 507
FEATURE
misc_feature
1..507
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..507
note = source = /note="Mammalian codon optimized LPG50179"
source
1..507
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 482
atgagcaacc ctgagatgtgaa ccacaggtac tggatgtgggtt acgcccgtac cctggccaaag 60
cggggccagag aggaaggcgca agtggcgtgtg ggccgcgtgc tggatgtgaa cgaacgggtg 120
atcgccgaaatg gatggaaatgg agccatcgcc ctgcgtatgacc ccacccgcata cgccgagatc 180
atggccctca gacaggcgccg catggatgtgtg cagaactaca ggtgtatgtca caccacccgt 240
taacgtgacat ttgagccatgtg tggatgtgtg gccggcgctgt tggtccactc tagaatcggt 300
cacctgggtt tccggcgccg gaacagcaag agaggagggtt ctgggtccctt gatgaacgtg 360
ctgaattacc ccggcatgaa tcatagagggtg gccattacag agggcgctgt gagatgaa 420
tgcagcgccata tgctgtgcga cttctaccgc cagcttagac aagtgaagaa cgccctgaaa 480
aagaccttga ggcgtatgcca ggagcag 507

SEQ ID NO: 483      moltype = DNA length = 504
FEATURE
misc_feature
1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature
1..504
note = source = /note="Mammalian codon optimized LPG50180"
source
1..504
mol_type = other DNA

```

---

---

-continued

---

```

SEQUENCE: 483
organism = synthetic construct
atgtccaatc ctgagcacga ccacgagtag tggatgcggc acgcccgtaa cctggcccg 60
cggggccagag atgaggcgca ggtccccgtg ggccgcgtgc tggctctcaa caaccaggtc 120
atccggagaag gctggaaaccg cgccatcgcc ctgcattgacc caacagccca tgctgaatc 180
atggccctga gacagggcgcc cctgggtctg cagaactacc ggctgtgtga tacaacctg 240
tacgtgaccc tggagccctg cgtgatgtg agccgcgtta tggtcacag ccggatccgc 300
acccctggct acggcggttag aaacggagaaa agaggcgccg ccggcagct gatgaacgtg 360
ctgggatatac ctggaaatgaa tcaccagggt aacgtgtacg gcccgggtgt ggctcaggac 420
tgttctgcga gactgtgcga ctttcacaga atgcctagac agcaaaaagaa ccagcagaga 480
gccgaactga aggccaaagg cgac 504

SEQ ID NO: 484      moltype = DNA length = 504
FEATURE           Location/Qualifiers
misc_feature      1..504
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..504
note = source = /note="Mammalian codon optimized LPG50181"
source            1..504
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 484
atgagcggacc ccgagctgaa tcacgaggat tggatgagac acgcccgtca actggcccg 60
agagccagag atgaggcgca agtgcacgtg ggccgcgtgc tggctctgaa caaccagggt 120
atccggagaag gctggaaacag ggccatcgca ctgcattgacc ctacagccca cggccaaatc 180
atggccctga gacagggcgcc cctgggtctg cagaactacc ggctgtgtga caccacctg 240
tacgtgaccc tggagccctg cgtgatgtg tggccgcgtca tggccgcgtcg cagaatccgc 300
acagtgtgt acggcggtcgca gaacggagaaa cggggcgccgt ctggcagct gctgaatgtg 360
cttccttacc cccggcatgaa ccaccagggtt aagggtgtacg gcgaaatgtgt ggcccgtgt 420
tgtagcgcga tgctgtgcga ctttcacaga atgcctagac agcagaaaaa ccagcaaaag 480
gccgagtgga agctgagcgcc cgag 504

SEQ ID NO: 485      moltype = DNA length = 513
FEATURE           Location/Qualifiers
misc_feature      1..513
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..513
note = source = /note="Mammalian codon optimized LPG50182"
source            1..513
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 485
atgagcaacc ccgagctgaa ccacgagtag tggatgagat acgcccgtac actggccaaa 60
agagccagag acgaggggaga agtgcacgtg ggccgcgtgc tggctctacca cgaccaatgt 120
atccggcgaag gctggaaacag ggccatcgca ctgcattgatc ctacccgccta cggccgatc 180
atggccctcc ggcaggggagg cctgggtctg cagaactacc gactgtgtca caccacactg 240
tacgtgaccc ttggagccctg tggatgtgc gccggccgtca tggtcacag cagaatttgc 300
agactggttt tggcggtcgca caactctaag cggggcgccgt ctggcagct gctgaatgtg 360
ctgaattacc ctggcatgaa ccaccaggatc gatatggaa aaggcggtgt gccggatgtg 420
tgcgcgcgca tgctgtgcga ctttcacgg ctggcttagaa tggatgtggaa tgcactgaag 480
cagtccctc cagacacgcac caacccgtcat gcc 513

SEQ ID NO: 486      moltype = DNA length = 3213
FEATURE           Location/Qualifiers
misc_feature      1..3213
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature      1..3213
note = source = /note="Mammalian codon optimized
nAPG07433.1"
source            1..3213
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 486
atgagagagc tggactacag aattggccctg gccatcgcc caaacagcat cggatggggc 60
gtgatcgagc tggctctggaa caaagaccgg gagagatacg agaagggtcgaa aatcgatggat 120
caaggcggtga gaatgttcga cagacccgtca cggccgtccatc cttggatgtgt 180
cccagaagaa tggcggtccatc cggccgtccatc cggccgtccatc gaaagggtcgaa 240
aacatccggc acctgtgtgt gcaacacccgcgtccatc cttggatgtgt 300
tacccctgtg caaaaagag catggacatc tggggcattc ggctcgacgg cttggatgtgt 360
cttcctcaatc atttcgatgt ggccgactgt ctgtatccacc tggatgtgt 420
aagtccaaaca gaaaggtgtgaa actggaaatg acagagacac gcaagggtgt gggatgtgt 480
caactgaacg agaaacgggtt gagttgtat agaaccgtgg gcgatgtg gatgtgg 540
cccgacttct ctaaatcgca taggaagaga aatagccccca acgaataatcg tttcgatgt 600

```

---

-continued

---

tcttagagccg	agctggaaaaa	ggaaatcggt	accctgttcg	ccgcccacgc	gagattccag	660
agcccttacg	ccagcaaaa	tctgcaggag	acatatctcg	agatctggac	ccaccaactg	720
ccttcgcca	gcggcaatgc	catctgaad	aaggtcggat	actgctcct	gttcaaaggc	780
aaagaaaagaa	ggattccaa	ggctacatac	actttcaat	acttctctgc	tctggacccag	840
gtgaatcgga	ccagactggg	acctgatttc	cagcccttca	ccaaggagca	acggaaatt	900
atcttgaca	acatgttcca	gaggacagat	tactacaaga	agaaaaaccat	ccccgagggt	960
acctactatg	acatacgaa	gtggctggaa	ttggacgaga	caattcgtt	caaggcctg	1020
aactacgacc	ctaacgagga	actaagaag	atcgagaaga	acgcctttat	caatctgaag	1080
gccttctacg	agatcaacaa	ggtgttggcc	aactacagcc	aaagaaccaa	cgagaccc	1140
tccacccctgg	atcacgaa	cacggctca	gcctgtaccc	tgtacaaaac	cgacaaggat	1200
atccgcagct	acctgtggaa	cagtcaaac	ctacctaaga	gatgtacga	cgacaactg	1260
atcgaggaa	tgctgtggct	gagtcacaca	aaggtcggcc	acctgttct	gaaaggccatc	1320
aaccacgtgc	tgtctatcat	gcagaagggc	aataacctaca	aggaagccgt	ggaccaactg	1380
ggctacgaca	ccagcggcc	taaaggaa	aaagggttcca	agttctggcc	accttattt	1440
gatgaardatca	cgaaatccat	cgtgaaagg	gcctgtaccc	aggccagaaa	agtggtaac	1500
gccataatta	gaagacacgg	atctctca	tccgtcaca	tcgagtcgc	cagagactg	1560
agcaagaacc	acgacgagcg	gacaaagatc	gtcagcgc	aggatgaaaa	ctacaagaaa	1620
aacaaggccg	ctatcgat	cctgtctgc	cacggcatcc	tgaacccatc	aggctacgc	1680
atcgtgagat	acaaactgtg	gaaggagcg	ggcgaacacg	ggccttacag	cctgaaggaa	1740
atccctgtcc	atacatttt	caacgcgt	aagaaggaa	gcaacggcc	ccctatctt	1800
gaagtggacc	acatctgcc	ctacagccag	tccttcatcg	actcctacca	caacaaggtc	1860
cttgtgtaca	gogacgaaa	ccggaaaaaa	ggcaacacaa	tcccttatac	ctacttcctg	1920
gaaaccaaca	aggatggaa	ggccttggaa	cggtacgtgc	ggagcaacaa	attttctcc	1980
aaaaaaaagg	gagatgttct	tctgaagccg	gttattcgtc	ctagagaatc	tgagctgatc	2040
aaagaacgc	acctgttacg	caccagatac	gcctctaccc	tcctgaagaa	cttcatcgag	2100
cagaacccgtc	agttcaagga	agccgaggaa	aaccccaaaa	aaagacgggt	gcaaaaccgt	2160
acaggcgtt	tcacccggcc	cttcagaaaa	cggtggggcc	tggagaa	ccggcaggag	2220
acataccccc	atcacgtat	ggacgc	atcg	tgatcagacca	ccacatggc	2280
accagagtga	ccgagacta	tcagatca	gaaagcaaca	agagctgaa	gaaggccat	2340
tttcctatgc	cttggaaagg	cttcgggac	gagctgtga	gccacttgc	ttctcagcc	2400
atcgccaa	aatcgcgg	ggaaatgtgaa	gcccgttac	agagctgtg	ctacatctt	2460
gtgtccagaa	tgcctaa	gatttac	ggcgtgtc	ataagcagac	catatcg	2520
aaaggaggaa	ttgacaagaa	gggcaaaaca	atcatcatcg	aaaggctgca	cctgaaggat	2580
atcaagtgc	acgagaacgg	agatttca	atggtggca	aggaacagga	catggccaca	2640
taccaagata	ttaaaccagag	atacgttgg	cacggcaaga	atagcaagaa	ggccttgcag	2700
acccccctgt	aaagcccg	aaaaaaagg	acaggttaac	tgtcaagcg	ggtgaaggt	2760
gaaggacagg	ccaagacgtt	tgtgagggaa	gtgaaacggcc	gagtggcc	aatatggcgt	2820
ctgggttag	ttgatttgg	tgagaaggat	gataagtact	acatggtccc	catctacgt	2880
ccagacacc	tgtgtggca	gctgc	ccagcttccaa	gggtatag	2940	
cagtggcgt	cactggata	cagtttacc	ttaaagtcc	gctgtaccc	ttatgatct	3000
gtgcggctgg	tcaaggggaa	tgaggatcg	ttctgtact	ttggcaccc	ggacatcgac	3060
agcgacagac	ttaacttca	ggacgtgaa	aaggcaagca	agaagaacga	gtaccgtac	3120
agttgaaaaa	ccatcgagga	cttggagaa	taaggatgg	gctgtctgg	cgatcta	3180
ctggtccgg	aggaaactcg	aagaaactt	cac			3213

SEQ ID NO: 487                  moltype = DNA length = 96  
 FEATURE                  Location/Qualifiers  
 misc\_feature                  1..96  
                           note = source = /note="Description of Artificial Sequence:  
                           Synthetic oligonucleotide"  
 misc\_feature                  1..96  
                           note = source = /note="Codon optimized linker"  
 source                  1..96  
                           mol\_type = other DNA  
                           organism = synthetic construct

SEQUENCE: 487  
 tccggcgggt ctccggccgg ctcttagtggg agtgagacgc caggaacgtc tgaatctgct 60  
 actcccaat ctacggccgg attcaggatggc ggtatgt 96

SEQ ID NO: 488                  moltype = AA length = 1323  
 FEATURE                  Location/Qualifiers  
 REGION                  1..1323  
                           note = source = /note="Description of Artificial Sequence:  
                           Synthetic polypeptide"  
 REGION                  1..1323  
                           note = source = /note="LPG50140-nAPG07433.1 protein  
                           sequence"  
 source                  1..1323  
                           mol\_type = protein  
                           organism = synthetic construct

SEQUENCE: 488  
 MAPKKKRKV D YKDHDG DYKD HDIDYK DDD KMSDLELNHE YWMRHALQLA KRARDEGEVP 60  
 VGAVLVLNQ VIGEGWNRAI GLHDP TAHAE IMLRQGLV LQNYRLIDTT LYVTPEPCVM 120  
 CSGAMVHSRI GTLVFGVRNS KRGAA GSLMN VLNYPGMNHQ VQIIDGV LAP ECSGLLCDFY 180  
 RMPRQVFNQQ KAESTSINGD SGGSSGGSSG SETPGTSESA TPESSGGSSG GSMRELDYRI 240  
 GLAIGTNSIG WGVIELSWNK DRERYEKVRI VDQGVRFMPDR AEMPKTGASL AEPRRRIARSS 300

---

-continued

---

RRRLNRSQR	KKNIRNLLVQ	HGVITQEELD	SLYPLSKKSM	DIWGIRLDGL	DRLLNHFEDA	360
RLLIHLAQR	GFKSNRKSSEL	KDTETGKVLS	SIQLNEKRRLS	LYRTVGEMWM	KDPDFSKYDR	420
KRNSPNEYVF	SVSRAELEKE	IVTLCFAAQRR	FQSPYASKDL	QETYLQIWTH	QLPFASGNAI	480
LNKVGYCSLL	KGKERRIPKA	TYTFQYFSAL	DQVNTRLGP	DFQPFTKEQR	EIIILNNMFQR	540
TDYYKKKTIP	EVTVYDIRKW	LELDETIQFK	GLNYDPNEEL	KKIEKKPFIN	LKAFYEINKV	600
VANYSERTNE	TFSTLDYDG1	GYALTIVYKTD	KDIRSYLKSS	HNLPKRCYDD	QLIEELLSLS	660
YTKFGHLSLK	AINHVLSIMQ	KGNTYKEAVD	QLGYDTSGLK	KEKRSKFLPP	ISDEITNPIV	720
KRALTQARKV	VNAIIRRHS	PHSVHIELAR	ELSKNHDERT	KIVSAQDENY	KKNKGAIISL	780
SEHGLNPTG	YDIVRYKLWK	EQGERCAYSL	KEIPADTFFN	ELKKERNGAP	ILEVHDILPY	840
SQSFIDSYH	KVLVYSDENK	KKGNRIPYT	FLETNKDWEA	FERYVRSNPKF	FSKKKREYLL	900
KRAYLPRESE	LIKERHLNDT	RYASTFLKNF	IEQNLQFKEA	EDNPRKRRVQ	TVNGVITAHP	960
RKRWGLEKDR	QBTYHLHAMD	AIIVACTDHH	MVTRVTVEYYQ	IKESNSKVK	PYFPMPWEGF	1020
RDELLSHLAS	QPIAKKISEE	LKAGYQSLDY	IFVSRMPKRS	ITGAAHQQT	MRKGIDKKG	1080
KTIIIRHLH	KDIKFDENGF	FKVMGKEODM	ATYEAIAKORY	LEHGKNSKKA	FETPLYKPSK	1140
KGTGNLKR	KVEGQAKSFV	REVNGVVAQN	GDLVRVDSL	GDKYMMVPI	YVPDTVCSEL	1200
PIKKVASSKG	YEQWLTLDNS	FTFKPSLYP	DLVRLVKGDE	DRFLYFGTLD	IDSDRLNFKD	1260
VNKPSKKN	RYSLKTIEDL	EKYEVGVLGD	LRLVRKETRR	NFHSGGSKRP	AATKKAGQAK	1320
KKK						1323

SEQ ID NO: 489  
 FEATURE moltype = AA length = 1318  
 REGION Location/Qualifiers  
 1..1318  
 note = source = /note="Description of Artificial Sequence:  
 Synthetic polypeptide"  
 REGION 1..1318  
 note = source = /note="LPG50141-nAPG07433.1protein sequence"  
 1..1318  
 source mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 489  
 MAPKKKRKVD YKDHDGDKD HDIDYKDDDD KMSNPELTHE HWMRYALTIA KRAREEGEVP 60  
 VGAVLVLNQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLIDTT LYVTFPCVM 120  
 CAGAMVHSRI GQLVFGVRNS KRGAAAGSLMN VLNVPGMNRH IEFTEGVLRD ECAAMLCDFY 180  
 ROPRQVNAL KTGNASGGSS GGSSGSETPG TSESATPESS GSSSGGSMRE LDYRIGLAIG 240  
 TNSIGWGVIE KTKNWDKDRERY EVKRVIDQGV RMPDRAEMPK TGASLAEPPR IARSSRRLN 300  
 RKSRQKKNR NLLVQHGKVIT QEEELSLYPL SKKSMDIWI GRLDRLLN HFEWARLLIH 360  
 LAQRRGFPSN RKSELKDTE GKVLSSIQLN EKRLSLYRTV GEMWMKDPDF SKYDRKRNSP 420  
 NEVVFSVSR A ELEKEITVTLF AAQRFQSPY ASKDLQETYL QIWTHQLPFA SGNAILNKVG 480  
 YCSLLKGKER RIPKATVLF YFSQDNQNR TRLQPDFQPF TKEQREIIILN NMFORTDDYK 540  
 KKTIPEVITYY DIRKWELEDE TIQFKGLNYD PNEELKKIEK KPFIINLKFY BINKVANYS 600  
 ERTNETFSTL DYDGIGYALT VYKTDKDIRS YLKSSHNLPK RCYDDQLIIE LLSLSYTKFG 660  
 HLSLKAHV LSIMQKNTY KEAVDQLGVD TSGLKKEKRS KFLPPISDEI TNPIVKRALT 720  
 QARKVVAII RRHGSFHSVH IELARELSKN HDERTKIVSA QDENYKKNKG AISILEHGI 780  
 LNPTGYDIVR YKLWKEQGER CAYSLKEIPA DTFNNELKKE RNGAPILEV D HILPYSQSFI 840  
 DSYHNKVLVY SDENRKKGNN IPYTYFLETN KDWAEFERYV RSNKFFSKKK REYLLKRAYL 900  
 PRESELIKER HLNDTRYAST FLKNFIEQNL QFKEAEDNPR KRRVQTVNGV ITAHFRKRWG 960  
 LEKDRQETYI HHAMDAIIVA CTDHHMVTRV TEYYQIKEBSN KSVKKPYFPM PWEGRFDELL 1020  
 SHLASQPIAK KISEELKAGY QSLDYIFVSR MPKRSITGAH HKQTIMRKGG IDKKGKTI 1080  
 ERLHLKDIKF DEENGDFKVMG KEQDMATYEA IKORYLEHKG NSKKAFTPL YKPSKKGTTGN 1140  
 LIKRVKVEQG AKSFVREVNG GVAQNGDLVR VDLFEKDCKY YMVPIYV PDT VCSELPKKV 1200  
 ASSKGYBQWL TLDSNPTFKF SLYPYDLVRL VKGDEDRLY FGTLIDSDR LNFKDVNKP 1260  
 KNEYRYSLSK TIEDLEKYEV GVLGDLRLV R KETRRRNHFSG GSKRPAATKK AGQAKKK 1318

SEQ ID NO: 490  
 FEATURE moltype = AA length = 1323  
 REGION Location/Qualifiers  
 1..1323  
 note = source = /note="Description of Artificial Sequence:  
 Synthetic polypeptide"  
 REGION 1..1323  
 note = source = /note="LPG50142-nAPG07433.1protein sequence"  
 1..1323  
 source mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 490  
 MAPKKKRKVD YKDHDGDKD HDIDYKDDDD KMSIPELNHD VWMRHALTIA KRAREEGEVP 60  
 VGAVLVLNQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLIDTT LYVTFPCVM 120  
 CAGAMVHSRI GQLVFGVRNS KRGAAAGSLIN VLNVPGMNRH VAITEGVLRE ECAAMLCDFY 180  
 ROPRQVNAL KKPAGDINA F SGGSSGGSS SETPGTSESA TPESSGGSSG GSMRELDYRI 240  
 GLAIGTNSIG WGVIELSWNK DRERYEKVRI VDQGVRMPDR AEMPKTGASL AEPRIARSS 300  
 RRRLNRSQR KKNIRNLLVQ HGVIQEEELD SLYPLSKKSM DIWGIRLDGL DRLLNHFEDA 360  
 RLLIHLAQR GFKSNRKSSEL KDTETGKVLS SIQLNEKRRLS LYRTVGEMWM KDPDFSKYDR 420  
 KRNPSPNEYVF SVSRAELEKE IVTLCFAAQRR FQSPYASKDL QETYLQIWTH QLPFASGNAI 480  
 LNKVGYCSLL KGKERRIPKA TYTFQYFSAL DQVNTRLGP DFQPFTKEQR EIIILNNMFQR 540  
 TDYYKKKTIP EVTVYDIRKW LELDETIQFK GLNYDPNEEL KKIEKKPFIN LKAFYEINKV 600  
 VANYSERTNE TFSTLDYDG1 GYALTIVYKTD KDIRSYLKSS HNLPKRCYDD QLIEELLSLS 660  
 YTKFGHLSLK AINHVLSIMQ KGNTYKEAVD QLGYDTSGLK KEKRSKFLPP ISDEITNPIV 720

-continued

---

KRALTQARKV	VNAIIIRRHGS	PHSVHIELAR	ELSKNHDERT	KIVSAQDENY	KKNKGAIISIL	780
SEHGLNPTG	YDIVRYKLWK	EQGERCAYSL	KEIPADTFFN	ELKKERNNGAP	ILEVDHILPY	840
SQSFIGDSYHN	KVLVYSDENR	KKGNRIPYTY	FLETNKDWEA	FERYVRSNKF	FSKKKREYLL	900
KRAYLPRERE	LIKERHLNDT	RYASTFLKNF	IEQNLQFKEA	EDNPRKRRVQ	TVNGVITAHF	960
RKRWGLEKDR	QETYLHHAMD	AIIVACTDHH	MVTRVTEYYQ	IKESENKSVKK	PYFPMPWEGF	1020
RDELLSHLAS	QPIAKKISEE	LKAGYQSLDY	IFVSRMPKRS	ITGAAHQQT	MRKGGRIDKKG	1080
KTIIIERLHL	KDIKFDENGD	FKMVGKEQDM	ATYEAIKQRY	LEHGKNSKKA	FETPLYKPSK	1140
KGTGNLTKRV	KVEGQAKSFV	REVNGVVAQN	GDLVRVLDLF	KDDKYYMVP	YVPDTVCSEL	1200
PKKVVASSKG	YEQWLTLDNS	FTFKFSLYPY	DLVRLVKGDE	DRFLYFGTLD	IDSRLNFKD	1260
VNKPSKCKNEY	RYSLKTIEDL	EKYEVGVLD	LRLVRKETTR	NFHSGGSKRP	AATKKAGQAK	1320
KKK						1323

SEQ ID NO: 491                    moltype = AA length = 1326  
 FEATURE                            Location/Qualifiers  
 REGION                            1..1326  
 note = source = /note="Description of Artificial Sequence:  
                                   Synthetic polypeptide"  
 REGION                            1..1326  
 note = source = /note="LPG50143-nAPG07433.1protein sequence"  
 source                            1..1326  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 491  
 MAPKKKRKV YKDHDG DYKD HDIDYK DDD KMSNP EPLNH YWMRYALT LA KRAR DEGE VP 60  
 V GAVL VL NDQ VIGEGWN RAI GLHD PTA HAE IMAL RQ GGLV LQ NYR LID TT LY VT FEP CVM 120  
 CAGAMVHS R I GRL VFG VRNS K RGAAG SLLN VL NYP GMN NH IE MEEG VL RD ECAAM LCDF Y 180  
 R QP RQ VFN AL KKS PPD INNL QAR SGSS GG SSG SET PG TS ES ATPE SS GG SSG SSM RE LD 240  
 YRIGLAIG TN SIG WG VIELS W NKDR ERY EK VRIV DQ GVR M FDRA E MP KT G ASLA E PR RIA 300  
 RSSRR RL NR K SQRK KN IR NL LV QH GVI TQE EL DLS LY PLS K KSMD I W G IRL DGL DR LL NH F 360  
 EW AR LL TH LA QR RG FK S N R SEL KDT ETG K VL SS IQ LNE K RLS LY RT VGE MM KDP D FSK 420  
 YDR K R W GLE K R FVS V S R A E L EKE IV TL F PA N Q RR FQ SP Y AS K DL Q E TY L QI WTH QLP FAS G 480  
 NAI LN K V CYC SL LG KER RI PK AT Y TF Q YF SAL DQ VN RTR LG PDF Q PFT K EQ RE I LNN M 540  
 F QRT D YY KKK T IPE V TY D I RK WL EDET I QFK GL NY DP N EEL K KIE KKP FIN LK AF YE I 600  
 NIK VV AN Y SER T NET F ST LD Y DG IGY AL T V Y KTD K DIR SYL KSSH NLP KRC Y DD QL I EELL 660  
 SLS Y TKR GH L KDR Q E TY L HH AMD A I VACT DHH M VTR VTE YY QI KES NKS V KK PYF PMP W 1020  
 PIV K R AL T Q A R K V NAI IR R HGSP HS V HIE LARE LSK NH D ERT KIV SA QD EN Y KKN KG AI 780  
 S IL SEH GIL N PT GY D I V R Y K LW KE Q GER CA YSL KE I PAD T FF NEL K KERN GAPI LEVD HI 840  
 LP YSQ SFI DS YH NK VLV VY SD EN R KKG N R I P YTF LET N K D WEA FERY VRS NK FFS K KRE 900  
 YLL K R P R ESE L KER HL ND T YR IAST YL K NPI E QN LQ F KEA D E N P R K R RV Q T VNG VIT 960  
 AH FR K R W GLE K DR Q E TY L HH AMD A I VACT DHH M VTR VTE YY QI KES NKS V KK PYF PMP W 1020  
 EG FR DELL SH LAS QP IA KKI SEEL KAG Y Q S LD YI F V S R M P KRS IT GAA H QT IM R KG GI D 1080  
 K K G K T III ER L H L K D I K F D E NG DF KM V G K B QD M AT Y E A I K Q R Y L E H G K N S K K A F E T P LY K 1140  
 PS KKG TGN L I K R KV VEG QAK SF VRE VNG GV A QNG D L V R V D L F E K D D K Y Y M V PI Y V P D T V C 1200  
 SEL P K V V A S SK G Y E Q W L T D N S FT K F S L Y P Y D L V R L V K G D E D R F L Y F G T L D I D S R L N 1260  
 F K D V N K P S K K NE Y R Y S L K T I E D L E K Y E V G V L G D L D L R V R K E T R R N F H S G G S K R P A A T K K A G 1320  
 QAK KKK  
 1326

SEQ ID NO: 492                    moltype = AA length = 1323  
 FEATURE                            Location/Qualifiers  
 REGION                            1..1323  
 note = source = /note="Description of Artificial Sequence:  
                                   Synthetic polypeptide"  
 REGION                            1..1323  
 note = source = /note="LPG50144-nAPG07433.1protein sequence"  
 source                            1..1323  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 492  
 MAPKKKRKV YKDHDG DYKD HDIDYK DDD KMSNP EPLTH D HWMR HAL T LA QR AR NE GE VP 60  
 V GAVL VL NDQ VIGEGWN RAI GLHD PTA HAE IMAL RQ GGLV LQ NYR LID TT LY VT FEP CVM 120  
 CAGAMVHS R I GRL VFG VRNS K RGAAG SLLN VL NYP GMN NH VE II GVL RD ECAAM LCDF Y 180  
 R HPR QV FN AL KKN AGT INT Q SG GSS GG SSG SET PG TSE SA TPE SGG SSG GSM RE LD Y RI 240  
 GLA IGT N S I G W G VIEL SW NK DR ERY EK VRI VD QG V R M P D R A E M P K T G A S L A E P R R I A R S S 300  
 R R R L N R K S Q R K K N I R N L L V Q H G V I T Q E E D L S I Y P L S K K S M D I W G I R L D G L D R L L N H F E W A 360  
 R L L I H L A Q R R G F K S N R K S E L K D T E T G K V L S S I Q L N E K R S L E L Y R T V G E M W M K D P D F S K Y D R 420  
 K R N S P N E Y V F V S V S R A B L E K E I V T L F A A Q R R F Q S P Y A K D L Q E T Y L Q I W T H Q L P F A S G N A I 480  
 L N K V G Y C S L L K G K E R R I P K A T Y T F Q Y F S A L D Q V N R T R L G P D F Q P F T K E Q R E I L N N M F Q R 540  
 T D Y K K K T I P E V T Y D I R K W L E L D E T I Q F K G L N Y D P N E E L K K I E K K P F I N L K A F Y E I N K V 600  
 V A N Y S E R T N E T F S T L D Y D G I G Y A L T V Y K T D K D I R S Y L K S S H N L P K R C Y D D Q L I E L L S L S 660  
 Y T K F G H L S L K A I N H V L S I M Q K G N T Y K E A V D Q L G Y D T S G L K K E K R S K F L P P I S D E I T N P I V 720  
 K R A L T Q A R K V V N A I I R R H G S P H S V H E L A R E L S K N H D E R T K I V S A Q D E N Y K K N G A I S I L 780  
 S E H G L N P T G Y D I V R Y K L W K E Q G E R C A Y S L K E I P A D T F F N E L K K E R N G A P I L E V D H I L P Y 840  
 S Q S F I D S Y H N K V L V Y S D E N R K K G N R I P Y T Y F L E T N K D W E A F E R Y V R S N K F F S K K K R E Y L L 900  
 K R A Y L P R E S E L I K E R H L N D T R Y A S T F L K N F I E Q N L Q F K E A E D N P R K R R V Q T V N G V I T A H F 960  
 R K R W G L E K D R Q E T Y L H H A M D A I I V A C T D H H M V T R V T E Y Y Q I K E S N K S V K K P Y F P M P W E G F 1020  
 R D E L L S H L A S Q P I A K K I S E E L K A G Y Q S L D Y I F V S R M P K R S I T G A H Q Q T I M R K G G I D K K G 1080

-continued

---

KTIIIERLHL KDIKFDENGD FKMVGKEQDM ATYEAIKQRY LEHGKNKKA FETPLYKPSK	1140
KGTGNLTKRV KVEGQAKSFV REVNGVVAQN GDLVRVDSL E KDDKYYMVPY YVPDTVCSEL	1200
PKKVVASSKG YEQWLTLDSN FTFKFSLYPY DLVRLVKGDE DRFLYFGTLD IDSDRNLNFKD	1260
VNKPSKNEY RYSLKTIEDL EKYEVGVLGD LRLVRKETRR NFHSGGSKRP AATKKAGQAK	1320
KKK	1323

SEQ ID NO: 493	moltype = AA length = 1320
FEATURE	Location/Qualifiers
REGION	1..1320
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1320
	note = source = /note="LPG50145-nAPG07433.1protein sequence"
source	1..1320
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 493	
MAPKKKRKV D YKDHDGDKD HDIDYKDDDD KMSDTELNEH YWMRH ALMLA KRAR DEGEV P	60
VGAVLVLNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV LQNYRLIDTT LYVTF PCVM	120
CAGAMVHSRI GHLVFGVRNS KRGAGSLIN VLNV PGHMNR VEIAEGVLAD ECSA MLCDFY	180
RHPRQVF NAL KQA AKHISGG SSGGSSGSET PGTSE SATPE SS GGSSGGSM RELDYRIGLA	240
IGTNSIGWGV IELSWN KDR E RYE KIVR DQ GVR MF DRAEM PKTG ASLAEP RRIAR SRRR	300
LN RKS QRKKN I RNLL VQHGV IT QEE LDLSY PL SKKSMDL GIR LDGL DR L NHF EWAR LL	360
IHLAQQRGFK SNRK SELK D ETG KV LSS IQ LNE KRLS LYR TV GEM MMKD P DFS K YDR KRN	420
SPNEYVFSVS RAE LEKE I VT LF AA QRRF QFS PY ASK DKL QET YL QI WTH QLP FAS GNAIL N	480
VG YC SLLKGK ER RI PKAT YT FQ YF SALD QV NR TRL GPD FQ PFT KE QRE II LNN M PQT DY	540
YKK KTI PEVT YD I R KWL E DETI QFK GLN YDP NEEL KK E KKP F INLK A FYE I KV VAN	600
Y SERT NET FTS TLD YD GIG YA LT VY KTD KDI RS YL KSS HNL PK CY DDQ LI E ELL SLS YTK	660
F GHLS LKAIN HV L SIM QKG N TYKE A VD QL G DQ RSK F LPPI SD EIT NP IV KRA	720
L TQARK VVNA IIRR HGSPH V HI ELARE L KNH DERT KIV SA QDEN YKKN KGA ISIL SEH	780
G ILN PTK VNA I RY KLW KEQG E RCAY SLK EI PAD TFF NEL KERN GA PILE VDH IL PYS QQS	840
FIDS YH NKV I VVS DEN RKK G NRP Y TPL E TNK DW EAPER VRS NKF FSK KK REY LLK RA	900
YLP RESE LIK ER HLN DTRY A STFL KNFIE Q NLQ FKEA EDN PR KRR VQT VN GV I TAH FR KR	960
WG LEK DRO ET YL HHAM DAII VACT DHM VTI R VT E YY QI KE SNK SVK KPY F PMP WEG FR D E	1020
L LSH LAS QPDI AK KISE EEL KA GY QSL D YF SR MP KRSIT G AA HK QT IM RK GG ID KKG KTI	1080
I IER LH KDV DI KFD ENGD FKM VKE QDM ATY EA IK QRY L EH G KNS KKA FET PLY KPS KKG T	1140
GNL I KRV KVE QO AKS FV REV NGV A QN GDL VR DLF EK D KYY M PIV YP DTV C ELP PK	1200
VVASS KG YEQ WLT LD NSFT F KFS L YP DL R LV KG D EDF R LY FG TLD I DS DRL NF DV NK	1260
PSK NEY RYS LKT IE D LE KY EV GVL GD RL VR KET RR NF DV NK SF GSK RPA AT KK AG Q	1320

SEQ ID NO: 494	moltype = AA length = 1325
FEATURE	Location/Qualifiers
REGION	1..1325
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1325
	note = source = /note="LPG50146-nAPG07433.1protein sequence"
source	1..1325
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 494	
MAPKKKRKV D YKDHDGDKD HDIDYKDDDD KMSDIELNEH YWMRH ALMLA KRAREE GEV P	60
VGAVLVLNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV LQNYRLIDTT LYVTF PCVM	120
CAGAMVHSRI GHLVFGVRNS KRGAGSLIN VLNV PGHMNR IE FTE GVLAD ECSA MLCDFY	180
RYPRQVF NAL KQA AKI NPA A QSGGSSGG SSG SET PG TSE SAT PESS GG SGG S MREL D	240
RIGLAIGTNS I GVG VIEL SW NKDR EY TN S RIV DQ GVR MF DRA EM PK TG A SLAE P RRI AR	300
SSR RRL NR KS QR KKN IRN LL VQ HVI TQEE LD SLY PLS KK SMDI WGI RL D GL DR LL NH F E	360
WAR LLI HLA Q RRG FKS NR KS EL KDT ETG KV LSS IQL NEK LSL YRT VGEM WM KDP DF SKY	420
DR KRNS PNEY VFS VSA RAE L KEI VTL F A Q RRF QSPY ASK DL QET YL QI W TH QL F PAS GN	480
A ILN KVG YCS LL KG KERR IP KAT YD VQ FCS ALD QVN RTR L GDF QPFT KE QRE II LN NM F	540
Q RTD YY KKK T IPE VTY Y DIR KWL EDET I Q FK GLN YD PNE EL KK I EKK P F IN LK AF YE IN	600
KV VANY SERT NET FST LD YD GIG YALT VY K TD KDI RSY LK S SHN L PK R CY DD QL IE ELL S	660
L SYT KFG HLS L KAI NHV L SI MQ KGN TY KEA VD QL GY DT SG L KKE KRS KFL PP IS D EIT N P	720
IV KRAL TQ AR KV VNA I RY K GSP H VME L ARE LSK HDE RT KIV SA QD NY KKN KGA IS	780
I LSE H GIL N P TG YD I VY K RL WKE Q H GRC AY SL K EIP ADT F F NEL K KER NG A PI LEV D HIL	840
P YSQ SF DIDS Y HNK VLV SDE NR KKG N RI PY T YF LET N KDW EA FER VY RSN KFF SK KRE Y	900
L L KRAY L PRE SEL I KER HLN DTR YAST FLK N FIE QN L QF K EA EDN PR K RR VQT VNG VITA	960
HFR KRW GLE K DR QET YL HHA MDA II VACT D HHM VTR VTE Y QI KES N KSV KKP Y PPM PWE	1020
GFR DELL SHL AS QP IAK KIS EEL KAG Y QSL DY IF VSR MPK RS IT GAA HK Q TIM RKG GID K	1080
KG KTI III ER L HL KDI KFD EN GDF K M V GKE Q DMAT Y EA IK Q RY LEH G KNS K KAF ET PLY K P	1140
S KKG GTGN L K RV K VEG QAK S F VRE VNG VVA QNG DL V R V D L F EK D KYY M PIV Y P DTV CS	1200
E LP KKV VASS KGY EQWL TL D NSFT KFS L Y P D V L V R L V K G DED R FLY FGT LD D S D R L NF	1260
KDV N KPS KKN EY RY S L K TIE D LE KY EV GVL GDL R L V R KET RR NF HS GG SK RPA AT KK AG Q	1320
AKKK	1325

SEQ ID NO: 495	moltype = AA length = 1327
----------------	----------------------------

-continued

---

FEATURE REGION	Location/Qualifiers 1..1327 note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1327 note = source = /note="LPG50147-nAPG07433.1protein sequence"
source	1..1327 mol_type = protein organism = synthetic construct
SEQUENCE: 495	
MAPKKKRKV D YKDHDGYKD HDIDYKDDDD KMSIPELNHD VWMRHALT LA KRAREEGEVP 60 VGAVLVLNQ Q VIGEGWNRAI GLHDPATAHE IMALRQGLV LQNYRLIDTT LYVTFEPVCVM 120 CAGAMVHSRI GQLVFGVRNS KRGAAAGSLMN VLNPYGMNH R VEITEGVLRD ECAAMLCDFY 180 RQPQRQVNAL KKPAGDINAL QNNRSGGSSG GSSGSETPGT SESATPESSG GSSGGSREL 240 DYRIGLAIGT NSIGWGVI E SWNKDREREY KVRIVDQGVR MFDRDRAEMPKT GASLAEPRI 300 ARSSSRRRLNR KSQRKKNIRN LLVQHGVIT Q EELDSLWPLS KKSMDIWGIR LDGLDRLLNH 360 FEWARLLIHL AQRRGFKS NR KSELKDTEG KVLSIQLNE KRLSLYRTVG EMWMKDPDFS 420 KYDRKRNSPN BYVFSVSR AE LEKEIVTLFA AQRRFQSPYA SKDLQETYLQ IWTHQLPFAS 480 GNAILNKG Y C SLLKGKERR I PIPKATYTFQY FSALDQVNRT RLGPDFQPF KEQREIIILNN 540 MFORTDYYKK KTIPEVTTYD IRKWELELD T IQFKGLNYPD NEELKKIEKK PFINLKAFYE 600 INKVVANYSE RTNETFSTLD YDGIGYALT V YKTDKDIRSY LKSSHNLPKR CYDDQLIEEL 660 LSLSYTKEFGH LSLKAHNVL SIMQKNTYK EAVDQLGYDT SGLKKEKR SK FLPPISDEIT 720 NPIVKRALT Q ARKVVNAIIR RHGSPHSVH ELARELSKNH DERTKIVSAQ DENYKKNKGA 780 ISILSEHGIL NPTGYDIVR K LWKEQGERC AYSLKEIPAD TFFFNELKKER NGAPILEVHD 840 ILIPYSQSFD SYHNKVLVYS DENRKKG NR PYTYFLETNK DWEAFERYVR SNKFFSKKKR 900 EYLLKRAYLP RESELIKERH LNDTRYASTP LKNFIEQNLIQ FKEAEDNPRK RRVQTVNNGVI 960 TAHFRKRWGL BKDRQETYLH HAMDAIVAC TDHHMVTRV T EYYQIKESNK SVKKPYFPMP 1020 WEGFRDELLS HLASQPIAKK I SEELKAGYQ SL DYIFVFSRM PKRSITGA AH KQTIMRKGGI 1080 DKKGKTIIIE RLHLKD I KFD ENGDFKMVGK EQDMATYEAI KQRYLEHGKN SKKAFETPLY 1140 KPSKKGTG NL I KRVKVEGQA KSFVREVN GNG V A QNGD LVRV DLFEKDDKYY MVPIV PDTV 1200 CSELPKKVVA SSKGYEOWLT LDNSFTFKFS LYPYDLVRLV KGDED RFLYF GTLDIDSDRL 1260 NPKDVNPKSK KNEYRYSLKT I E DLEK YEVG VLGLDRLV RK ETRRN FHSGG SKRPAATKKA 1320 GQAKKKK 1327	
SEQ ID NO: 496	moltype = AA length = 1322
FEATURE REGION	Location/Qualifiers 1..1322 note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1322 note = source = /note="LPG50148-nAPG07433.1protein sequence"
source	1..1322 mol_type = protein organism = synthetic construct
SEQUENCE: 496	
MAPKKKRKV D YKDHDGYKD HDIDYKDDDD KMSNPEFTHE YWMRHALT LA RRARDEGEVP 60 VGAVLVLNQ Q VIGEGWNRAI GLHDPATAHE IMALRQGLV LQNYRLIDTT LYVTFEPVCVM 120 CAGAMVHSRI GQLVFGVRNS KRGAAAGSLMN VLNPYGMNH R VKTIGGV LAP ECSG LLCDFY 180 RMPRQVFNQQ KAE LKSINDS GGSSGSSG ETPGTSESAT PESSGGSSGG SMREL D YRIG 240 LAIGTNSIGW G VIEL SWNKD RERYE KVRIV DQGV RMFDRA EMPKT GASLA EPRRIARSSR 300 RRLNRK SQRK KNIRN LVLQV GH G VITRQ ELS L YPLS KKSMD IWGIR LDGLD RLL NHF EWAR 360 L L IHLA QRRG FKS NRK SELK DTETG KV LSS I QLN EK RSL YRTV GEM WMK DP DF SKY DRK 420 RNSPNEVFS VSRAE LEKEI V TFLFAA QRSS QSPY ASK D LQ ETYLQ IWT HQ LPF ASGNAIL 480 NKVG YCS LK GKERR IPKAT YTFQ YFS ALD QVN RTRL GP FQP FTKE QRE I IL NN M F QRT 540 DYYKKK TIPE V TYD Y DIRK WL ELD EY KF QK LG NYD PNEEL K KIE KK PFI NL KAF YE IN KV 600 AN YSERT NET F STL D YD GIG Y ALTV YK TDK DIRS YL KSH N LPK R CY DDQ LIE ELL SLSY 660 TKPGHLS LKA I NHV LSI MQ G NTY KEA VDQ LG YD TS GLK K EKR SKF LP PI SDE ITN PIV K 720 RALT QARK VV NAI I RR HG SP HSV HIEL ARE LSK NHD ERTK IV SAQ DEN YK KNKG AIS ILS 780 E H GIL NPT GY D IV RY K L W K E OG CAY S L K EIP ADT F FNE L K KER NG API LEV D HIL PYS 840 QSFIDSYH NK LV VY SD EN RK KGN RPI YF LET N K DWE AF ERY VRSN KFF SKK R EY LLK 900 RAY LPRESEL I KER HLN DTR YAST FLK NI E QNL QF KBE A DN PRK R RV Q VNG VITAH FR 960 KR WG L EK DR Q ETYL HHAM DA I IVA CT DHM V T R V TE YY QI KES NKS V K KP YFP MP WEG FR 1020 DELL SHLAS Q PIA KKI SEEL KAG Y QSL D YI F VSR MP KRSI TGA AH KQ TIM RKG GI D KKG K 1080 TII IER LH LK DI KFD E N GDF KM V GKE QDMA T YEA I K Q R YL EHG KNS KKA F ETPL YK PSK K 1140 GTGN LIK RVK VEG QAK SF VR EV NGG VQA QNG DL V R VD L F EK DD K Y Y M VPI Y VP D T V C S ELP 1200 KKV VASS KGY E QWL TLD NSF TFKF SLY PVD L VRL VKG DED RFLY FGT LDI DSD R LNF KDV 1260 NPKSKKNEYR YSL KTI EDLE KYEV GVL GDL R LVR KET RRN FHSGG SKR PA ATK KAG QAK K 1320 KK 1322	
SEQ ID NO: 497	moltype = AA length = 1321
FEATURE REGION	Location/Qualifiers 1..1321 note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1321 note = source = /note="LPG50149-nAPG07433.1protein sequence"

---

---

-continued

---

```

source          1..1321
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 497
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSDAELTHE YWMRHALTAA QRARDEGEVP 60
VGAVLVLNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV QONYRLDDTT LYVTPEPCVM 120
CAGAMVHSRI GRLIFGVRNS KRGAAAGSLIN VLNVPGMNRH VEVVEGILRD ECAGMLCDFY 180
RQPRQVNAL KKGATDINSQ GSSGGSSGE TPGTSESATP ESSGGSSGS MRELDYRIGL 240
AIGTNSIGWG VIELSWNKDR ERYEKVIRVD QGVRMFDRAE MPKTGASLAE PRRIARSRR 300
RINRKSRKQK NIRRNLVQHQ VITOQEELDSL YPLSKKSMDI WGIRLGLDR LLNHPREWRL 360
LIHLAQRRGF SKNRKSELKD TETGKVLSI QLNEKRLSLY RTVGEWMKD PDFSKYDRKR 420
NSPNEYFVSV SRAELEKEIV TLFAAQRRFQ SPYASKDLQE TYLQIWTQL PFASGNAILN 480
KVGYCSSLKG KERRIPKATY TFQYFSALDQ VNRTTRLGPDF QPFTKEQREI ILNNMFQRTD 540
YYKKKTIPEV YYDIRKWL LDETIQFKGL NYDPNEELKK IEKKPFINLK AFYEINKVVA 600
NYSSERTNETF STLDYDGIGY ALTVYKTDKD IRSYLNKSHN LPKRCYDDQL IEELLSLSYT 660
KFGHLSLAI NHVLSIMQKG NTYKEAVDQL GYDTSGLKEE KRSKFLPPIS DEITNPIVKR 720
ALTQARKVNV AIIRRHGSPH SVHIELAREL SKNHDERTKI VSAQDENYKK NGKAISILSE 780
HGILNPCTYD IVRYKLWKEQ GERCAYSLKE IPADTFFNELL KKERNAGAPIL EVDHILPYSQ 840
SFIDSYHNK LVYSDENRKK GNRIPYTYFL ETNPKDWEAFA RYVRSNKFSS KKKREYLK 900
AYLPRESELI KERHLNDTRY ASTFLKNFIE QNLQFKEAAD PNRKRRVQTV NGVITAHFRK 960
RWGLEKDRQE TYLHHAMDAI IVACTDHMHV TRVTEYYQIK ESNKSVKPY FPMPWEGFRD 1020
ELLSHLASQP IAKKISEELK AGYQSALDQ VSRMPKRSIT GAHKQTIMR KGGIDKKKGKT 1080
IIIERLHLKD IKFDENGDFK MVGKEQDMAT YEAIKQRYLE HGKNSKKAFE TPLYKPSKKG 1140
TGNLIKRVKV EGQAKSFVRE VNGGVAQNGD LVRVDSLFEKD DKYMMVPIYV PDTVCSELPK 1200
KVASSKGYE QWLTLDSNFT FKFSLYPYDL VRLVKGDEDR FLYFGTLDID SDRLNFKDVN 1260
KPSKKNEYRY SLKTIEDLEK YEVGVLGDLR LVRKETRRNF HSGGSKRPAA TKKAGQAKKK 1320
K                                         1321

SEQ ID NO: 498      moltype = AA length = 1321
FEATURE           Location/Qualifiers
REGION            1..1321
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION            1..1321
note = source = /note="LPG50150-nAPG07433.1protein sequence"
source             1..1321
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 498
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSDAELTHE YWMRHALTAA QRARDEGEVP 60
VGAVLVLNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV QONYRLDDTT LYVTPEPCVM 120
CAGAMVHSRI GRLIFGVRNS KRGAAAGSLIN VLNVPGMNRH VEVVEGILRD ECAGMLCDFY 180
RQPRQVNAL KKGATDVLG SSSGGSSGE TPGTSESATP ESSGGSSGS MRELDYRIGL 240
AIGTNSIGWG VIELSWNKDR ERYEKVIRVD QGVRMFDRAE MPKTGASLAE PRRIARSRR 300
RINRKSRKQK NIRRNLVQHQ VITOQEELDSL YPLSKKSMDI WGIRLGLDR LLNHPREWRL 360
LIHLAQRRGF SKNRKSELKD TETGKVLSI QLNEKRLSLY RTVGEWMKD PDFSKYDRKR 420
NSPNEYFVSV SRAELEKEIV TLFAAQRRFQ SPYASKDLQE TYLQIWTQL PFASGNAILN 480
KVGYCSSLKG KERRIPKATY TFQYFSALDQ VNRTTRLGPDF QPFTKEQREI ILNNMFQRTD 540
YYKKKTIPEV YYDIRKWL LDETIQFKGL NYDPNEELKK IEKKPFINLK AFYEINKVVA 600
NYSSERTNETF STLDYDGIGY ALTVYKTDKD IRSYLNKSHN LPKRCYDDQL IEELLSLSYT 660
KFGHLSLAI NHVLSIMQKG NTYKEAVDQL GYDTSGLKEE KRSKFLPPIS DEITNPIVKR 720
ALTQARKVNV AIIRRHGSPH SVHIELAREL SKNHDERTKI VSAQDENYKK NGKAISILSE 780
HGILNPCTYD IVRYKLWKEQ GERCAYSLKE IPADTFFNELL KKERNAGAPIL EVDHILPYSQ 840
SFIDSYHNK LVYSDENRKK GNRIPYTYFL ETNPKDWEAFA RYVRSNKFSS KKKREYLK 900
AYLPRESELI KERHLNDTRY ASTFLKNFIE QNLQFKEAAD PNRKRRVQTV NGVITAHFRK 960
RWGLEKDRQE TYLHHAMDAI IVACTDHMHV TRVTEYYQIK ESNKSVKPY FPMPWEGFRD 1020
ELLSHLASQP IAKKISEELK AGYQSALDQ VSRMPKRSIT GAHKQTIMR KGGIDKKKGKT 1080
IIIERLHLKD IKFDENGDFK MVGKEQDMAT YEAIKQRYLE HGKNSKKAFE TPLYKPSKKG 1140
TGNLIKRVKV EGQAKSFVRE VNGGVAQNGD LVRVDSLFEKD DKYMMVPIYV PDTVCSELPK 1200
KVASSKGYE QWLTLDSNFT FKFSLYPYDL VRLVKGDEDR FLYFGTLDID SDRLNFKDVN 1260
KPSKKNEYRY SLKTIEDLEK YEVGVLGDLR LVRKETRRNF HSGGSKRPAA TKKAGQAKKK 1320
K                                         1321

SEQ ID NO: 499      moltype = AA length = 1323
FEATURE           Location/Qualifiers
REGION            1..1323
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION            1..1323
note = source = /note="LPG50151-nAPG07433.1protein sequence"
source             1..1323
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 499
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSDLELNH YWMRHALQLA QRARDEGEVP 60
VGAVLVYNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV QONYRLDDTT LYVTPEPCVM 120

```

---

-continued

---

CSGAMVHSRI	GTLVFGVRNE	KRGAAGSLMN	VLRYPGMNHQ	VQIIDGV LAP	ECSGLLCDFY	180
CMPRQQQNQQ	KAESTSPGD	SGGSSGGSSG	SETPGTSESA	TPESSGGSSG	GSMRELDYRI	240
GLAIGTNSIG	WGVIETSWNK	DRERYEKVRI	VDQGVRMFDR	AEMPKTGASL	AEPRIARSS	300
RRLRNRSQR	KKNIRNLVQ	HGVITQEELD	SLYPLSKKSM	DIWGIRLDGL	DRLLNHFEWA	360
RLLIHLAQRR	GFKSNRKSEL	KDTETGKVL	SIQLNEKRLS	LYRTVGEMWM	KDPDFSKYDR	420
KRNSPNEYVF	SVSRAELEKE	IVTLEFAAQR	FQSPYASKDL	QETYLIQWTH	QLPFASGNAI	480
LNKVGYCSLL	KGKERRIPKA	TYTFQYFSAL	DQVNRTRLGP	DFQPFTEQR	EIILNMMFQR	540
TDYKKKTTIP	EVTYYDIRK	LELDETIQPK	GLNYDPNEEL	KKIEKKPFIN	LKAFYEINKV	600
VANYSSERTNE	TFSTLDYDG	GYALTVYKT	KDIRSYLKSS	HNLPKRCYDD	QLIEELLSLS	660
YTKEFGHLSL	A1NHVLSIMQ	KGTYKEAVD	QLGYDTGSLK	KEKRSKFLPP	ISDEITNPIV	720
KRALTQARV	VNAIIRRHGS	PHSVHIELAR	ELSKNHDERT	KIVSAQDENY	KKNKGAI SIL	780
SEHGIILNPTG	YDIVRYKLW	EQGERCAYSL	KEIPADTFFN	ELKKERNNGAP	ILEVDHILPY	840
SQSFIDSYHN	KVLVYSDENR	KKGNRIPYTY	FLETNKDWEA	FERYVRSNKF	FSKKKREYLL	900
KRAYLPRESE	LIKERHLNDT	RYASTFLKNF	IEONLQFKEA	EDNPKRKRVQ	TVNGVITAHF	960
RKRWGLEKDR	QETYLHHAMD	AIIAVTRTEYYQ	IKESENKSVKK	PYFPMPWEGF	1020	
RDELLSHLAS	QPIAKKISEE	FTFKFSLYPY	DLVRLVKGDE	DRFLYFGTLD	IDSRLNFKD	1260
PKVVASSKKG	YEQWLTLDNS	TFKFSLYPY	DLVRLVKGDE	DRFLYFGTLD	IDSRLNFKD	1260
VNKPSKNEY	RYSLKTIEDL	EKYEVGLGD	LRLVRKETRR	NPHSGGSKR P	AATKKAGQAK	1320
KKK						1323

```

SEQ ID NO: 500      moltype = AA length = 1321
FEATURE           Location/Qualifiers
REGION            1..1321
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION            1..1321
note = source = /note="LPG50152-nAPG07433.1protein sequence"
source             1..1321
mol_type = protein
organism = synthetic construct

```

SEQUENCE: 500						
MAPKKKRKV D YKDHDG DYKD	H DIDYKDDDD	KMSDNE LNHE	YWMRH ALGLA	KRAREEGEV P	60	
VGAVLV LNNQ V IGE GWN RAI	G LHDPTA HAE	I MALR QGGLV	L QNYR LLDT T	LYVT FEP CVM	120	
CAGAMVHSRI GTLVFGVR NS	KRGAAGSL MN	V LNYPGM NHQ	V EIVEG ILSE	SCA MLCD FY	180	
RQPRAVK NAL KKAADPA ASG	GSSGGSSG SE	T PGTSESAT P	E SSGGSSG GS	M RE LDYR IGL	240	
AIGTNSIG W VIEL SWN KDR	E RYK VRI VD	Q GVR MF DRA E	M PKT GAS LAE	P RR IAR SRR	300	
RLLNRK SQR KK NIRM LLV QHG	V ITQ EELDSL	Y PL SKKS MDI	W GIRD GLDR	L LNHP EWAR L	360	
LIHLA QR GFG KSNR KSEL KE	T ETG KV LSS I	Q LNE KRL SLY	R TV GEW M K D	P DF SKY DR KR	420	
NSPNEYF VSV SRAE LEK IV	T LFAA Q RRF Q	S P Y ASK DL Q E	T YLQ IWT HQL	P FAS GNA IL N	480	
KVGYCSLLKG KERR IPK AT	TF QY FS ALD Q	V NR TR L GPD F	Q PFT K E Q R E I	I LNN M F Q R T D	540	
YYKKKT IPE V YY DIR K W	L D E T I Q F K G L	N YDP NE EL K K	I E K K P F I N L K	A F Y E I N K V V A	600	
NY SERT NET F KER H L ND T R Y	A S T F L K N F I E	Q NL Q F K E A E D	N P R K R V Q T V	N G V I T A H F R K	660	
RW GLE KDR Q E TYL HHAM DAI	I VACT DHH M V	Y DS YL K S H N	L P K R C Y D D Q L	I E E L L S L S Y T	660	
ELL SH LAS Q P I AK KI SE E L	A G Y Q S L D Y I F	V S R M P K R S I T	G A A H Q T I M R	K G G I D K K G K T	720	
III ER L H L K D I K F D E N G D F K	M V G K E Q D M A T	Y E A I K Q R Y L E	H G K N S K K A F E	T P L Y K P S K K G	1140	
T G N L I K R V K V E Q Q A K S F V R E	V N G V V A Q N G D	L V R V D L F E K D	D K Y Y M V P I Y V	P D T V C S E L P K	1200	
KV VASS K G Y E Q W L T D N S F T	F K F S L Y P Y D L	V R L V K G D E D R	F L Y F G T L D I D	S D R L N F K D V N	1260	
KPSK KNEY RY SLK TIE D L E K	Y E V G V L G D L R	L V R K E T R R N F	H S G G S K R P A A	T K K A G Q A K K K	1320	
K						1321

```

SEQ ID NO: 501      moltype = AA length = 1318
FEATURE           Location/Qualifiers
REGION            1..1318
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION            1..1318
note = source = /note="LPG50153-nAPG07433.1protein sequence"
source             1..1318
mol_type = protein
organism = synthetic construct

```

SEQUENCE: 501						
MAPKKKRKV D YKDHDG DYKD	H DIDYKDDDD	KMSDTE FTHE	H WMR HAL TL A	QRAR DEGE VP	60	
VGAVLV LNNQ V IGE GWN RAI	G LHDPTA HAE	I MALR QGGLV	L QNYR LLDT T	LYVT FEP CVM	120	
CAGAMVHSRI GHLVFGVR NS	KRGAAGSL MN	V LGYPGM NHQ	V QVSEG VLAT	ECSMLCD FY	180	
RAPRLV K NAL KEK AR SGG SS	GGSSGSET PG	T SESAT PESS	GGSSGGSMRE	L DYR IGL AIG	240	
TNSIGWVIE L SWN KDR E Y	E KV RIV DQ CV	R MFD RA E M P K	T GAS L AEP RR	I AR S R R R L N	300	
RKSQRKKNIR NLLVQHGVIT Q E E L D S L Y P L	S KKSMDI W G I	R LD GLD R L N	H F E W A R L L I H		360	
LAQRRG F K S N R K S E L K D T E T	E K V L S S I Q L N	E K R L S L Y R T V	G E M W M K D P D F	S K Y D R K R N S P	420	
NEYVFSV SRA E L E K I V T L F	A A Q R F Q S P Y	A S K D L Q E T Y L	Q I W T H Q L P F A	S G N A I L N K V G	480	

-continued

---

YCSLLKGKER	RIPKATYTTFQ	YFSALDQVNR	TRLGPDFQPF	TKEQREIILN	NMFQRTDYYK	540
KKTIPEVTYY	DIRKWLLELD	TIQFKGLNYD	PNEELKKIEK	KPFINLKAFY	EINKVANYS	600
ERTNETFSTL	DYDGIGYALT	VYKTDKDIRS	YLKSSHNLPK	RCYDDQLIEE	LLSLSYTKFG	660
HISLKLAINHV	LSIMQKGNTY	KEAVDQLGYD	TSGLKKEKRS	KFLPPISDEI	TNPIVKRALT	720
QARKVVNAII	RRHGSPhSVH	IELARELSKN	HDERTKIVSA	QDENYKKNKG	AISILSEHGI	780
LNPTGYDIVR	YKLWKEQGER	CAYSLKEIPA	DTFFNELKKE	RNGAPILEVD	HILPYSQSFI	840
DSYHNKVLVY	SDENRKKGNR	IPYTYFLETN	KDWEAFERYV	RSNKFFSKKK	REYLLKRAYL	900
PRESELIKER	HLNDTRYAST	FLKNFIEQNL	QFKEAEDNPR	KRRVQTVNGV	ITAHFRKRWG	960
LEKDRQETYL	HHAMDAIIVA	CTDHMMVTRV	TEYYQIKESN	KSVKPYFPM	PWEGRDELL	1020
SHLASPIAK	KISEELKAGY	QSLDYIFVR	MPKRSITGAA	HQQTMRKGG	IDKKGKTTIII	1080
ERLHLKDIKF	DENGDPKMVG	KEQDMATYE	IKQRYLEHKGK	NSKKAFTPL	YKPSKKGTCN	1140
LIKRVKVEGQ	AKSFVREVNG	GVAQNGDLVR	VDLFEKDKEY	YMPVIVPDT	VCSELPKVU	1200
ASSKGYEQWL	TLDNSFTFKF	SLYPYDLVRL	VKGDEDRFLY	FGTLIDSDR	LNFKDVKPS	1260
KKNEYRYSLK	TIEDLEYEV	GVLGDLRLVR	KETRRRNHFSG	GSKRPAATKK	AGQAKKK	1318

SEQ ID NO: 502	moltype = AA	length = 1325
FEATURE	Location/Qualifiers	
REGION	1..1325	
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION	1..1325	
source	note = source = /note="LPG50154-nAPG07433.1protein sequence" 1..1325	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 502		
MAPKKKRKV D YKDHDG DYKD HDIDYK DDDD KMS ESE FTHE HWMRH ALT LA RR AREE GE VP	60	
VGA VLV LNN Q VIGEG WNR A I GLH DPTA HAE IM AL RQ GGL V LQ NYR LL D ST LY VTF PC VM	120	
CAG AMV HGR I GNL VPG VRS N KRG A GS LMN VVG YPG MNH Q INV IE GV LAE ECS AML CDF Y	180	
RAP RL VKN AL KEK ARGN NN P N KS GG SS GG S SG SET PG TSE SAT PE SSS G SGG S MRE LD Y	240	
RIGLA IG TNS I GWG VIEL SW NK DR ER YEV K RIV DQ GVR MF DRA EPM KT GA SLA EPR RI AR	300	
SS RRL NR KS Q KKN IRN LL VQ HG VIT QEE L DLS LY PLS KK SMD I WGI RL D GLD RLL NH F E	360	
WAR LLI HLA Q RRG FKS NR KS EL KDT ETG KV LSS I QL NE KR LSL Y RTV GEM WM KDP DF SKY	420	
DR KR NS PNEY VFS VSR A ELE KIV TL FA AQ RRF QSP V ASK D LQ ET YL QI W TH QLP FAS GN	480	
AIL NKV GY CS L LKG KERR IP KAT YD QY FS AL DQ VNR TRL GPD FQ PFT KE Q RE I I L N NM F	540	
Q RTD YY KKT I PEV TY YD IR KWL EDE TI F FK GLN Y DP NE EL KK I EK KP F IN LKA F YE IN	600	
KV VANY SERT NET FST LD YD GIG Y AL TV YK TD KDI RS YL KSSH NLP KRC Y DD QLIE ELL S	660	
L SYT KF GH LS LK AIN HV LSI M QKG NTY KEA VD QL GY DT SG LK KE KRS KFL PPI SDE IT NP	720	
I V KRAL TQ AR KV VNA I I RR GSP HVS HIE LARE LSK HNE RT KIV SA QD E NY KKN KG A IS	780	
I LSE HGL I LP N TGY DIV RY KL WKE Q GERC AY SL KE I PA DFT F NEL KER NG API LEV DH IL	840	
P YSQ FIDS Y HNK VLV Y SDE NR KKG NRI PY TF LETN KDW EAF ERY VRS N KFF SKK RE Y	900	
L LK RAY L PRE SEL I KER HLN D TRY A STF LK NF I EQ NL QFK EA E DNP R KRR V QT VNG VITA	960	
H FKR RW L EK DR Q ET Y LH HA MDA I T VACT D HHM VTR VTE Y QI KES NKS V KKP YP PM PWE	1020	
G FR DELL SH L AS QPI AKK IS EEL KAG Y QSL DY I F VSR MPK R SIT QGA HK Q TIM RKG GID K	1080	
K GKT III IER HL KDI KFD EN GDF KMV GKB Q DMAT YEA IK Q RYLE H GKN SK KAF E TPLY K P	1140	
S KKG TG N LIK RV KVEG QAK S FV REV NGG VA QNG DLV R VDL FE KDD KY MV PI YV PDT VCS	1200	
E LP KKV VASS KGY E QWL TLD NS TFK FSL Y PY DLV R LV KVG DED RFLY FGT LD I DSD R LNF	1260	
KDV NKPS KKN EY RY S L K TIE D LE K YEV GVL GDL RL VR KET RRN FH SG GS K RPA AT KAG Q	1320	
AKKK		1325

SEQ ID NO: 503	moltype = AA	length = 1318
FEATURE	Location/Qualifiers	
REGION	1..1318	
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION	1..1318	
source	note = source = /note="LPG50155-nAPG07433.1protein sequence" 1..1318	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 503		
MAPKKKRKV D YKDHDG DYKD HDIDYK DDDD KMS NPEL THE HWM RYALT LA KR AREE GE VP	60	
VGA VLV LNN Q VIGEG WNR A I GLH DPTA HAE IM AL RQ GGL V LQ NYR LL D ST LY VTF PC VM	120	
CAG AMV HSR I GOL VPG VRS N KRG A GS LMN VLN YPG MNH I EFT EG VL RD ECA AML CDF Y	180	
RQ PRL VKN AL KTG NGS GSS GG SSG SET PG TSE SAT PE SSS G SGG S MRE LD YR I GLA I G	240	
TNS I G WGV IE L SWN KDR EY E KRV IVD QGV RMP DRA E M P TGA S LAE PRR IAR S RRL N	300	
RKS QRK KN IR NLL VQ HG VIT QEE L D SLY PL SK K SMD I WGI RLD GLD RLL N H F E WAR L L I H	360	
LA Q RRG FKS N RKS E LKD TET GKV L SSI QLN E KRL SLY RT V GEM W M KDP DF SKY DR KRN SP	420	
NE YV F VS RA E L E K I V TLF AA Q RR FQ SPY AS KDL Q ET YL QI WTH QLP F A SGN A I L NK VG	480	
YCS L LKG KER RIP KAT YT TF Q YFS AL DQ VNR TRL GPD FQ PF TKE Q RE I I L N M F Q RT D YY K	540	
KK T IPE V TYY D IR KWL L ELD E TI QF KGL NYD PNE EL KK IE K P F I N LKA FY EINKVAN Y S	600	
ERT NET FST L DY DGIG YALT VY KTD KDIRS YL KSS HNL PK RCY DD QLIE E LLS LSY TKFG	660	
HIS LKL AIN HV L S I M QKG NTY K E A V D Q L G Y D T S G L K K E K R S K F L P P I S D E I T N P I V K R A L T	720	
QARK VV NA II RRG SP HS VH I E L A R E L S K N H D E R T K I V S A QDEN YKK NKG A I S I L S E H G I	780	
LN PTG Y DIV R YKL WKE QGER CAY SL K E I PA DTF FN E L K K RNG A P I L E V D H I L P Y P S F S I	840	
DSY HN KVL VY SD E N R KKG NR I P Y T Y F L E T N K D W E A F E R Y V R S N K F F S K K R E Y L L K R A Y L	900	

SEQ ID NO: 504	moltype = AA	length = 1318
FEATURE	Location/Qualifiers	
REGION	1..1318	
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"	
REGION	1..1318	
source	note = source = /note="LPG50155-nAPG07433.1protein sequence" 1..1318	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 504		
MAPKKKRKV D YKDHDG DYKD HDIDYK DDDD KMS NPEL THE HWM RYALT LA KR AREE GE VP	60	
VGA VLV LNN Q VIGEG WNR A I GLH DPTA HAE IM AL RQ GGL V LQ NYR LL D ST LY VTF PC VM	120	
CAG AMV HSR I GOL VPG VRS N KRG A GS LMN VLN YPG MNH I EFT EG VL RD ECA AML CDF Y	180	
RQ PRL VKN AL KTG NGS GSS GG SSG SET PG TSE SAT PE SSS G SGG S MRE LD YR I GLA I G	240	
TNS I G WGV IE L SWN KDR EY E KRV IVD QGV RMP DRA E M P TGA S LAE PRR IAR S RRL N	300	
RKS QRK KN IR NLL VQ HG VIT QEE L D SLY PL SK K SMD I WGI RLD GLD RLL N H F E WAR L L I H	360	
LA Q RRG FKS N RKS E LKD TET GKV L SSI QLN E KRL SLY RT V GEM W M KDP DF SKY DR KRN SP	420	
NE YV F VS RA E L E K I V TLF AA Q RR FQ SPY AS KDL Q ET YL QI WTH QLP F A SGN A I L NK VG	480	
YCS L LKG KER RIP KAT YT TF Q YFS AL DQ VNR TRL GPD FQ PF TKE Q RE I I L N M F Q RT D YY K	540	
KK T IPE V TYY D IR KWL L ELD E TI QF KGL NYD PNE EL KK IE K P F I N LKA FY EINKVAN Y S	600	
ERT NET FST L DY DGIG YALT VY KTD KDIRS YL KSS HNL PK RCY DD QLIE E LLS LSY TKFG	660	
HIS LKL AIN HV L S I M QKG NTY K E A V D Q L G Y D T S G L K K E K R S K F L P P I S D E I T N P I V K R A L T	720	
QARK VV NA II RRG SP HS VH I E L A R E L S K N H D E R T K I V S A QDEN YKK NKG A I S I L S E H G I	780	
LN PTG Y DIV R YKL WKE QGER CAY SL K E I PA DTF FN E L K K RNG A P I L E V D H I L P Y P S F S I	840	
DSY HN KVL VY SD E N R KKG NR I P Y T Y F L E T N K D W E A F E R Y V R S N K F F S K K R E Y L L K R A Y L	900	

-continued

---

PRESELIKER	HLNDTRYAST	FLKNFIEQNL	QFKEAEDNPR	KRRVQTNGV	ITAHFRKRWG	960
LEKDRQETYL	HHAMDAIIA	CTDHHMVTRV	TEYYQIKESN	KSVKKPYFPM	PWEGRFDELL	1020
SHLASQPIAK	KISEELKAGY	QSLDYIFVSR	MPKRSITGAA	HQQTIMRKGG	IDKKGKTIII	1080
ERLHLKDIKF	DENGDFKMVG	KEQDMATYEA	IQKORYLEHKG	NSKKAFETPL	YKPSKKGTTGN	1140
LIKRVKVEGQ	AKSFREVNG	GVAQNGDLVR	VDLFEKDCKY	YMVPIYVPDT	VCSELPKKVV	1200
ASSKGYEQWL	TLDNSFTFKF	SLYPYDLVRL	VKGDEDRLY	FGTLIDSDR	LNFKDVNPKS	1260
KKNEYRYSLK	TIEDLEKYEV	GVLGDLRLVR	KETRRRNFHSG	GSKRPAATKK	AGQAKKKK	1318

SEQ ID NO:	504	moltype = AA	length = 1320
FEATURE		Location/Qualifiers	
REGION		1..1320	
		note = source = /note="Description of Artificial Sequence:	
		Synthetic polypeptide"	
REGION		1..1320	
		note = source = /note="LPG50156-nAPG07433.1protein sequence"	
source		1..1320	
		mol_type = protein	
		organism = synthetic construct	

SEQUENCE:	504					
MAPKKKRKV	YKDHDGYKD	HDIDYKDDDD	KMSDPELNHE	YWMRHALQLA	KRAREEGEVP	60
VGAVLVLNQQ	VIGEGWNRAI	GLHDPTAAHE	IMALRQGGLV	LQNQYRLLDTT	LYVTFEPVCVM	120
CSGAMIHSRI	CTVVFGRNR	KRGAAGSSLIN	VLRYPGMNHQ	VNVLGGLAP	ACSEMLCEFY	180
RMPRQQKNRQ	KAESKLSSGG	SSGGSSGSET	PGTSESATPE	SSGGSSGGS	RELDYRIGLA	240
IGTNSIGWGV	IELSINKDRE	RYEKVIRVDQ	GVRMFDRADM	PKTGASLAEP	RRIARSSRRR	300
LNKRSQRKRN	IRNLLVQHGV	ITQEELDSLY	PLSKKSMDI	GIRLDGLDR	LNHFEWARLL	360
IHLAQRRGFK	SNRKSELKDT	ETGKVLISSIQ	LNEKRLSLYR	TVGEMWMKDP	DFSKYDRKRN	420
SPNEYVFSVS	RAELEKEIV	LFAAQRRFQS	PYASKDQ	YLOQIWTQLP	FASGNAILNK	480
VGYCSSLKKG	ERRIPKATYT	FQYFSALDQV	NRTRLGPDFQ	PFTKEQREII	LNNMQRDTY	540
YKKKTIPEVT	YYDIRKWEL	DETIOFKGLN	YDPNEELKKI	EKKPFINLKA	FYEINKVAN	600
YSERTNETFS	TLDYDGIGYA	LTIVYKTDKDI	RSYLKSSHNL	PKRCYDDQLI	EELLSLSYTK	660
FGHLSLKAIN	HVLSSIMQKGN	TYKEAVDQLG	YDTSGLKEK	RSKFLPPISD	EITNPIVKRA	720
LTCARKVUNA	IIRRHGSPHS	VHIELARELS	KHDERTIV	SAQDENYKKN	KGAISILSEH	780
GILNPTGYDI	VRYKLWKEQG	ERCA	YSLKEI	PADTFFNELK	KERNGAPILE	840
FIDSYHNKVL	VYSDENRKKG	NRIPYTYFLE	TNKDWEAPE	YVRSNKF	KKREYLLKRA	900
YLPRESELIK	ERHLDNTRYA	STFLKNFIEQ	NLQFKEAEDN	PRKRRVQTVN	GVITAHFRKR	960
WGLEKDRQET	YHHAMDAII	VACTDHMVT	RVTEYYQIKE	SNKSVKPYF	PMPWEGFRDE	1020
LISHLASQPI	AKKISEELKA	GYQSLDYIFV	SRMPKRSITG	AAHKQTMRK	GGIDKKGTI	1080
IIERLHLKDI	KFDENGDFKM	VGKEQDMATY	EAIKQRYLEH	GKNSKKAFET	PLYKPSKKGT	1140
GNLICKRVE	QOAKSFVREV	NGGVAQNGDL	VRVDSLFEKDD	KYVMVPIYVP	DTVCSELPKK	1200
VVASSKGYEQ	WLTLDNSFTF	KFSLYPYDLV	RLVKGDEDRF	LYFGTLDIDS	DRLNFKDVNK	1260
PSKKNEYRYS	LKTIEDLEKY	EVGVLGDLR	VRKETRRNFH	SGGSKRPAAT	KKAGQAKKK	1320

SEQ ID NO:	505	moltype = AA	length = 1320
FEATURE		Location/Qualifiers	
REGION		1..1320	
		note = source = /note="Description of Artificial Sequence:	
		Synthetic polypeptide"	
REGION		1..1320	
		note = source = /note="LPG50157-nAPG07433.1protein sequence"	
source		1..1320	
		mol_type = protein	
		organism = synthetic construct	

SEQUENCE:	505					
MAPKKKRKV	YKDHDGYKD	HDIDYKDDDD	KMSDNELNHE	HWMRHALLA	QRAREEGEVP	60
VGAVLVLNQQ	VIGEGWNRAI	GLHDPTAAHE	IMALRQGGMV	LQNQYRLLDTT	LYVTFEPVCVM	120
CAGAMVHSRI	GQLVFGVRNS	KRGAAGSSLIN	VLNYPGMNH	VEITEGVLA	DCSSMLCDFY	180
RHPREQKNAL	KRAHSNSGG	SSGGSSGSET	PGTSESATPE	SSGGSSGGS	RELDYRIGLA	240
IGTNSIGWGV	IELSINKDRE	RYEKVIRVDQ	GVRMFDRADM	PKTGASLAEP	RRIARSSRRR	300
LNKRSQRKRN	IRNLLVQHGV	ITQEELDSLY	PLSKKSMDI	GIRLDGLDR	LNHFEWARLL	360
IHLAQRRGFK	SNRKSELKDT	ETGKVLISSIQ	LNEKRLSLYR	TVGEMWMKDP	DFSKYDRKRN	420
SPNEYVFSVS	RAELEKEIV	LFAAQRRFQS	PYASKDQ	YLOQIWTQLP	FASGNAILNK	480
VGYCSSLKKG	ERRIPKATYT	FQYFSALDQV	NRTRLGPDFQ	PFTKEQREII	LNNMQRDTY	540
YKKKTIPEVT	YYDIRKWEL	DETIOFKGLN	YDPNEELKKI	EKKPFINLKA	FYEINKVAN	600
YSERTNETFS	TLDYDGIGYA	LTIVYKTDKDI	RSYLKSSHNL	PKRCYDDQLI	EELLSLSYTK	660
FGHLSLKAIN	HVLSSIMQKGN	TYKEAVDQLG	YDTSGLKEK	RSKFLPPISD	EITNPIVKRA	720
LTCARKVUNA	IIRRHGSPHS	VHIELARELS	KHDERTIV	SAQDENYKKN	KGAISILSEH	780
GILNPTGYDI	VRYKLWKEQG	ERCA	YSLKEI	PADTFFNELK	KERNGAPILE	840
FIDSYHNKVL	VYSDENRKKG	NRIPYTYFLE	TNKDWEAPE	YVRSNKF	KKREYLLKRA	900
YLPRESELIK	ERHLDNTRYA	STFLKNFIEQ	NLQFKEAEDN	PRKRRVQTVN	GVITAHFRKR	960
WGLEKDRQET	YHHAMDAII	VACTDHMVT	RVTEYYQIKE	SNKSVKPYF	PMPWEGFRDE	1020
LISHLASQPI	AKKISEELKA	GYQSLDYIFV	SRMPKRSITG	AAHKQTMRK	GGIDKKGTI	1080
IIERLHLKDI	KFDENGDFKM	VGKEQDMATY	EAIKQRYLEH	GKNSKKAFET	PLYKPSKKGT	1140
GNLICKRVE	QOAKSFVREV	NGGVAQNGDL	VRVDSLFEKDD	KYVMVPIYVP	DTVCSELPKK	1200
VVASSKGYEQ	WLTLDNSFTF	KFSLYPYDLV	RLVKGDEDRF	LYFGTLDIDS	DRLNFKDVNK	1260
PSKKNEYRYS	LKTIEDLEKY	EVGVLGDLR	VRKETRRNFH	SGGSKRPAAT	KKAGQAKKK	1320

-continued

---

SEQ ID NO: 506	moltype = AA length = 1322
FEATURE	Location/Qualifiers
REGION	1..1322
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1322
	note = source = /note="LPG50158-nAPG07433.1protein sequence"
source	1..1322
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 506	
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSNPEHHNE YWMRHALLA QRARDEGEVP 60	
VGAVLVYNQQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLLDTT LYVTFEPVCV 120	
CSGAMVHSRI CTLVFGVRNE KRGAAAGSLMN VLGYPGMNNHQ VOIIGGV LAP DCSGLLCDFY 180	
RMPRQKNNQQ ETPGTSESAT PESSGGSSGG SMRELDYRIG 240	
LAIGTNSIGW CVIELSWNKD RERYEKVIRV DQGVRMFDR A EMPKTGASLA EPRIARISSR 300	
RRLNRKSQRK KNIRNLVQH GVTQEELDS LYPLSKKSM D IWGIRLDGLD RLLNHFEWAR 360	
LILHLAQQRG FKSNRKSELE DTETGKVLS IQLNEKRLSL YRTVGEWMK DPDFS KYDRK 420	
RNSPNEYVFS VSRAELEKEI VTLFAAQRRF QSPYASKDLQ ETYLQIWT HQ LPFASGNAIL 480	
NKVGYCSLLK GKERRIPKAT YTQYQFSALD QVNRTTRLGP D QOPFTKEQRE II LNMMFQRT 540	
DYYKKKTIPE VTYYDIRKWL ELD E TIQFKG L NYD PNEELK KIEKKPFI NL KAFYEINKVV 600	
ANYSERTNET FSTLBYDGIG Y ALTVYKTDK DIRSYLKSH NLPKRCYDDQ LIEELL SLSY 660	
TKFGHLSLRK INHVLSIMQK GNTYKEAVDQ LGYDTSGLKK EKR SKFLPPI SDEITNP IVK 720	
RALTQARKVV NAIIRRHGSP HSVHIELARE LSKNHDERTK IVSAQDENYK KNKAISILS 780	
EHGILNPTGY DIVRYKLWKE QGERCAYSLK EIPADTF FNE LKKER NGAPI LEVDHILPYS 840	
QSFIDSYH NK VL VYSDENRK KG NRIPYTYF LETN KDW EAF ERYVRSN KFF SKKKREYLLK 900	
RAYL PRESEL I KERH LNDTR YASTFLKNFI EQNLQFKEAE DNPRK RRVQ VNGVITA HFR 960	
KR WGLEK DRQ ETYLHHAMDA IIVACTDHMM VTRVTEY YQI KESNKS V KKP YFPMPWEGFR 1020	
DELLSHLASQ PIAKKISEEL KAGYQSLDYI FVS RMPKRSI TGA AHQQTIM RKGGIDKKGK 1080	
TIIIERHLK DIKF DENGDF KMVGK E QDMA TYEA IKQ RYL EH GKN SKAF ETPL YKPSKK 1140	
GTGNLI KRVK VEGQAKSF VR EVNGGVAQNG DLVRV DLF EK D DKY YMV PIY VPDT VCS ELP 1200	
KV VASSK GY EQWLTLDNSF TFKFSLYPYD LV RLVKGD E RFLYFGTL D I DSDR LNF KDV 1260	
NKPSKKNEYR YSLK TIEDLE KYEVGV LGD L R LVRKETRRN FHSGGS KRPA ATK KAGQAKK 1320	
KK 1322	
SEQ ID NO: 507	moltype = AA length = 1322
FEATURE	Location/Qualifiers
REGION	1..1322
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1322
	note = source = /note="LPG50159-nAPG07433.1protein sequence"
source	1..1322
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 507	
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSDLELNHE YWMRHALSLA KRARDEGEVP 60	
VGAVLVLNQQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLLDTT LYVTFEPVCV 120	
CSGAMVHSRI CTLVYGV RNE KRGAAAGSLMN VLGYPGMNNHQ VOIIGGV LAP DCSGLLCDFY 180	
RMPRQKNNQQ ETPGTSESAT PESSGGSSGG SMRELDYRIG 240	
LAIGTNSIGW CVIELSWNKD RERYEKVIRV DQGVRMFDR A EMPKTGASLA EPRIARISSR 300	
RRLNRKSQRK KNIRNLVQH GVTQEELDS LYPLSKKSM D IWGIRLDGLD RLLNHFEWAR 360	
LILHLAQQRG FKSNRKSELE DTETGKVLS IQLNEKRLSL YRTVGEWMK DPDFS KYDRK 420	
RNSPNEYVFS VSRAELEKEI VTLFAAQRRF QSPYASKDLQ ETYLQIWT HQ LPFASGNAIL 480	
NKVGYCSLLK GKERRIPKAT YTQYQFSALD QVNRTTRLGP D QOPFTKEQRE II LNMMFQRT 540	
DYYKKKTIPE VTYYDIRKWL ELD E TIQFKG L NYD PNEELK KIEKKPFI NL KAFYEINKVV 600	
ANYSERTNET FSTLBYDGIG Y ALTVYKTDK DIRSYLKSH NLPKRCYDDQ LIEELL SLSY 660	
TKFGHLSLRK INHVLSIMQK GNTYKEAVDQ LGYDTSGLKK EKR SKFLPPI SDEITNP IVK 720	
RALTQARKVV NAIIRRHGSP HSVHIELARE LSKNHDERTK IVSAQDENYK KNKAISILS 780	
EHGILNPTGY DIVRYKLWKE QGERCAYSLK EIPADTF FNE LKKER NGAPI LEVDHILPYS 840	
QSFIDSYH NK VL VYSDENRK KG NRIPYTYF LETN KDW EAF ERYVRSN KFF SKKKREYLLK 900	
RAYL PRESEL I KERH LNDTR YASTFLKNFI EQNLQFKEAE DNPRK RRVQ VNGVITA HFR 960	
KR WGLEK DRQ ETYLHHAMDA IIVACTDHMM VTRVTEY YQI KESNKS V KKP YFPMPWEGFR 1020	
DELLSHLASQ PIAKKISEEL KAGYQSLDYI FVS RMPKRSI TGA AHQQTIM RKGGIDKKGK 1080	
TIIIERHLK DIKF DENGDF KMVGK E QDMA TYEA IKQ RYL EH GKN SKAF ETPL YKPSKK 1140	
GTGNLI KRVK VEGQAKSF VR EVNGGVAQNG DLVRV DLF EK D DKY YMV PIY VPDT VCS ELP 1200	
KV VASSK GY EQWLTLDNSF TFKFSLYPYD LV RLVKGD E RFLYFGTL D I DSDR LNF KDV 1260	
NKPSKKNEYR YSLK TIEDLE KYEVGV LGD L R LVRKETRRN FHSGGS KRPA ATK KAGQAKK 1320	
KK 1322	
SEQ ID NO: 508	moltype = AA length = 1320
FEATURE	Location/Qualifiers
REGION	1..1320
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1320

-continued

---

```

source          note = source = /note="LPG50160-nAPG07433.1protein sequence"
               1..1320
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 508
MAPKKKRKVD YKDHDGYKD HDIDYKDDDD KMSDHEFNGE YWMRHALTLA KRAREEGEVP 60
VGAVLVLNNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV LQNYRLIDAT LYVTFEPCVM 120
CAGAMVHSRI SRLVFGVRNS KRGAAAGSLMN VLNPYGMNHR VEITEGLILAE SCSAMLCDFY 180
RWPREVKNAL KKARQEESGG SSGGSSGSET PGTSESATPE SSGGSSGSM RELDYRIGLA 240
IGTNSIGWGV IELSWNKDR ERYKVRIVDQ GVRMFDRADM PKTGASLAEP RRIARSSRR 300
LNPKSQRKRN IRNLLVQHGV ITQEELDLSY PLSKKSMDIW GIRLGLDRL LNHFEWARLL 360
IHLAQRRGFK SNRKSELKDT ETGKVLLSIQ LNEKRLSLYR TVGEMWMKDP DFSKYDRKRN 420
SPNEYVFSVS RAELEKEIVT LFQAAQRFFQSY ASKDLQET YLQIWTQHQLP FASGNAILNK 480
VGYCSLLKGK ERRIPKATYT FQYFSALDQV NRTRLGPDPQ PFTKEQREII LNNMFPRTDY 540
YKKKTIPEVT YYDIRKWEL DETIQFKGLN YDPNEELKKI EKKPFINLKA FYEINKVAN 600
YSERTNETFS TLDYDGIGYA LTIVKTDKDI RSYLKSSHNL PKRCYDDQLI EELLSLSYTK 660
FGHLSLKAIN HVLSIMQKGN TYKEAVDQLG YDTSGLKEK RSKFLPPISD EITNPIVKRA 720
LTQARKVVNA IIRRHGSPHS VHIELARELS KNHDERTKIV SAQDENYKKN KGAISILSEH 780
GILNPTGYDI VRYKLWKEQG ERCAYSLKEI PADTFFNEKL KERNGAPILE VDHILPYSQS 840
FIDSYHNKVL VYSDENRKKG NRIPYTYFLE TNKDWEAER YVRSNKFSSK KKREYLLKRA 900
YLPRESELIK ERHLDNDTRYA STFLKNFIEQ NLQFKEAEDN PRKRRVQTVN GVITAHFRKR 960
WGLEKDRQET YLHHAMDAII VACTDHHMVT RVTEYYQIKE SNKSVKKPYF PMPWEGFRDE 1020
LLSHLASQPI AKKISEELKA GYQSLDYIFV SRMPKRSITG AAHKQTIMRK GGDKKGKTI 1080
IIERLHLKDI KFDENGDFKM VGKEQDMATY EAIKQRYLEH GKNSSKAFET PLYKPSKKG 1140
GNLIKRVKVE QOAKSFVREV NGGAQNGDL VRVDSLFEKDD KYYMVPYIYV DTVCELPKK 1200
VVASSKGYEQ WLTLDNSFTF KFSLYPYDLV RLVKGDEDRF LYFGTLDIDS DRLNFKDVNK 1260
PSKKNEYRYS LKTIEDLEKY EVGVLGDLRL VRKETRRRNHF SGGSKRPAAT KKAGQAKKK 1320

SEQ ID NO: 509      moltype = AA length = 1320
FEATURE          Location/Qualifiers
REGION           1..1320
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION           1..1320
note = source = /note="LPG50161-nAPG07433.1protein sequence"
1..1320
mol_type = protein
organism = synthetic construct

SEQUENCE: 509
MAPKKKRKVD YKDHDGYKD HDIDYKDDDD KMSQTELTHE YWMRHALTLA QRARDEGEVP 60
VGAVLVLNNQ VIGEGWNRAI GLHDPTAHAE IMALRQGGLV LQNYRLDDTT LYVTFEPCVM 120
CAGAMVHSRI GTLVFGVRNS KRGAVGSLMN ITGYPGMNHR VQVIEGLILAE ECSAMLCAFY 180
RQPRLVKNAL KEAAKTASGG SSGGSSGSET PGTSESATPE SSGGSSGSM RELDYRIGLA 240
IGTNSIGWGV IELSWNKDR ERYKVRIVDQ GVRMFDRADM PKTGASLAEP RRIARSSRR 300
LNPKSQRKRN IRNLLVQHGV ITQEELDLSY PLSKKSMDIW GIRLGLDRL LNHFEWARLL 360
IHLAQRRGFK SNRKSELKDT ETGKVLLSIQ LNEKRLSLYR TVGEMWMKDP DFSKYDRKRN 420
SPNEYVFSVS RAELEKEIVT LFQAAQRFFQSY ASKDLQET YLQIWTQHQLP FASGNAILNK 480
VGYCSLLKGK ERRIPKATYT FQYFSALDQV NRTRLGPDPQ PFTKEQREII LNNMFPRTDY 540
YKKKTIPEVT YYDIRKWEL DETIQFKGLN YDPNEELKKI EKKPFINLKA FYEINKVAN 600
YSERTNETFS TLDYDGIGYA LTIVKTDKDI RSYLKSSHNL PKRCYDDQLI EELLSLSYTK 660
FGHLSLKAIN HVLSIMQKGN TYKEAVDQLG YDTSGLKEK RSKFLPPISD EITNPIVKRA 720
LTQARKVVNA IIRRHGSPHS VHIELARELS KNHDERTKIV SAQDENYKKN KGAISILSEH 780
GILNPTGYDI VRYKLWKEQG ERCAYSLKEI PADTFFNEKL KERNGAPILE VDHILPYSQS 840
FIDSYHNKVL VYSDENRKKG NRIPYTYFLE TNKDWEAER YVRSNKFSSK KKREYLLKRA 900
YLPRESELIK ERHLDNDTRYA STFLKNFIEQ NLQFKEAEDN PRKRRVQTVN GVITAHFRKR 960
WGLEKDRQET YLHHAMDAII VACTDHHMVT RVTEYYQIKE SNKSVKKPYF PMPWEGFRDE 1020
LLSHLASQPI AKKISEELKA GYQSLDYIFV SRMPKRSITG AAHKQTIMRK GGDKKGKTI 1080
IIERLHLKDI KFDENGDFKM VGKEQDMATY EAIKQRYLEH GKNSSKAFET PLYKPSKKG 1140
GNLIKRVKVE QOAKSFVREV NGGAQNGDL VRVDSLFEKDD KYYMVPYIYV DTVCELPKK 1200
VVASSKGYEQ WLTLDNSFTF KFSLYPYDLV RLVKGDEDRF LYFGTLDIDS DRLNFKDVNK 1260
PSKKNEYRYS LKTIEDLEKY EVGVLGDLRL VRKETRRRNHF SGGSKRPAAT KKAGQAKKK 1320

SEQ ID NO: 510      moltype = AA length = 1321
FEATURE          Location/Qualifiers
REGION           1..1321
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION           1..1321
note = source = /note="LPG50162-nAPG07433.1protein sequence"
1..1321
mol_type = protein
organism = synthetic construct

SEQUENCE: 510
MAPKKKRKVD YKDHDGYKD HDIDYKDDDD KMSNPELNHD YWMRHALSLA KRAREEGEVP 60
VGAVLVRNNE VIGEGWNRAI GLHDPTAHAE IMALRQGGMV LQNYRLIDTT LYVTFEPCVM 120
CAGAMVHSRI GQLVFGVRNS KRGAAAGSLMN VLNPYGMNHR VEIVEGVLRD ECAGMLCDFY 180

```

---

-continued

---

RQPLRVKNAQ	KKGAEPLISG	GSSGGSSGSE	TPGTSESATP	ESSGGSSGGS	MRELDYRIGL	240
AIGTNSIGW	VIELSWNKDH	ERYEIKVIRD	QGVRMFDRAE	MPKTGASLAE	PRRIARSSRR	300
RLNRKSQRKK	NIRNLVQHQ	VITQEELDSL	YPLSKKSMDI	WGIRLDGLDR	LLNHFEWARL	360
LIIHLAQRRGF	KSNRKSELKD	TETGKVLSI	QLNEKRLSLY	RTVGEMWMKD	PDFSKYDRKR	420
NSPNEYVFSV	SRAELEKEIV	TLFAAQRRFQ	SPYASKDLQE	TYLQIWTQHQL	PFASGNAILN	480
KVGYCSLLKG	KERRIPKATY	TFQYFSALDQ	VNRTRLGPDF	QPFTKBEORET	IILNNMFQRTD	540
YYKKKTIPEV	YYDIRKWL	LDETIQFKGL	NYDPNEELKK	IEKKPFINLK	AFYEINKVVA	600
NYSSERTNETF	STLDYDGIG	ALTVYKTDKD	IRSYLKSSH	LPKRCYDDQL	IEELLLSLSYT	660
KFGHLSLKAI	NHVLSIMQKG	NTYKEAVDQL	GYDTSGLKE	KRSKFLPPIS	DEITNPIVKR	720
ALTQARKVNV	AIIRRHHGSPH	SVHIELAREL	SKNHDERTKI	VSAQDENYKK	NKGAI SILSE	780
HGILNPTGYD	IVRYKLWKEQ	GERCAYSLKE	IPADTFFNELL	KKERNGAPIL	EVDHILPYSQ	840
SFIDSYHNKV	LVYSDENRK	GNRIPYTYFL	ETNPKDWEAFT	RYVRNSNKFFS	KKKREYLILK	900
AYLPRESELI	KERHLNDTRY	ASTFLKNFIE	QNLQFKEAED	NPRKRRVQTV	NGVITAHFRK	960
RWGLEKDRQE	TYLHHAMDAI	IVACTDHHMV	TRVTEYYQIK	ESNKSVKKPY	FPMPWEGFRD	1020
ELLSHLASQP	IAKCISEELK	AGYQSLDYFV	VSRMPKRSIT	GAHKQTIMR	KGGIDKKGKT	1080
IIIERLHLKD	IKFDENGDFK	MVGKEQDMAT	WNGKNSKKAFT	TPLYKPSKKG	1140	
TGNLIKRVKV	EGQAKSFVRE	VNGGVAQNGD	LVRVLDFEKL	DKYYMVIYV	PDTVCSELPK	1200
KVASSKGYE	QWLTLDNSFT	FKFSLYPYDL	VRLVKGDEDR	FLYFGTLDID	SDRLNFKDVN	1260
KPSKKNEYRY	SLKTIEDLEK	YEVGVLGDLR	LVRKETRRNF	HSGGSKRPA	TKKAGQAKKK	1320
K						1321

```

SEQ ID NO: 511      moltype = AA  length = 1326
FEATURE          Location/Qualifiers
REGION           1..1326
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
REGION           1..1326
                  note = source = /note="LPG50163-nAPG07433.1protein sequence"
source            1..1326
                  mol_type = protein
                  organism = synthetic construct

```

```

SEQUENCE: 511
MAPKKKRKV D YKDHDG DYKD HDIDYK DDDD KMSNPE LHNE YWMRYALT LA KRAR DEGE VP 60
VGAVLVN DQ VIGEGWN RAI GLHDPTA HAE IMA LRQG GLV LQNY RL DTT LYVT FEP CVM 120
CAGAMVHSR I GRLV PGV RNS KRG AAG SLLN VL NPG GMN HH I EME EG VL RD ECA AML CDF Y 180
RQPRMVKN AL KKS PPD SPNL QAR SGSS GSSG SET PG TS ES AT PESS GG SSGG SM REL D 240
YRIGLAIG T N SIG WGV IELS WNK DRERY EK VRIV DQ GVR M FD RAEM PK TG ASLA EP R P RIA 300
RSSR RRL NRK SQR KK N IRNL LVQ HG VIT QE EL DS LY PLS K KS MDI NG I RL D GL DR LL NH F 360
EWAR L L NRK QR GQ FKS NR S SEL K D T ET GK VL SS SI QL NEK RLS LY RT VGE MM WKD P DF SK 420
YDRKR N S PNE YF VS V S R A E L E K I V T F Q Y FV SAL D QV N R T R LG PD F Q P FT K EQ RE I L N NM 480
NAI LN K V G Y C SLL K G K E R R I P K A T Y T F Q Y FV SAL D QV N R T R LG PD F Q P FT K EQ RE I L N NM 540
F Q R T D Y Y K K T I P E V T Y Y D I R K W L E L D E T I Q F K G L N Y D P N E E L K K I E K K P F I N L K A F Y E I 600
N K V V A N Y S E R T N E T F S T L D Y D G I G Y A L T V Y K T D K D I R S Y L K S S H N L P K R C Y D D Q L I E E L L 660
SLS Y T K F G H L S L K A I N H V L S I M Q K G N T Y K E A V D Q L G Y D T S G L K N E K R S K F L P P I S D E I T N 720
P IV K R A L T Q A R K V V N A I I R H H G S P H V S H I E L A R E L S K N H D E R T K I V S A Q D E N Y K K N G A I 780
S I L S E H G I L N P T G Y D I V R Y K L W K E Q G E R C A Y S L K E I P A D T F F N E L K K E R N G A P I L E V D H I 840
L P Y S Q F S I D Y H N K V L V Y S E N R K K G N K R I P T Y T F L E T M K D W E A F E R Y V R S N K F F S K K K R E 900
Y L L K R A Y L P R E S E L I K R H L D T R Y A S T L K N F I E Q N L Q P K A E D N P R K R R V Q T V N G V I T 960
A H F R K R W G L E K D R Q E T Y L H H A M D A I I V A C T D H H M V T R V T E Y Q O I K E S N K S V K K P Y F P M P W 1020
E G F R D E L L S H L A S Q P I A K K I S E E L K A G Y Q S L D Y I F V S R M P K R S I T G A A H K Q T I M R K G G I D 1080
K K G K T I I I E R L H L K D I K F D E N G D F K M V G K E Q D M A T Y E A I K Q R Y L E H G K N S K K A F E T P L Y K 1140
P S K K G T G N L I P R K V K V E G Q A K S F V R E V N G G V A Q N G D L V R V D L F K E D R D K Y Y M V P I Y P D T V C 1200
S E L P K V V A S K G Y E Q W L T L D N S F T F K F S L Y P D L V R L V K G D E D R F L Y F G T L D I D S R L N 1260
F K D V N K P S K K N E Y R S L K T I E D L E K Y E V G V L G D L R L V R K E T R R N F H S G G S K R P A A T K K A G 1320
Q A K K K K 1326

```

```

SEQ ID NO: 512      moltype = AA  length = 1322
FEATURE          Location/Qualifiers
REGION           1..1322
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
REGION           1..1322
                  note = source = /note="LPG50164-nAPG07433.1protein sequence"
source            1..1322
                  mol_type = protein
                  organism = synthetic construct

```

```

SEQUENCE: 512
MAPKKKRKV D YKDHDG DYKD HDIDYK DDDD KMSNPE FTHE YWMRH AL TL A RRAR DEGE VP 60
VGAVLVN N Q VIGEGWN RAI GLHDPTA HAE IMA LRQG GLV LQNY RL DTT LYVT FEP CVM 120
CSGAMVHSR I GTLV PGV RNE KRG AAG SLMN VL GYP GMN HQ VKT IGG VL AP ECG SLL CDF Y 180
RMP RQQ KN Q Q K A E L K S S G D S G G S S E T P G T S E S A T P E S S G G S S G S M R E L D Y R I G 240
LAIGTNSIG W G V I E L S W N K D R E Y E K V R I V D Q G V R M F D R A E M P K T G A S L A E P R R I A R S S R 300
RRLNRKSQRK KNIRNLVQH G V I T Q E E L D S L Y P L S K K S M D I I W G I R L D G L D R L L N H F E W A R 360
LIIHLAQRRG F K S N R K S E L K D T E T G K V L S I Q L N E K R L S L Y R T V G E M W M K D P D F S K Y D R K 420
R N S P N E Y V F S V S R A E L E K I V T L F A A Q R R F Q S P Y A S K D L Q E T Y L Q I W T Q H L P F A S G N A I L 480
N K V G Y C S L L K G K E R R I P K A T Y T F Q Y F S A L D Q V N R T R L G P D F Q P F T K B Q R E I I L N N M F Q R T 540

```

---

-continued

---

DYYKKKTIPE	VTYYDIRKWL	ELDETIQFKG	LNYDPNEELK	KIEKKPFINL	KAFYEINKVV	600
ANSERTNET	FSTLDYDGIG	YALTIVYKTDK	DIRSYLKSSH	NLPKRCYDDQ	LIEELLSLSY	660
TKFGHSLKA	INHVLSIMQK	GNTYKEAVDQ	LGYDTSGLKK	EKRSKFLPPI	SDEITNPIVK	720
RALTQARKVV	NAIIRRHHGP	HSHVIELARE	LSKNHDERTK	IVSAQDENYK	KNKGAIISLS	780
EHGILNPNTGY	DIVRYKWLKE	QGERCAYSLK	EIPADTFFNE	LKKERNGAPI	LEVDHILPYS	840
QSIFIDSYHK	VLVYSDENRK	KGNRIPYTF	LETNPKDWEAF	ERYVRSNKFPF	SKKKREYLLK	900
RAYLPRESEK	IKERHLNDTR	YASTFLKNFI	EQNLOQFKEAE	DNPRTKRRVQT	VNGVITAHFR	960
KRWGLEKDRQ	ETYLHHAMDA	IIVACTDHMH	VTRVTEYYQOI	KESNKSVKKP	YFPMPMPEGFR	1020
DELLSHLASQ	PIAKKISEEL	KAGYQSLDYI	FVSRMPKRSI	TGAAHQQTIM	RKGGIDKKGK	1080
TIIIERHLHK	DIKFDENGDF	KMVKGQDMA	TYBAIKQRYL	EHGKNSKKAF	ETPLYKPSKK	1140
GTGNLIKRVK	VEGQAKSFVR	EVNGVVAQNG	DIVRVDLFKE	DDKYYMVPYI	VPDTVCSELP	1200
KKVVAASSKGY	EQWLTLDSNF	TFKFSLYPYD	LVRLVKGDED	RFLYFTGLDI	DSDRLNFKDV	1260
NKPSKKNEYR	YSLKTIEDLE	KYEVGVLGDL	RLVRKETRRN	FHSGGSKRPA	ATKKAGQAKK	1320
KK						1322

SEQ ID NO: 513                  moltype = AA    length = 1319  
 FEATURE                          Location/Qualifiers  
 REGION                          1..1319  
 note = source = /note="Description of Artificial Sequence:  
                                   Synthetic polypeptide"  
 REGION                          1..1319  
 note = source = /note="LPG50165-nAPG07433.1protein sequence"  
 source                          1..1319  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 513  
 MAPKKKRKVD YKDHDGDKD HDIDYKDDDD KMSDNEPNHE YWMRHALTLA QRARDEGEVP 60  
 VGAVLVLDNQ VIGEGWNRAI GLHDPATAHE IMALRQGGMV LQNYRLINAT LYVTPEPCVM 120  
 CAGAMVHSRI GHVVPGVRNS KRGAAAGSLMN VLNPYGMNHR VEVTEGVLR ECAAGMLCDFY 180  
 REPREQFNAL RAKAQKASGGS SGGSSGSETP GTSESATPES SGGSSGSSMR ELDYRIGLAI 240  
 GTNSIGWGV1 ELSWNKDRER YEKVRIVDQG VRMFDRDRAEMP KTGASLAEPR RIARSSRRRL 300  
 NRKSQRKKNI RNLLVQHGV1 TQEELDSLVP LSKKSMDIWG IRLDGLDRLL NHFEWARLLI 360  
 HLAQRRGFKS NRKSELKDTE TGKVLSIQL NEKRLSLYRT VGEMWMKDPD FSKYDRKRNS 420  
 PNEYVFSVSR AELEKEIVTL FAAQRQFQSP YASKDLQETY LQIWTHQLPF ASGNAILNKV 480  
 GYCSLLKGKE RRIPKATYTF QYFSALDQVN RTRLGPDFQF FTKEQREIIL NNMFQRTDYY 540  
 KKKTIPEVTY YDIRKWLLED ETIQFKGLNY DPNEELKKIE KKPFINLKAF YEINKVWANY 600  
 SERTNETFST LDYDGIGYAL TVYKTDKDIR SYLKSSHNLN KRCYDDQDIE ELLSSLYSTKF 660  
 GHLSLKAINT VLSIMQKGNT YKEAVDQLGV DTSGLKEKRR SKFLPPISDE ITNPIVKRAL 720  
 TQARKVNVNAI IRRHGSQPHSV HIELARELSK NHDERTKIVS AQDENYKKNN GAISILSEHG 780  
 ILNPTGYDIV RYKLWKEQGE RCAYKLSEK ADTFFNELLK ERNGAPILEV DHILPYSQSF 840  
 IDSYHNKVLV YSDENRKKGKRN RIPYTYFLET NKDWEAFERY VRSNKPFSSK KREYLLKRAY 900  
 LPRESELIKE RHLNDTRYAS TFLKNFIEQN LOFKEAEDNP RKRRVQTENG VITAHFRKRW 960  
 GLEKDRQETY LHHAMDAIV ACTDHDMVTR VTEYYYQIKES NKSVKKPYFP MPWEGFRDEL 1020  
 LSHLASLKEKAG YQSLDYIFVS RMPKRSITGA AHKQTIMRKG GIDKKKGKTI 1080  
 IERLHLKDIK PDENGDFKMY GKEQDMATYE AIKQRYLEHG KNSKKAFTP LYKPKKGKGTG 1140  
 NLIKRKVKEG QAKSFVREVN GGVAQNGDLV RVDLFEKDDK YMVPIYVVD TVCSELPKKV 1200  
 VASSKGYEQW LTLDNSFTF FSLYPLYDVLV LVKGDEDRF1 YFGTLIDSD RLNFKDVNKP 1260  
 SKKNEYRYSL KTIEDLEKYE VGVLGDLRLV RKETRRNPHS GGSKRPAAATK KAGQAKKK 1319

SEQ ID NO: 514                  moltype = AA    length = 1324  
 FEATURE                          Location/Qualifiers  
 REGION                          1..1324  
 note = source = /note="Description of Artificial Sequence:  
                                   Synthetic polypeptide"  
 REGION                          1..1324  
 note = source = /note="LPG50166-nAPG07433.1protein sequence"  
 source                          1..1324  
 mol\_type = protein  
 organism = synthetic construct

SEQUENCE: 514  
 MAPKKKRKVD YKDHDGDKD HDIDYKDDDD KMSDNELNHE YWMRHALTLA QRARDEGEVP 60  
 VGAVLVLNQQ VIGEGWNRAI GLHDPATAHE IMALRQGGMV LQNYRLIDAT LYVTPEPCIM 120  
 CAGAMVHSRI GQVVFGRVNS KRGAAAGSLIN ILNPYGMNHR DVTEGVLSE RCANMLCDFY 180  
 REPRLOPNAQ RAKAEKAGNA ASGGSSGGSS GSETPGTSES ATPESSGGSS GGSMRELDR 240  
 IGLAIGTN1 GWGVLSWNR KDRERYEKV IVDQGVRMFD RAEMPKTGAS LAEPRRIARS 300  
 SRRRLRNRSQ RKKNIRNLVQ QHGVITQEEEL DSYLPLSKS MDIWGIRLDG LDRLLNHFEW 360  
 ARLLIHLAQR RGFKNSRKSE LKDDETGKL SSIQLNEKRL SLYRTVGEW MKDPDFSKYD 420  
 RKRNSPNEYV FSVSRAELEK EIVTLFAAQR RFQSPYASKD LQETYQLQIWT HQLPFASGNA 480  
 ILNKVGYCL LKGKERRIPK ATYTFQYFSA LDQVNRTRLG PDFQPFKEQ REIILNNMFQ 540  
 RTDYYKKKTI PEVTTYYDIRK WLELDETIQF KGLNYDPNEE LKKIEKKPFI NLKAFYEINK 600  
 VVANYSERTN ETFSTLDYDG IGYALTIVYKT DKDIRSYLKS SHNLPKRCYD DQLIBELLSL 660  
 STYKFGHSL KAINHVLSIM QKGNTYKEAV DQLGYDTSGL KKEKRKFLP PISDEITNPI 720  
 VVKALTQARK VVNAAIRRHH SPHSVHIELA RELSKNHDERTK IVSAQDEN YKKNKGAISI 780  
 LSEHGINP1 GYDIVRYKWL KEQGERCAYS LKEIPADTFF NELKKERNGA PILEVDHILP 840  
 YSQSFIDSYH NKVLVYSDEN RKKGNRIPYT YFLETNKDWAE AFERYVRSNK FFSKKREYI 900  
 LIKRAYLPRES ELIKERHLND TRYASTFLKN FIEQNLQPKE AEDNPRKRRV QTNGVITAH 960

---

-continued

---

FRKRWGLEKD RQE <sup>T</sup> YLHHAM DAIIVACTDH HMVTRVTEYY QIKESNKSVK KPYFPMPWEG	1020
FRDELLSHLA S <sup>Q</sup> PIAKISE ELKAGYQS <sup>L</sup> D YIFVSRMPKR SITGAHKQT IMRKGGIDKK	1080
GK <sup>T</sup> IIERLH LDKIKFDENG DF <sup>K</sup> MVGKEQD MATYEAIKQR YLEHGKNSKK AFETPLYKPS	1140
KKGTGNL <sup>I</sup> IKR VKVEQOAKSF VREVN <sup>G</sup> GAQ <sup>N</sup> NGDLVRV <sup>D</sup> LF EKDDKYVMVP IYV <sup>P</sup> DTVCSE	1200
LPKKVVASSK GYEQWLTL <sup>D</sup> N SFTFKFSLYP YDLVRLVKG <sup>D</sup> ED <sup>R</sup> FLYFGTL DIDSDRLNF <sup>K</sup>	1260
DVN <sup>K</sup> PSKKN <sup>E</sup> YRYSLKTIED LEKYEVGV <sup>L</sup> G DLRLVRKETRR RN <sup>F</sup> HSGGSKR PAATKKAGQA	1320
KKKK	1324

SEQ ID NO: 515	moltype = AA length = 1323
FEATURE	Location/Qualifiers
REGION	1..1323
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1323
	note = source = /note="LPG50167-nAPG07433.1protein sequence"
source	1..1323
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 515	
MAPKKKRKV <sup>D</sup> YKDHDG <sup>D</sup> YKD HDIDYK <sup>DDDD</sup> KMSNPELTH HWMRH <sup>A</sup> LT <sup>A</sup> QRARNEGEVP	60
VGAVLV <sup>L</sup> LNQ <sup>Q</sup> VIGEGWNRAI GLHDPTA <sup>H</sup> AE IMALRQ <sup>G</sup> GLV LQN <sup>Y</sup> RL <sup>I</sup> DTV LYVT <sup>F</sup> EP <sup>C</sup> VM	120
CAGAMVHSRI GQLVFGVRNS K <sup>R</sup> GAAGS <sup>L</sup> IN VLN <sup>Y</sup> PGM <sup>N</sup> HR VEII <sup>E</sup> GV <sup>L</sup> R <sup>D</sup> ECAAMLCDFY	180
RHPRLVKN <sup>A</sup> L KKNAGT <sup>S</sup> T <sup>P</sup> TQ SGG <sup>N</sup> SSG <sup>S</sup> SETPGTSESA PTESSGGSS <sup>S</sup> GSMRELDYRI	240
GLAIGT <sup>N</sup> SIG WGVI <sup>E</sup> LSWN <sup>K</sup> DRERYEKV <sup>R</sup> I VDQGVRMPDR <sup>R</sup> AEMP <sup>K</sup> TGASL AEP <sup>R</sup> R <sup>I</sup> ARSS	300
RRRLNRKSQR KKNIRNLLVQ HG <sup>V</sup> ITQEE <sup>E</sup> LD SLYPLSKKS <sup>M</sup> DIWGIRLDGL DR <sup>L</sup> LNHF <sup>E</sup> WA	360
RLLIHLAQ <sup>R</sup> R GFKSNRKS <sup>E</sup> LD KDTETGKV <sup>L</sup> S SIQ <sup>L</sup> NEKR <sup>L</sup> LYRTVGEMWW KDPDPSKYDR	420
KRNPSPNEYVF SVSRAELEKE I <sup>V</sup> TVLFAAQR <sup>R</sup> FQSPYASK <sup>D</sup> L Q <sup>E</sup> TYLQIW <sup>H</sup> TH QLPFASGN <sup>A</sup> I	480
LNKVG <sup>C</sup> YCS <sup>L</sup> L KGKERRIPKA TYTFQYFS <sup>A</sup> L DQVN <sup>R</sup> TRLGP DFQPFTKEQR EII <sup>L</sup> NNMFQR	540
TDYYKK <sup>T</sup> TIP EVTYYDIRK <sup>W</sup> LELDETIQFK <sup>R</sup> GLNYDPNEEL KKIEKKPF <sup>I</sup> N LKAFYEINKV	600
VANYSERTNE TFST <sup>L</sup> DYDG <sup>I</sup> I GYALT <sup>V</sup> YKTD KDIRSYLK <sup>S</sup> HNL <sup>P</sup> KRCYDD QLIEELL <sup>L</sup> SLS	660
YT <sup>K</sup> FGHLSLK AINHVL <sup>S</sup> IM QGNTYKEA <sup>D</sup> QL G <sup>Y</sup> QDTS <sup>G</sup> KL KEKRSKF <sup>L</sup> PP ISDEITNPIV	720
KRALTQARKV VN <sup>A</sup> II <sup>I</sup> R <sup>R</sup> HGS PHSVHIELAR ELSKNHD <sup>E</sup> RT KIVSAQDENY KKN <sup>K</sup> GAISIL	780
SEHGLNPTG YDIV <sup>R</sup> YKLW <sup>K</sup> EQGERCAYSL KEIPAD <sup>T</sup> FFN ELKKERN <sup>G</sup> AP I <sup>L</sup> EV <sup>D</sup> HILPY	840
SQSFIDSYHN KVLVYSDEN <sup>R</sup> KKG <sup>N</sup> RIPY <sup>T</sup> FLETNKDWEA <sup>R</sup> FERYVRSNKF <sup>R</sup> FSKKKREYLL	900
KRAYL <sup>P</sup> RESE LIKERHLNDT RYASTLV <sup>K</sup> YKTD KDIRSYLK <sup>S</sup> HNL <sup>P</sup> KRCYDD QLIEELL <sup>L</sup> SLS	960
RKRWGLEKDR QBTY <sup>L</sup> HAMD AII <sup>V</sup> ACTDH <sup>M</sup> MVTRVTEYYQ <sup>I</sup> IKESNKS <sup>V</sup> VKK PYFPMPWEGF	1020
RDELLSHLAS QPIAKISE <sup>R</sup> D LKAGYQS <sup>L</sup> DY IFVSRMPK <sup>R</sup> S ITGA <sup>A</sup> HKQT <sup>I</sup> MRKGGIDKK <sup>G</sup>	1080
KTII <sup>I</sup> ERL <sup>H</sup> L KDIKFDENG <sup>D</sup> FKMVGKEQDM <sup>R</sup> ATYEAIKQR <sup>I</sup> LEHGKNS <sup>K</sup> KA FETPLYKPS <sup>K</sup>	1140
KGTGNL <sup>I</sup> KR <sup>R</sup> VKEGQAKSF <sup>R</sup> REVNGGVAQ <sup>N</sup> GDLVRLV <sup>D</sup> LF EKDDKYVMVP <sup>I</sup> YVPDTVCSEL	1200
PKKV <sup>V</sup> ASSKG YEQWLTL <sup>D</sup> N SFTFKFSLYP <sup>R</sup> DLVRLVKG <sup>D</sup> ED <sup>R</sup> FLYFGTL <sup>D</sup> ISDRLNF <sup>K</sup> D	1260
VN <sup>K</sup> PSKKN <sup>E</sup> YRYSLKTIED <sup>R</sup> LEKYEVGV <sup>L</sup> G <sup>R</sup> LRLVRKETRR <sup>R</sup> NFHSGGSKR <sup>R</sup> AATKKAGQAK	1320
KKK	1323

SEQ ID NO: 516	moltype = AA length = 1320
FEATURE	Location/Qualifiers
REGION	1..1320
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1320
	note = source = /note="LPG50168-nAPG07433.1protein sequence"
source	1..1320
	mol_type = protein
	organism = synthetic construct

SEQUENCE: 516	
MAPKKKRKV <sup>D</sup> YKDHDG <sup>D</sup> YKD HDIDYK <sup>DDDD</sup> KMSDTEL <sup>N</sup> H <sup>E</sup> YWMRHALMLA K <sup>R</sup> ARDEGEVP	60
VGAVLV <sup>L</sup> LNQ <sup>Q</sup> VIGEGWNRAI GLHDPTA <sup>H</sup> AE IMALRQ <sup>G</sup> GLV LQN <sup>Y</sup> RL <sup>I</sup> DTV LYVT <sup>F</sup> EP <sup>C</sup> VM	120
CAGAMVHSRI GNLVFGVRNS K <sup>R</sup> GAAGS <sup>L</sup> IN VLN <sup>Y</sup> PGM <sup>N</sup> HR VEIAEGVLAD ECSAMLCDFY	180
RHPRQQQN <sup>A</sup> L KQA <sup>A</sup> KHDSGG SSGGSSGSET PGTSESAT <sup>P</sup> SSGGSSGSM RELDYRIGLA	240
IGTNSIGWGV IELSWNKD <sup>R</sup> RYE <sup>K</sup> V <sup>R</sup> IDVQ <sup>R</sup> GVRMF <sup>D</sup> RAEM <sup>R</sup> PKTGASLAEP <sup>R</sup> R <sup>R</sup> IARSSRR	300
LNKRSQR <sup>I</sup> KRN <sup>I</sup> 1RN <sup>L</sup> VQHGV <sup>I</sup> ITQEE <sup>L</sup> DL <sup>S</sup> Y PLSKKMS <sup>D</sup> ML GIRL <sup>G</sup> DL <sup>R</sup> LNHFEWAR <sup>L</sup>	360
IHLAQR <sup>R</sup> GF <sup>K</sup> SNRSKELKD <sup>R</sup> ETGKV <sup>L</sup> SS <sup>I</sup> Q LNEKRLS <sup>L</sup> Y <sup>R</sup> TVGEMMMKD <sup>R</sup> DFSKYDRK <sup>R</sup>	420
SPNEYVFSVS RAELEKE <sup>I</sup> VT LF <sup>A</sup> Q <sup>R</sup> RFQ <sup>S</sup> PYASK <sup>D</sup> LQET <sup>R</sup> YLQIW <sup>H</sup> QLP <sup>R</sup> FASGNAIL <sup>N</sup> LN <sup>K</sup>	480
VGYCSL <sup>L</sup> KG <sup>K</sup> ER <sup>R</sup> IPKAT <sup>T</sup> YFQYF <sup>S</sup> ALDQ <sup>V</sup> NR <sup>R</sup> TRLGPD <sup>F</sup> Q <sup>R</sup> PFTKEQRE <sup>I</sup> LNNNM <sup>R</sup> QRT <sup>D</sup> Y	540
YKKK <sup>T</sup> PEV <sup>T</sup> YYDIRK <sup>L</sup> WE <sup>I</sup> DETI <sup>R</sup> QFK <sup>G</sup> LN <sup>R</sup> YDPNEELKK <sup>I</sup> EKKP <sup>R</sup> FINLK <sup>R</sup> FYEINKV <sup>A</sup> N	600
YSERTNETFS TLDYD <sup>G</sup> IGYA <sup>R</sup> LT <sup>V</sup> Y <sup>K</sup> D <sup>T</sup> KD <sup>I</sup> RS <sup>L</sup> Y <sup>K</sup> Q <sup>S</sup> SHNL <sup>R</sup> PRKCYDDQ <sup>L</sup> I E <sup>L</sup> LLS <sup>L</sup> SY <sup>T</sup> K	660
F <sup>G</sup> HLSL <sup>I</sup> KAIN <sup>R</sup> HVLSIMQ <sup>K</sup> G <sup>N</sup> TYKEAVDQ <sup>L</sup> Q <sup>R</sup> YD <sup>T</sup> S <sup>G</sup> GL <sup>K</sup> KE <sup>R</sup> RSKFL <sup>P</sup> PI <sup>S</sup> D EITNP <sup>I</sup> IV <sup>K</sup> R <sup>A</sup>	720
LTQARKV <sup>V</sup> N <sup>A</sup> IIRR <sup>H</sup> GSPHS <sup>R</sup> VHIELARELS <sup>R</sup> KNHD <sup>E</sup> RT <sup>I</sup> KV <sup>R</sup> SAQDENY <sup>K</sup> N <sup>R</sup> KGAISIL <sup>S</sup> HE	780
GILN <sup>P</sup> TYGYDI <sup>R</sup> VRYKLW <sup>K</sup> E <sup>Q</sup> ERC <sup>A</sup> YSL <sup>K</sup> E <sup>I</sup> PAD <sup>T</sup> FFNE <sup>L</sup> K <sup>R</sup> KERN <sup>G</sup> AP <sup>I</sup> E <sup>R</sup> VDHIL <sup>P</sup> SQS <sup>R</sup>	840
FIDSYHN <sup>K</sup> V <sup>L</sup> VYSDE <sup>N</sup> R <sup>K</sup> K <sup>G</sup> NR <sup>I</sup> PY <sup>T</sup> YF <sup>L</sup> E TNKDWEA <sup>F</sup> ER YVRSN <sup>K</sup> FFSK <sup>R</sup> KKREYLL <sup>K</sup> R <sup>A</sup>	900
YLPRESEL <sup>I</sup> ER <sup>H</sup> LN <sup>D</sup> TRYA <sup>R</sup> STFLKN <sup>F</sup> IEQ <sup>R</sup> NLQFKEAEDN <sup>R</sup> PRKRRVQ <sup>T</sup> TVN <sup>R</sup> GVITA <sup>H</sup> FR <sup>K</sup> R	960
WGLEKDRQ <sup>E</sup> T YLHH <sup>I</sup> MDAII <sup>R</sup> VACTDH <sup>M</sup> MT <sup>R</sup> VTE <sup>Y</sup> YQ <sup>I</sup> KE <sup>R</sup> SN <sup>K</sup> SV <sup>K</sup> K <sup>P</sup> Y <sup>R</sup> PMP <sup>W</sup> GR <sup>F</sup> R <sup>D</sup> E	1020
LLSHLASQ <sup>P</sup> I AKKISEELKA <sup>R</sup> GYQSL <sup>D</sup> Y <sup>I</sup> F <sup>V</sup> SRMPKRS <sup>I</sup> IT <sup>R</sup> AAHKQT <sup>I</sup> MRK <sup>R</sup> GGIDKK <sup>G</sup> K <sup>T</sup> I	1080
I <sup>I</sup> ERL <sup>H</sup> L <sup>I</sup> KD <sup>I</sup> KFDENGDFKM <sup>R</sup> VGK <sup>E</sup> QDM <sup>A</sup> TY <sup>R</sup> EA <sup>I</sup> KQ <sup>R</sup> Y <sup>L</sup> E <sup>R</sup> GKNSKKAFET <sup>R</sup> PLY <sup>K</sup> PSK <sup>K</sup> GT <sup>R</sup>	1140
GN <sup>L</sup> IKR <sup>V</sup> K <sup>V</sup> E GQAKSF <sup>V</sup> REV <sup>R</sup> NGGVAQNGD <sup>R</sup> VRV <sup>D</sup> LF <sup>E</sup> K <sup>D</sup> D Y <sup>K</sup> YMP <sup>V</sup> I <sup>Y</sup> P <sup>R</sup> D <sup>T</sup> V <sup>C</sup> SEL <sup>P</sup> KK <sup>R</sup>	1200
VVASSKG <sup>Y</sup> EQ <sup>R</sup> WLTL <sup>D</sup> N <sup>I</sup> S <sup>F</sup> TF <sup>R</sup> K <sup>F</sup> FSLY <sup>P</sup> YD <sup>L</sup> V R <sup>L</sup> V <sup>G</sup> DE <sup>R</sup> DR <sup>F</sup> LY <sup>G</sup> FT <sup>G</sup> LD <sup>I</sup> D <sup>S</sup> D <sup>R</sup> LN <sup>F</sup> KDV <sup>N</sup> K	1260
PSKKNEYRYS <sup>R</sup> LKTIED <sup>R</sup> LEKYEVGV <sup>L</sup> G <sup>R</sup> LRLVRKETRR <sup>R</sup> NFHSGGSKR <sup>R</sup> AATKKAGQAK <sup>R</sup>	1320

---

-continued

---

```

SEQ ID NO: 517      moltype = AA length = 1325
FEATURE          Location/Qualifiers
REGION           1..1325
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
REGION           1..1325
                  note = source = /note="LPG50169-nAPG07433.1protein sequence"
source            1..1325
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 517
MAPKKKRKV YKDHDGYKD HDIDYKDDDD KMSDIELNHE YWMRHALMLA KRAREEGEV 60
VGAVLVLNQV VIGEGWNRAI GLHDPATAHE IMALRQGLV LQNYRLIDTT LYVTPEPCVM 120
CAGAMVHSRI GQLVFGVRNS KRGAAAGSLIN VLNPYGMNHR VAITEGVLAE SCSAMLCDFY 180
RYPQQQONTL QKAAKANPPA AQSGGSSGGG SGSETPGTSE SATPESSGGG SGGSMRELDDY 240
RIGLAIGTNS IGVGVIELSW NKDRERYEV RIVDQGVRMF DRAEMPKTGA SLAEPRRIAR 300
SSRRRLNRKS QRKKNIRNLN VHGVITQEE LDSLYPLSKK SMDIWRGLD GLDRLLNHFE 360
WARLLHLAQ RRGFKSNRKS ELKDTEGTVK LSSIQLNEKRL LSILYRTVGEW WMKDPDFSKY 420
DRKRNSPNEY VEVSVRAELE KEIVTLFAQQ RRQSPYASK DLQETYLQIW THQLFASGN 480
AILNKVGYCS LLKGKERRIP KATYTFQYFS ALDQVNTRAL GDPFQPFTE QREIILNNMF 540
QRTDYKKKT IPEVITYDIR KWLELDETIQ FKGLNYPDNE ELKKIEKKPF INLKAFYEIN 600
KVANYNSERT NETFSTLDYD GIGYALTIVK TDKDIRSYLK SSHNLPKRCY DDQLLIEELLS 660
LSYTKFGHLS LKAINHVLSI MQKGNTYKEA VDQLGYDTSG KKKEKRSKFL PPISDEITNP 720
IVKRALTQAR KVVAIIRRHH GSPHSVHIEL ARELSKNHDE RTKIVSAQDE NYKKNKGAI 780
ILSEHGLNLP TGYDIVRYKL WKEQGERCAY SLKEIPADTF FNELKKERNG APILEVDHIL 840
PYQSFDISY HNKVLVYSDE NRKKGNNRIPY TYPLETNKDW EAFTERVRSN KFFSKKKREY 900
LILKRAYLPRE SELIKERHLN DTRYASTFLK NFIEQNLQFK EAEDDNPRKRR VQTVNNGVITA 960
HFRKRWGLEK DRQETYLHHA MDAAIVACTD HHMVTRVTEY YQIKESNSKSV KKPYFPMPWE 1020
GFRDELLSHL ASQPIAKKIS EELKAGYSL DYLIVFSRMPK RSITGAAHKQ TIMRKGGIDK 1080
KGKTTIIERL GDFKMGKQED DMATYEAIKQ RYLEHGKNSK KAFETPLYKP 1140
SKKGTGNLNIK FVKVEGQAKS FVREVNNGVA QNGDLVRLVDL PEKDDKYMM PIYVPDTVCS 1200
ELPKVVVASS KGYEQWLTL NSFTFKFSLY PYDLVRLVKG DEDRFLYFGT LDIDSDRLNF 1260
KDVNKPSKKN EYRYSLKTIE DLEYEVGVL GDLRLVRKET RRNFHSGGSK RPAATKKAGQ 1320
AKKKK                                         1325

SEQ ID NO: 518      moltype = AA length = 1319
FEATURE          Location/Qualifiers
REGION           1..1319
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
REGION           1..1319
                  note = source = /note="LPG50170-nAPG07433.1protein sequence"
source            1..1319
                  mol_type = protein
                  organism = synthetic construct
SEQUENCE: 518
MAPKKKRKV YKDHDGYKD HDIDYKDDDD KMSDNELNHE RWMRHALTLA QRARDEGEV 60
VGAVLVYQONQ VIGEGWNRAI GLHDPATAHE IMALRQGLV LQNYRLIDTT LYVTPEPCVM 120
CAGAMVHSRI GQLVFGVRNS KRGAAAGSLIN VLNPYGMNHR VAITEGVLAE SCSAMLCDFY 180
RHPREQQNRL RRAAQSSGG SGGSSGSETP GTSESATPES SGCGSSGSSMR ELDYRIGLAI 240
GTNSIGWVNI RLSWNKDRER YEKVRIVDQG VRMFDRDRAEMP KTGASLAEP RIARSSRRL 300
NRKSQRKKNI RNLLVQHGV1 TQEELDSLVP LSKKSMIDIW IRLDGLDRLL NHFEWARLL 360
HLAQRRGFKS NRKSELKDTE TGKVLSSIQL NEKRLSLYRT VGEMWMKDPP FSKYDRKRNS 420
PNEYVFSVSR AELEKEIVTL FFAQRQFQSP YASKDLQETY LQIWTQHOLPV ASGNAILNKV 480
GYCSLLKGKE RRIPKATYTQF YQFSALDQVN RTRLGPDFQF FTKEQREIIL NNMFQRTDYY 540
KKKTIPEVTY YDIRKWELEL ETIQFKGLN DPNEELKKIE KKPFINLKA F YEINKVWANY 600
SERTNETFST LDYDGIGYAL TVYKTDKDIR SYLKSSHNLN KRCYDDQLIE ELLSSLSTYKF 660
GHLSLKAIND VLSIMQKGNT YKEAVDQLGY DTSGLKEKRR SKFLPPSIDE ITNPIVKRAL 720
TQARKVNVAI RRRHGSQPHS V HIELAKRNS NHDERTKIVS AQDENYKKNK GAISILSEHG 780
IILNPTGYDIV RYKLWKEQGE RCAYSLKEIP ADTFFNELLKK ERNGAPILEV DHILYPSQSF 840
IDSYHNKVLV YSDENRKKGK RIPPYTYFLET NKDWEAFERY VRSNKFFSKK KREYLLKRAY 900
LPRSELIKE RHLNDTRYAS TFLKRNFIEQN LQPKEAEDNP RKRRVQTENG VITAHFRKRW 960
GLEKDRQETY LHHAMDAIIV ACTDHMVT R VTEYYQIKES NKSVKKPYFP MPWEGFRDEL 1020
LSHLASQPIA KKISEELKAG YQSLDYIFVS RMPKRSITGA AHKQTIMRKG GIDKKGKTII 1080
IERLHLKDIK FDENGDFKMV GKEQDMATYE AIKQRYLEHG KNSKKAFTP LYKPSKKGTG 1140
NLIKRVKVEG QAKSFVREVN GGVAQNGDLV RVDLFEKDDK YYMVPIYVPD TVCSELPKKV 1200
VASSKGYEWL LTLDNSFTFK FSLVLPYDVLV LVKGDEDRLF YFGTLDIDSD RLNFKDVNKP 1260
SKNEYRYSL KTIEDLEKYE VGVLGDLRLV RKETRRNFSH GS SKRPAATK KAGQAKKKK 1319

SEQ ID NO: 519      moltype = AA length = 1320
FEATURE          Location/Qualifiers
REGION           1..1320
                  note = source = /note="Description of Artificial Sequence:
                           Synthetic polypeptide"
REGION           1..1320

```

-continued

---

```

source          note = source = /note="LPG50171-nAPG07433.1protein sequence"
               1..1320
               mol_type = protein
               organism = synthetic construct

SEQUENCE: 519
MAPKKKRKVD YKDHDG DYKD HDIDYKDDDD KMSDLEL NDE YWMRH ALTLA KRAREEGE VP 60
VGAVLV LNNQ VIGEGWN RAI GLHDPTAHAE IMALRQ GGLV LQNYRLL DAT LYVT FEP CVM 120
CS GAMVHSRI ARLVFGVRNS KRGAA GSLMN VL NYPGMN H R VEISEGV LAE SCSAMLCDFY 180
RW PPREVK NAL KKAREQNSGG SSGGSSGSET PGTSESATPE SSGGSSGSM RELDYRIGLA 240
IGTNSIGWGV IELSWNKDR E RYE KRVIDVQ GVRMF DRAEM PKTGASLA E RRRIARSSRR 300
LN RKSQRK N IRNLLVQHG V I TQEELDSL Y PLSKKSMDI W GIRL GDLR L NHF EWAR LL 360
IHLAQRRGFK SNRKS ELKDT ETGKV LSSIQ LNEKRLS LYR TVGEMW M KDF DFSKY DRKRN 420
SPNEYVFSVS RAELEKEIVT LF AAQ RRFQ S PYASKDL QET YLQI WTHQLP FASGNAIL N 480
VG YCSL LGK ERRIPKAT Y FQYFSALDV N RTRLGP D F PFTKE QREII LNNM P QRTD Y 540
YKKK TIP E VT YDIRKWL E DETIQFKGLN YDPNEELK K EKKPFINLK A FYEINKVAN 600
YSERTNETFS TLDYDGIGYA LT VYKTD KDI RSYLKSSH NL PKRCY DDQLI E ELLS LS YTK 660
FGHLSL KAIN HVLSIMQKG N TYKEA VDQLG YDT SGLKKE RSKFLPP ISD EITN PIVK RA 720
LTQARKV VNA IIRR HGSPHS VHI LARELS KNHDERTK IV SAQDEN YKKN KGAISIL SEH 780
GILNPTGYDI VRYKLW KEQG ERCAYSLK EI PADTFFNE LK KERN GAPILE VDHILP YQS 840
FIDS YHNKVL VYSDENRKKG N RIPTYFLE TNK DWEA F VRSNKF FSK KKREYLLK RA 900
YLPRESEL IK ERHLNDTRYA STFLKNFIEQ NLQFKEAEDN PRKRRV QT VN GVITA HFRKR 960
WGLEKDRQET YLHHAMDAI VACTDHHM V RTVE YYQI KE SNKSVK KPYF PMPWEGF RDE 1020
LLSHLASL QPI AKKISEELKA GYQS LDYI SRMPKRSIT D AAHK QTIMRK G GIDKKG KGT 1080
I IERLHLK QDI KFDENGDF KM VGKE QDMAT Y EA IKQ RY LEH G KNSKKA FET PLYKPSK G 1140
GNL I KRVK VE GQAKSF VREV NGGVAQNGDL VRV DLF EKDD KYYMVPI Y DTVCELPK 1200
VVA SS KG YEQ WLTLDNSFTF KFSL VPYD LV RL VKG DEDRF LYFGT L DIDS DRLN FKDV NK 1260
PSKKN EY RYS LKTIE DE KY EVGVL DRL VRKETRRN FHH SG SKR PAAT K KAGQAK KKK 1320

SEQ ID NO: 520      moltype = AA length = 1323
FEATURE           Location/Qualifiers
REGION            1..1323
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..1323
note = source = /note="LPG50172-nAPG07433.1protein sequence"
1..1323
mol_type = protein
organism = synthetic construct

SEQUENCE: 520
MAPKKKRKVD YKDHDG DYKD HDIDYKDDDD KMSDLEL NDE YWMRH ALLA KRAR DEEGE VP 60
VGAVLV LNNQ VIGEGWN RAI GLHDPTAHAE IMALRQ GGLV LQNYRLL DAT LYVT FEP CVM 120
CS GAMVHSRI ARLVFGVRNS KRGAA GSLMN VL NYPGMN H R VEISEGV LAE SCSAMLCDFY 180
RM PQQK NQO Q KAES TSSRGD SGGSSGSS SETPGTSE SA TPESSGSSG GSMREL DYRI 240
GLAIGT NSIG WGVI ELSWNK DR ERYEV K VRI VDQG V R MFDR AEPK TGA SL AEP RRIAR SS 300
R RRLN RKSQR KKNIRN LLVQ HG VIT QEELD SLYPLS KKS M DIW GIRD GL DRLN HFWEA 360
R LLIHLA QR GFKS NRK SEL KDTETG KVLS SIQLNE KRLS LYRTV GEM WM KDPDF SKYDR 420
KR NSP EYVF SVS RABE LEKE I VTLFAA RQ FQSPY ASK DL QETYLQI WTH QLPF ASGNAI 480
LN KVGC SLL KG KERR IPKA TYTF QYFSAL DQVN RTRL GP DFQPF TKE Q EII LNNM FQR 540
TD YKKK TIP E VTYY DIRK W L ELD E TIQ F K GL NYDPN B E L K KIE KKP FIN LKA F Y EINKV 600
VAN YSERT NE TF STL D YDGI GYALT VY KTD KDI RSYL KSS HNL PKRCY DD QLIE ELLS L 660
YT KFGH LSLK AINHV L SIMQ KGNT YKE AVD QLG YDTS GLK KEK RSKF LPP ISDE ITN PIV 720
KR ALTQ ARK V NAI IIRH MG PSV HIELAR ELS KNH DERT KIV SAQDEN KKN KGAISIL 780
SE H GILN PTG YD I VY KWL E EQGERC AYSL KEI PADT FFN EL KKER NGAP I LEH G HILP Y 840
SQ SFIS DYH N KVL VY SDEN R KKG N RI PYT FLET N KDW EA FERY VRSN KF FS KKK REY LL 900
KRAY L PRESE LIK ERL H LNDT RY ASTR FL KNF IE QNL QF KEA EDN P R K R RV Q TV NGV IT AHF 960
RKR WGLE KDR QET YLHH AMD A II VACT D H M VTR VTE YYQ I KES N KSVK PYFP MPWEGF 1020
R DELL SHLAS QPIAK KIS E LKAG YQSL D Y IF VSRM P KRS IT GAAH KQTI MRKG GID KKG 1080
K TII IERL HL KDI KFDENG D FK MVG K E QDM AT YEA IK QRY LEH G KNS KKA FET PLY KPSK 1140
KG TGN L KRV KVEG QAK SF V REV NGV A QN GDL VRL DLF E KDD KYYM VP Y VPD TV CSEL 1200
PK KVVA SS KG YEQ WLTLDNSFTF KFSL VPYD LV RL VKG DEDRF LYFGT L DIDS DRLN FKD 1260
V NKPSKKN EY RYS LKTIE DE KY EVGVL DRL VRKETRRN FHH SG SKR PAAT K KAGQAK KKK 1320
KKK

SEQ ID NO: 521      moltype = AA length = 1316
FEATURE           Location/Qualifiers
REGION            1..1316
note = source = /note="Description of Artificial Sequence:
                     Synthetic polypeptide"
REGION            1..1316
note = source = /note="LPG50173-nAPG07433.1protein sequence"
1..1316
mol_type = protein
organism = synthetic construct

SEQUENCE: 521
MAPKKKRKVD YKDHDG DYKD HDIDYKDDDD KMSDTELTHE YWMRH ALMLA QRAR DEEGE VP 60
VGAVLV LNNR VIGEGWN RAI GLHDPTAHAE IMALRQ GGLV LQNYRLL DAT LYVT FEP CVM 120

```

---

-continued

---

CAGAMVHGR	GTLVFGVRNL	KRGAAGSLMN	VLNYPGMNHR	VEIVEGTLSD	ECSGMLCEFV	180
RQPRLAFAQ	KQASGGSSGG	SSGSETPGTS	ESATPESSGG	SSGGSMRELQ	YRIGLAIGTN	240
SIGWGVIELS	WNKDRERYEK	VRIVDQGVRM	FDRAEMPKTG	ASLAEPRIA	RSSRRRLNRK	300
SQRKKNIRNL	LVQHGVITQE	ELDSLYPLSK	KSMDIWGI	DGLDRLLNHF	EWARLILHLA	360
QRRGFKSNRK	SELKDTEGK	VLSSIQLNEK	RLSLYRTVG	MWMKDPDFSK	YDRKRNSPNE	420
YFESVSRAEL	EKEIVTFLFAA	QRRFQSPYAS	KDLQETYLQI	WTHQLFFASG	NAILNKVGYC	480
SLLKGKERRI	PKATYTFQYF	SALDQVNTR	LGPDFQPF	EQREIILNNM	FQRTDDYYKK	540
TPEPEVTTYDI	RKWLELDETT	QFKGLNYDPN	EELKKIEKPP	FINLKAFYEI	NKVVANYSER	600
TNETFSTLDY	DGIGYALT	KTDKDIRSYL	KSSHNLPKRC	YDDQLIEELL	SLSYTKFGHL	660
SLKAINHVLS	IMQKGNTYKE	AVDQKRSK	LPPISDEIT	PIVKRALTQA	720	
RKVVNAIIRR	HGSPHSVHIE	LARELSKNHD	ERTKIVSAQD	ENYKKNKGA	SILSEHGILN	780
PTGYDIVRYK	LWKEQGERCA	YSLKEIPADT	FFNELKERN	GAPILEVDH	LPYSQSFDIS	840
YHNKVLVYSD	ENRKGNRIP	YTYPFLETNKD	WEAFERYVRS	NKFFSKKRE	YLLKRAYLPR	900
ESELIKERHL	NDTRYASTFL	KNFNQNLQF	KEAADNPRKR	RVQTVNGVIT	AHFRKRWGLE	960
KDRQETLYHH	AMDAIIVACT	DHHMVT	YYQIKESNKS	VKPKYFPMPW	EGFRDELISH	1020
LASQPIAKI	SEELKAGYQ	LDYIFVSRMP	KRSITGAHK	QTIMRKGGID	KKGKTIIEER	1080
LHLKDIKFDE	NGDFKVMVGKE	QDMATYEAIK	QRYLEHGN	KKAFETPLYK	PSKKGTLGNI	1140
KRVKVEGQAK	SFVREVNGGV	AQNQDVLVRD	LFEKDDKYMM	VPIYVPDTVC	SELPKKVVAS	1200
SKGYEQWLTL	DNSFTFKFSL	YPIYDVLRLVK	GDEDRLFLYFG	TLDIDSDRLN	FKDVNKPSK	1260
NEYRYSLKT	EDLEKYEVGV	LGDLRLV	TRRNFHSGGS	KRPAATKKAG	QAKKK	1316

---

SEQ ID NO: 522	moltype = AA	length = 1327
FEATURE	Location/Qualifiers	
REGION	1..1327	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1327	
	note = source = /note="LPG50174-nAPG07433.1protein sequence"	
source	1..1327	
	mol_type = protein	
	organism = synthetic construct	

---

SEQUENCE: 522						
MAPKKKRKVD YKDHDGDKD	HDIDYKDDDD	KMSIPELNHD	VWMRHALT	LA KRAREEGEV	P 60	
VGAVLVLNQ	VIGEGWNRAI	GLHDPTAHE	IMALRQGGLV	LQNYRLIDTT	LYVTFEP	120
CAGAMVHSR	QOLVFGVRNS	KRGAAGSLMN	VLNYPGMNHR	VEISEGVLAG	SCSAMLCDFY	180
RQPRLVKNAL	KKPAGDPSAL	QNNRGGSSG	GSSGSETPGT	SESATPESSG	GSSGSMREL	240
DYRIGLAIGT	NSIGWGVI	SWNKDRERYE	KVRIVDQGVR	MFDRAEPMKT	GASLAEPRI	300
ARSSRRRLNR	KSQRKKNIRN	LLVQHGVITQ	EELDSLYPLS	KSMIDIWGI	LDGLDRLLNH	360
FEWARLILH	AQRRGFKSNS	KSELQDQV	KVLSSIQLNE	KRLSLYRTVG	EMWMKDPDF	420
KYDRKRNSPN	EVFESVSR	LEKEIVTFLA	AQRFQSPYAS	SKDLQETYLQ	IWTHQLPFAS	480
GNAILNKVPGY	CSLLKGKERR	IPKATYTFQY	FSALDQVNRT	RLGPDFQPF	KEQREII	540
MFQRTDYKK	KTIPEVTTYD	IRKWL	ELDET IOPKGLN	YDPD NEELKKIEKK	PFINLKAFYE	600
INKVVANYSE	RTNETFSTLD	YDGIGYALT	YKTDKDIRSY	LKSSHNLPKR	CYDDQLIEEL	660
LSLSYTKFGH	LSLKAINHVL	SIMQKGNTYK	EAVDQQLGDY	SGLKKEKRSK	FLPPISDEIT	720
NPIVKRALTQ	ARKVVMAIIR	RHGPSHVH	ELARELSKNH	DERTKIVSAQ	DENYKKNKGA	780
ISILSEHGIL	NPTGYDIVRY	KLWKEQGERC	AYSLKEIPAD	TF	FNELKKER NGAPILEVDH	840
ILPYSQSFSID	SYHNKVLVYS	DENRKKG	YR	PTYFLETNK	DWEAFERYVR SNKFFSKKK	900
EYLLKRAYL	RESELIKERH	LNDTRYASTFL	LKNFIEQNLQ	P KEAEDNPRK	RRVQTVNGVI	960
TAHFRKRWGL	EKDRQBTYHL	HAMDAAIVAC	TDHHMVT	TY	YQIKESNK SVKPKYFPMP	1020
WEGFRDELLS	HLASQPIAK	ISEELKAGYQ	S LDYIFVSRM	PKRSITGAH	KQTIMRKGGI	1080
DKKGKTTI	RLHLKDIFK	ENGDFKVMV	EQDMATYEAI	KQRYLEHGN	SKKAFETPLY	1140
KPSKKGTTI	I KRVKVGQ	KSFVRENVNG	VAQNGDVLVR	DLFEKDDKYY	MVPIYVPDTV	1200
CSLELPKKVVA	SSKGYBQWL	LDNSFTFKFS	L	YPIYDVLRLV	GDEDRLFLY	1260
NFKDVNKPSK	KNEYRYSLKT	I EDLEKYEV	VLGDLRLV	GTDIDSDRL	1320	
QQAKKK						1327

---

SEQ ID NO: 523	moltype = AA	length = 1320
FEATURE	Location/Qualifiers	
REGION	1..1320	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1320	
	note = source = /note="LPG50175-nAPG07433.1protein sequence"	
source	1..1320	
	mol_type = protein	
	organism = synthetic construct	

---

SEQUENCE: 523						
MAPKKKRKVD YKDHDGDKD	HDIDYKDDDD	KMSDLELNDE	YWMRHALT	LA KRAREEGEV	P 60	
VGAVLVLNQ	VIGEGWNRAI	GLHDPTAHE	IMALRQGGLV	LQNYRLIDAT	LYVTFEP	120
CAGAMVHSR	ARLVFGVRNS	KRGAAGSLMN	VLNYPGMNHR	VEISEGVLAG	SCSAMLCDFY	180
RWPREVKNAL	KKAREQNSGG	SSGGSSGET	PGTSESATPE	SSGGSSCGSM	RELDYRIGLA	240
IGTNSIGWV	I ELSWNKDR	RYEKVIR	DQ GVRMFDRAEM	PKTGASLAEP	RRIARSSRR	300
LNKRSQRKRN	IRNLLVQHGV	ITQEELDSL	Y PLSKKMSDIW	GIRLDGLDRL	LNHFEWARLL	360
IHLAQRRGFK	SNRKSELKDT	ETGKV	LNEKRLSLYR	TVGEMWMKDP	DFS	420
SPNEYVFSVS	RAELEKEIVT	LFAAQRRFQS	PYASKDLQET	YLOQIWTHQLP	FASGNAILNK	480
VGYCSLLKGK	ERRIPKATYT	FQYFSALDQV	NRTRLGPDFQ	PFTKEQREII	LNNMFQRTDY	540

-continued

---

YKKKTIPEVT	YYDIRKWEL	DETIQFKGLN	YPDNEELKKI	EKKPFINLKA	FYEINKVAN	600
YSERTNETFS	TLDYDGIGYA	LTVYKTDKDI	RSYLKSSHNL	PKRCYDDQLI	EELLSSLSTYTK	660
FGHLSLKA	HVLSIMQKGN	TYKEAVDQLG	YDTSGLKKEK	RSKFLPPISD	EITNPIVKRA	720
LTQARKVNA	IIRRHGSPHS	VHIELARELS	KNHDERTIV	SAQDENYKKN	KGAISILSEH	780
GILNPTGVYDI	VRYKLWKEQG	ERCAYSLKEI	PADTFFNELK	KERNGAPILE	VDHILPYSQS	840
FIDSYHKV	VYSDENRKKG	NRIPTYTYFLE	TNKDWEEAFER	YVRSNKFSS	KKREYLLKRA	900
YLPRESELIK	ERHLNDTRYA	STFLKNFIEQ	NLQFKEAEDN	PRKRNQVTVN	GVITAHFRKR	960
WGLEKDRQET	YLHHAMDAII	VACTDHMMVT	RVTEYYQIKE	SNKSVKKPYF	PMPWEGFRDE	1020
LLSHLASQP	AKKISEELKA	GYQSLDYIFV	SRMPKRSITG	AAHKQTIMRK	GGIDKKGKTI	1080
IIERLHLKDI	KFDENGDFKM	VGKEQDMATY	EAIKQRYLEH	GKNSKKAFF	PLYKPSKGT	1140
GNLIKRVKVE	WQAKSFWREV	NCGVAQNGDL	VRVDSLFEKD	KYVMVPIYVP	DTVCSELPKK	1200
VVASSKGYEQ	WLTLDNSFTF	KFSLYPLDV	RLVKGDEDRF	LYFGTLIDS	DRLNPKDVNK	1260
PSKNEYRYS	LKTIEDLEKY	EVGVGLDRL	VRKETRRNFH	SGGSKRPAAT	KKAGQAKKK	1320

```

SEQ ID NO: 524      moltype = AA length = 1307
FEATURE          Location/Qualifiers
REGION           1..1307
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..1307
note = source = /note="LPG50176-nAPG07433.1protein sequence"
source            1..1307
mol_type = protein
organism = synthetic construct

```

SEQUENCE: 524						
MAPKKKRKV	YKDHDGDKD	HIDIDYKDDDD	KMSDIEQNHE	YWMRHALVLA	KRAREEGEVP	60
VGAVLVLNQ	VIGEGWNRAI	GLHDPTAHAE	IMALRQGGLV	LQNYRLIDTT	LYVTPEPCVM	120
CAGAMVHGR	GSLVFGVRNS	KRGAAGSLIN	VLYNPGMNNHR	VEMLTEGVLAD	ECSAMLCDFY	180
RHPRSGGSSG	GGSGSETPGT	SESATPESSG	GSSGGSMREL	DYRIGLAIGT	NSIGWGVIEL	240
SWNKDRERYE	KVRIIVDQGV	MFDRAEMPKT	ASLAEPRRI	ARSSRRRLNR	KSQRKKNIRN	300
LLVQHGVI	EQEELDSLYPLS	KKSMIDWPKR	LDGLDRLLNH	FEWRLLIHL	AQRRGFKSNR	360
KSELKDTEG	KVLSIQLNE	KRLSLYRTVG	EMWMKDPDS	YKDRKRNSPN	EYVFSVSRAE	420
LEKEIVLFA	AQRRFQSPYA	SKDLQETYQL	IWTHQLPFAS	GNAILNKVGY	CSLLKGKERR	480
IPKATYTFCY	FSALDQVNRT	RLGPDFQPF	KEQREIIIINN	MQRDYYKK	KTIPEVTTYD	540
IRKWLEDET	IQFKGLNYDP	NEELKNEEKK	PFINLKAFYDE	INKVVANYSE	RTNETFSTLD	600
YDGGIGYALTV	YKTDKDIFRSY	LKSSHNLPKR	CYDDQLIBEL	LSSLSTKFGH	LSLKAHNHWL	660
SIMQKGNTYK	EAVDQLGYDT	SGLKKEKRSK	FLPPISDEIT	NPIVKRALTQ	ARKVVNAIIR	720
RHGPSPHSVH	ELARELSKHN	DERTKIVSAQ	DENYKKNGA	ISILSEHGI	NPTGYDIVRY	780
KLWKEQGERC	ASLKEIPAD	TFNRELKKER	NGAPILEVHD	ILPQSQSFID	SYHNKVLVYS	840
DENRKKGNRI	PYTYFLETNK	DWEAFERYVR	SNKFFSKKKR	EYLLKRAYLP	RESELIKERH	900
LNDTRYASTF	LKNFIEBQNLO	FKEAEDNPRK	RRVQTVNGVI	TAHFRRKRWGL	EKDRQETYLH	960
HAMDAAIVAC	TDHHMVTRVT	EYYQIKESNK	SVKPYFPMP	WEGFRDELLS	HLASQPIAKK	1020
ISEELKAGYQ	SLDYIFVSRM	PKRSITGAH	KOTIMRKCGGI	DKKGKTTIE	RLHLDKIDKF	1080
ENGDFKVMVGK	EQDMATYEAII	QKRYLEHGNK	SKKAFETPLY	KPSKKGTCNL	IKRVKVEGQA	1140
KSFVREVNNG	VAQNGDLVRV	DLFEKDDKYV	MVPIYVPDTV	CSELPKKVVA	SSKGYEQWLT	1200
LDNSFTFKFS	LYPYDLVRLV	KGDEDRLFLYF	GTLIDIDSRL	NFKDVNPKSK	KNEYRYSLKT	1260
IEDLEKEYVG	VLGDLRLVRK	ETRRNHFHSGG	SKRPAATKKA	QAKKKK		1307

```

SEQ ID NO: 525      moltype = AA length = 1322
FEATURE          Location/Qualifiers
REGION           1..1322
note = source = /note="Description of Artificial Sequence:
Synthetic polypeptide"
REGION           1..1322
note = source = /note="LPG50177-nAPG07433.1protein sequence"
source            1..1322
mol_type = protein
organism = synthetic construct

```

SEQUENCE: 525						
MAPKKKRKV	YKDHDGDKD	HIDIDYKDDDD	KMCNPERDHE	YWMRHALTIA	QRARDEGEVP	60
VGAVLVLNQ	VIGEGWNRAI	GLHDPTAHAE	IMALRQGMV	LQNYRLDDTT	LYVTPEPCVM	120
CSGAMVHSRI	GTLVFGVRNE	KRGAAGSLLN	VLYGPGMNHQ	VKTIGGV LAP	ACSALLCDFY	180
RMPRQQKNQO	KAELKLSNDS	GGSSGGSSGS	ETPGTSESAT	PESSGGSSGG	SMRELDYRIG	240
LAIGTNSIGW	GVIELSWNKO	RERYEKVRI	DQGVRMFDR	EMPKTGASLA	EPRRIARSSR	300
RRLNKRKSQRK	YKRNLLVQH	QVITQKSMR	LYPLSKKSM	IWGIRLDGLD	RLLNHFEWR	360
LILIHLAQQRG	PKSNRKESEL	DTETGKVLLS	IQLNEKRLSL	RTVGEWMWK	DPDFSKYDRK	420
RNSPNEYVFS	VSRAELEKEI	VTLFAAQRRF	QSPYASKDLQ	ETYLOIWTHQ	LPFASGNAIL	480
NIVGYCSLLK	CKERRIPKAT	YTFQYFSALD	QVNTRTRLGPD	FQPFTEBQRE	IILNMMFQRT	540
DYKKKTIPE	VTYYDIRKWL	ELDETIQFKG	LNYDPNEELK	KIEKKPFI	KAFYEINKVV	600
ANYSERTNET	FSTLDYDGIG	YALTQYKTDK	DIRSYLKSS	NLPKRCYDDQ	LIEELSSLSTY	660
TKFGHLSLKA	IHNVLSIMQ	GNTYKEAVDQ	LGYDTSGLKK	EKRSKFLPPI	SDEITNPIVK	720
RALTQARKVV	NAIIRRHGSP	HSVHIELARE	LSKNHDERTK	IVSAQDENYK	KNKGAISILS	780
EHIGILNPTGY	DIVRYKLWKE	QGERCAYS	EIPADTFFNE	LKKERNGAPI	LEVHDILPYS	840
QSFIDSYH	VLYVSDENRK	KGNRIPYTF	LETNKDWEAF	ERYVRSNKF	SKKKREYLLK	900
RAYLPRESEL	IKERHLNDTR	YASTFLKNFI	EQNLOFKEAE	DNPRKRRVQ	VNGVITAHFR	960
KRWGLEKDRQ	ETYLHHAMDA	IIVACTDHMH	VTRVTEYYQI	KESNKSVKP	YFPMPWEGFR	1020

---

-continued

---

DELLSHLASQ	PIAKKISEEL	KAGYQSLDYI	FVSMPKRSTI	TGAAHKQTIM	RKGIDKKGK	1080
TIIERHLHK	DIKFDENGDF	KMVGEQDMA	TYBAIKQRYL	EHGKNSKAF	ETPLYKPSKK	1140
GTGNLIKRVK	VEGQAKSFVR	EVNGGVAQNG	DLVRVDSLFEK	DDKYYMVPYI	VPDTCSEL	1200
KKVASSKGY	BQWLTLDNSF	TFKFSLYPYD	LVRLVKGDED	RFLYFTGLDI	DSDRLNFKDV	1260
NKPSKKNEYR	YSLKTIEDLE	KYEVGVLGDL	RLVRKETRRN	FHSGGSKRP	ATKKAGQAKK	1320
KK						1322

SEQ ID NO: 526	moltype = AA	length = 1316			
FEATURE	Location/Qualifiers				
REGION	1..1316				
	note = source = /note="Description of Artificial Sequence:				
	Synthetic polypeptide"				
REGION	1..1316				
	note = source = /note="LPG50178-nAPG07433.1protein sequence"				
source	1..1316				
	mol_type = protein				
	organism = synthetic construct				
SEQUENCE: 526					
MAPKKKRKV D YKDHDGYKD	HDIDYKDDDD	KMSAIELNHE	YWMRHAGLAA	QRARDEGEVP	60
VGAVLVYQNO VIGEGWNRAI	GLHDPATAAE	IMALRQGGLV	LQNYRLIDTT	LYVTPEPCVM	120
CAGAMVHSRI GRRVFGVRNS	KRGAAGSLMN	VLNYPGMNH	VEVTEGVLAG	ECSAMLCDFY	180
RAPRAQFNAQ KRPSSGSSGG	SSGSETPGTS	ESATPESSGG	SSGGSMRELD	YRIGLAIGTN	240
SIGWGVIERS WNKDRERYEK	VRIVDQGVR	FDRAEMPKT	ASLAEPERRIA	RSSRRRLNRK	300
SQRKKNIRNL LVQHGVITQE	ELDSLYPLSK	KSMDIWGI	DLGDRLLNHF	EWARLLIHIA	360
QRRGFKSNRK SELKDTETGK	VLSIQLNEK	RLSLYRTVGE	MWMKDPDFSK	YDRKRNSPNE	420
YVFSVSRAL EKEIVTLFAA	QRRQFSPYAS	KDLQETYLQI	WTHQLPFAAG	NAILNKVGYC	480
SLLKGKERRI PKATYTFQYI	SALDQVNTR	LGPDFQPF	FQREIILNNM	FQRTDYYKKK	540
TIPEVTYVDI FKWLELDETI	QFKGLNYDPN	EELKKIEKPP	FINLKAFYEI	NKVVANYSER	600
TNETFSTLDY DGIGYALTVY	KTDKDIRSYL	KSSHNLPKRC	YDDQLIEELL	SLSYTKFGHL	660
SLKAINHVLS IMQKGNTYKE	AVDQLGYDTS	GLKKEKRSKF	LPPISDEITN	PIVKRALTQA	720
RKVVAIIIRR HGSPHSVHIE	LARELSKNHD	ERTKIVSAQD	ENYKKNGAI	SILSEHGLN	780
PTGYDIVRYK LWKEQGERCA	YSLKEIPADT	FFNELLKERN	GAPILEVDHI	LPYSQSFDIS	840
YHNKVLVYSD ENRKGNRIP	YTYPFLETND	WEAFERYVRS	NKFFSKKKRE	YLLKRAYLPR	900
ESELIKERHL NDTRYASTFL	KNFIEQNLQF	KEAADNPRKR	RVQTVNGVIT	AHFRKRWGLE	960
KDRQETLYHH AMDAIIAVCT	DHHMVTTRVTE	YYQIKESNKS	VKKPYPFPMPW	EGFRDELISH	1020
LASQPIAKKI SSELKAGYQS	LDYIFVSRMP	KRSITGAHK	QTIMRKKGID	KKGKTIIDIER	1080
LHLKDIKFDE NGDFKMGVKE	QDMATYEAIK	QRYLEHGKNS	KKAFETPLYK	PSKKGTTGNLI	1140
KRVKVEGQAK SFVREVNGGV	AQNGLDVRD	LFEKDDKYMM	VPIYVPTVC	SELPKKVVAS	1200
SKGYEQWLTL DNSFTFKFSL	YPYDLVRLVK	GDEDRFLYFG	TLDIDSDRLN	FKDVNPKSKK	1260
NEYRYSLTKI EDLEKEYEVG	LGDLRLVKE	TRRNFHSGGS	KRPAATKKAG	QAKKK	1316

SEQ ID NO: 527	moltype = AA	length = 1323
FEATURE	Location/Qualifiers	
REGION	1..1323	
	note = source = /note="Description of Artificial Sequence:	
	Synthetic polypeptide"	
REGION	1..1323	
	note = source = /note="LPG50179-nAPG07433.1protein sequence"	
source	1..1323	
	mol_type = protein	
	organism = synthetic construct	
SEQUENCE: 527		

MAPKKKRKV D YKDHDGYKD	HDIDYKDDDD	KMSNPELNHE	YWMRYALTIA	KRAREEVEP	60
VGAVLVLNER VIGEGWNRAI	GLHDPATAAE	IMALRQGGMV	LQNYRLIDTT	LYVTPEPCVM	120
CAGAMVHSRI GHLVFGVRNS	KRGAAGSLMN	VLNYPGMNH	VAITEGVLRD	ECAAMLCDFY	180
RQPQVKNAL KKTLSDSSEQ	SGGSSGKLMN	SETPGTSESA	TEPSSGGSSG	GSMRELDYRI	240
GLAIGTN SIG WGVIELSWNK	DRERYEVKRI	VQDGVRMPDR	AEMPKTGASL	AEPRIARSS	300
RRRLNRKSQR KKNIRNLLVQ	HGVITQEELD	SLYPLSKKS	DIWGIRLDGL	DRLLNHFEWA	360
RILLIHLAQRR GFKSNRKSEL	KDTETGKVLS	SIQLNEKRLS	LYRTVGEMWW	KDPDFSKYDR	420
KRNPNEVYVF SVSRAELEKE	IVTLLFAAQR	FQSPYASKDL	QETYQLQIWTH	QLPFASGNAI	480
LNKVGCVSLL KGKERRIPKA	TYTFCYFSAL	DQVNRTLGP	DFQPFKTEQR	EII1NNMFQR	540
TDYYKKK TIP EVTYYDIRK	LELDETIQFK	GLNYDPNEEL	KKIEKKPFIN	LKAFYEINKV	600
VANYSERTNE TFSTLDYDG	GYALTVYKT	KDIRSYLKSS	HNLPKRCYDD	QLIEELLSLS	660
YTKFGHLSLK AINHVLSIMQ	KGNTYKEAVD	QLGYDTSGLK	KEKRSKFLLP	ISDEITNPIV	720
KRALTQARKV VNAAIRRHS	PHSVHIELAR	ELSKNHDERT	KIVSAQDENY	KKNKGAIISIL	780
SEHGIILNPTG YDIVRYKLW	EQGERCAYSL	KEIPADTFNN	ELKKERNNGAP	ILEVDPHILY	840
SQSFIDSYHN KVLVYSDENR	KKGNRIPYTY	FLETNKDWEA	FERYVRSNKF	FSKKKREYLL	900
KRAYLPRESE LIKERHLNDT	RYASTFLKNF	IEQNLQFKEA	EDNPFRKRVQ	TVNGVITAHF	960
RKRWGLEKDR QETYLYHAMD	AIIVACTDH	MVTRVTEYYQ	IKESNKSVKI	PYFPMPWEGF	1020
RDELLSHLAS QPIAKKISEE	LKAGYQSLDY	IFVSRMPKRS	ITGAAHQKTI	MRKGIDKKG	1080
KTIIERHLH KDIKFDENGD	FKMGKQEODM	ATYEAIKORY	LEHGKNSKA	FETPLYKPSK	1140
KGTGNLIKRV KVEGQAKSFV	REVNGGVAQ	GDLVRVDSLFE	DDKYYMVPYI	YVPDTCSEL	1200
PIKKVASSKGY YEQLTLDNS	FTFKFSLYPY	DLVRLVKGDE	DRFLYFTGLD	IDSDRLNFKD	1260
VNKPSKKNEY RYSLKTIEDL	EKYEVGVLGD	RLRLRKETRR	NFHGGSKRP	AATKKAGQAK	1320
KKK					1323

-continued

---

SEQ ID NO: 528	moltype = AA length = 1322
FEATURE	Location/Qualifiers
REGION	1..1322
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1322
	note = source = /note="LPG50180-nAPG07433.1protein sequence"
source	1..1322
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 528	
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSNPEHDHE YWMRHALNLA QRARDEGEVP 60	
VGAVLVLNQQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLLDTT LYVTFEPVCV 120	
CSGAMHSRI GTVVYGVNRNE KRGAAAGSLLN VLQYPGMNNHQ VKVIGEV LAP ACSAMLCDFY 180	
RMPRQKNNQQ RAEKLAQGDS GGSSGGSSGS ETPGTSESAT PESSGGSSGG SMRELDYRIG 240	
LAIGTNSIGW CVIELSWNKD RERYEKVIRV DQGVRMFDR A EMPKTGASLA EPRIARISSR 300	
RRLNRKSQRK KNIRNLVQH GVTQEELDS LYPLSKKSM D IWGIRLDGLD RLLNHFEWAR 360	
LILHLAQRG FKSNRKSELE DTETGKVLS IQLNEKRLSL YRTVGEWMK DPDFS KYDRK 420	
RNSPNEYVFS VSRAELEKEI VTLFAAQRRF QSPYASKDLQ ETYLQIWT HQ LPFASGNAIL 480	
NKVGYCSLLK GKERRIPKAT YTQYQFSALD QVNRTTRLGP D QOPFTKEQRE II LNMMFQRT 540	
DYYKKKTIPE VTYYDIRKWL ELD E TIQFKG L NYD PNEELK KIEKKPFI NL KAFYEINKVV 600	
ANYSERTNET FSTLBYDGIG YALTVYKTDK DIRSYLKSH NLPKRCYDDQ LIEELL SLSY 660	
TKFGHLSKA INHVLSIMQK GNTYKEAVDQ LGYDTSGLKK EKR SKFLPPI SDEITNP IVK 720	
RALTQARKVV NAIIRRHS GP HSVHIELARE LSKNHDERTKT IVSAQDENYK KNKGAI SILS 780	
EHGILNPNTGY DIVRYKLWKE QGERCAYSLK EIPADTF FNE LKKER NGAPI LEVDHILPYS 840	
QSFIDSYH NK VL VYSDENRK KG NRIPYTYF LETN KDW EAF ERYVRSN KFF SKKKREY LLK 900	
RAYL PRESEL I KERH LNDTR YASTFLKNFI EQNLQFKEAE DNPRK RRVQ VNGVITA HFR 960	
KR WGLEK DRQ ETYLHHAMDA IIVACTDHMM VTRVTEY YQI KESNKS V KKP YFPMPWEGFR 1020	
DELLSHLASQ PIAK KISEEL KAGYQSLDYI FVS RMPKRSI TGA AHQQTIM RKGGIDKKGK 1080	
TIIIERLHLK DIKF DENGDF KMVGK E QDMA TYEA IKQ RYL EH GKN SKAF ETPL YKPSKK 1140	
GTGNLI KRVK VEGQAKSF VR EVNGGVAQNG DLVRV DLF EK D DKY YMV PIY VPDT VCS ELP 1200	
KV VASSK GY EQWLTLDNSF TFKFSLYPYD LV RLVKGD E RFLYFGTL D I DSDR LNF KDV 1260	
NKPSKKNEYR YSLK TIEDLE KYEVGV LGD L RLRKETRRN FHSGGS KRPA ATKKAGQAKK 1320	
KK	1322
SEQ ID NO: 529	moltype = AA length = 1322
FEATURE	Location/Qualifiers
REGION	1..1322
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1322
	note = source = /note="LPG50181-nAPG07433.1protein sequence"
source	1..1322
	mol_type = protein
	organism = synthetic construct
SEQUENCE: 529	
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSDPELNHE YWMRHALQLA QRARDEGEVP 60	
VGAVLVLNQQ VIGEGWNRAI GLHDPATAHAE IMALRQGGLV LQNYRLLDTT LYVTFEPVCV 120	
CSGAMIHSRI GTVVYGVNRNE KRGAAAGSLLN VLQYPGMNNHQ VKVIGEV LAP ACSAMLCDFY 180	
RMPRQKNNQQ RAEKLAQGDS GGSSGGSSGS ETPGTSESAT PESSGGSSGG SMRELDYRIG 240	
LAIGTNSIGW CVIELSWNKD RERYEKVIRV DQGVRMFDR A EMPKTGASLA EPRIARISSR 300	
RRLNRKSQRK KNIRNLVQH GVTQEELDS LYPLSKKSM D IWGIRLDGLD RLLNHFEWAR 360	
LILHLAQRG FKSNRKSELE DTETGKVLS IQLNEKRLSL YRTVGEWMK DPDFS KYDRK 420	
RNSPNEYVFS VSRAELEKEI VTLFAAQRRF QSPYASKDLQ ETYLQIWT HQ LPFASGNAIL 480	
NKVGYCSLLK GKERRIPKAT YTQYQFSALD QVNRTTRLGP D QOPFTKEQRE II LNMMFQRT 540	
DYYKKKTIPE VTYYDIRKWL ELD E TIQFKG L NYD PNEELK KIEKKPFI NL KAFYEINKVV 600	
ANYSERTNET FSTLBYDGIG YALTVYKTDK DIRSYLKSH NLPKRCYDDQ LIEELL SLSY 660	
TKFGHLSKA INHVLSIMQK GNTYKEAVDQ LGYDTSGLKK EKR SKFLPPI SDEITNP IVK 720	
RALTQARKVV NAIIRRHS GP HSVHIELARE LSKNHDERTKT IVSAQDENYK KNKGAI SILS 780	
EHGILNPNTGY DIVRYKLWKE QGERCAYSLK EIPADTF FNE LKKER NGAPI LEVDHILPYS 840	
QSFIDSYH NK VL VYSDENRK KG NRIPYTYF LETN KDW EAF ERYVRSN KFF SKKKREY LLK 900	
RAYL PRESEL I KERH LNDTR YASTFLKNFI EQNLQFKEAE DNPRK RRVQ VNGVITA HFR 960	
KR WGLEK DRQ ETYLHHAMDA IIVACTDHMM VTRVTEY YQI KESNKS V KKP YFPMPWEGFR 1020	
DELLSHLASQ PIAK KISEEL KAGYQSLDYI FVS RMPKRSI TGA AHQQTIM RKGGIDKKGK 1080	
TIIIERLHLK DIKF DENGDF KMVGK E QDMA TYEA IKQ RYL EH GKN SKAF ETPL YKPSKK 1140	
GTGNLI KRVK VEGQAKSF VR EVNGGVAQNG DLVRV DLF EK D DKY YMV PIY VPDT VCS ELP 1200	
KV VASSK GY EQWLTLDNSF TFKFSLYPYD LV RLVKGD E RFLYFGTL D I DSDR LNF KDV 1260	
NKPSKKNEYR YSLK TIEDLE KYEVGV LGD L RLRKETRRN FHSGGS KRPA ATKKAGQAKK 1320	
KK	1322
SEQ ID NO: 530	moltype = AA length = 1325
FEATURE	Location/Qualifiers
REGION	1..1325
	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1325

-continued

---

```

source          note = source = /note="LPG50182-nAPG07433.1protein sequence"
                1..1325
                mol_type = protein
                organism = synthetic construct

SEQUENCE: 530
MAPKKKRKVD YKDHDGDYKD HDIDYKDDDD KMSNPELNHE YMRYALTIA KRARDEGEVP 60
VGAVLVYHDQ VIGEGWNRAI GLHDPTAAHE IMALRQGGLV LQNYRLIDTT LYVTFEPCVM 120
CAGAMVHSRI GRLVEGVRNS KRGAAGSLLN VLNYPGMNHQ IDMEEGVLRD ECAAMLCDFY 180
RLPRIVKNAL QSPPDSTNL HASGGSSGGS SGSETPGTSE SATPESSGGS SGGSMRELDY 240
RIGLAIGLNS LGWGVIELSW NKDRERYEKV RIVDQGVFRMF DRAEMPKTGA SLAEPRIAR 300
SSRRLNRKS QRKKNIRNLN VQHGVITQEE LDLSLYPLSKK SMDIWGIRLD GLDRLLNHFE 360
WARLLIHLAQ RRGFKSNRKS ELKDTEGKV LSSIQLNEKR LSLYRTVGEN WMKDPPFSKY 420
DRKRNSPNEY VFSVSRAELE KEIVTLFAQQ RRFQSPYASK DLQETYLQIW THQLPFASGN 480
A1LNKVGCS LLKGKERRIP KATYTFQYFS ALDQVNTRL GDPFQFTKE QREIILNNMF 540
QRTDYYKKKT IPEVTVYDII KWLLEDETIQ FKGLNYPDPN ELKKIEKKPF INLKFYEIN 600
KVVANYSERT NETFSTLDYD GIGYALTIVK TDKDIRSYLK SSHNLPKRCY DDQLIEELLS 660
LSYTKFGHLS LKAINHVLSI MQKGNTYKEA VDQLGYDTSG LKKEKRSKFL PPISDEITNP 720
IVKRALTQAR KVNVNIAIRR GSPHSVHIEL ARELSKNHDE RTKIVSAQDE NYKKNKGAI 780
ILSEHGLNP TGDIVRYKL WKEQGERCAY SLKEIPADTF FNELKKERNG APILEVDHIL 840
PSYQSFDISY HNKVLVYSDE NRKKGNRIPY TYFLETNKDW EAFERVVRSN KFFSKKKREY 900
LLKRAYLPLR SELIKERHLM DTRYASTFLK NFIEQNLQFK EAEDNPRKRR VQTVNGVITA 960
HIFKRWGLEK DRQETYLHHQ MDAIIVACTD HHMVTRVTEY YQIKESNKS  KKPYPPMPWE 1020
GFRDELLSHL ASQPIAKKIS EELKAGYQSL DYIFVSRMPK RSITGAHKQ TIMRKGGIDK 1080
KGKTI111ERL HLKD1KFDEN GDFKMMVGKEQ DMATYEAIKQ RYLEHGKNSK KAFETPLYKP 1140
SKKGTGNLIK RVKVEGQAKS FVREVNGVA QNGDLVRVLD FEKDDKYYMV PIYVPDTVCS 1200
ELPKKVVASS KGYEQLTLD NSFTKFSLY PYDVLVRLVKG DEDRFLYFGT LDIDSDRLNF 1260
KDVNKPSKKN EYRYSLKTIE DLEYEVGVL GDLRLVRKET RRNFHSGGSK RPAATKKAGQ 1320
AKKKK                                         1325

SEQ ID NO: 531      moltype = RNA length = 130
FEATURE           Location/Qualifiers
misc_feature      1..130
                  note = source = /note="Description of Artificial Sequence:
                  Synthetic polynucleotide"
misc_feature      1..130
                  note = source = /note="SGN000139"
source            1..130
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 531
aggttttaat gccccaggcct gtcatagttc cattaaagcc aaaagtggct ttgatgttcc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctcg cccattgaaa tgggcttctc 120
cccatttatt                                         130

SEQ ID NO: 532      moltype = RNA length = 130
FEATURE           Location/Qualifiers
misc_feature      1..130
                  note = source = /note="Description of Artificial Sequence:
                  Synthetic polynucleotide"
misc_feature      1..130
                  note = source = /note="SGN000143"
source            1..130
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 532
catggcagta cattagagca gtcatagttc cattaaagcc aaaagtggct ttgatgttcc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctcg cccattgaaa tgggcttctc 120
cccatttatt                                         130

SEQ ID NO: 533      moltype = RNA length = 130
FEATURE           Location/Qualifiers
misc_feature      1..130
                  note = source = /note="Description of Artificial Sequence:
                  Synthetic polynucleotide"
misc_feature      1..130
                  note = source = /note="SGN000186"
source            1..130
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 533
ggacagtgcg catctccctg gtcatagttc cattaaagcc aaaagtggct ttgatgttcc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctcg cccattgaaa tgggcttctc 120
cccatttatt                                         130

SEQ ID NO: 534      moltype = RNA length = 130
FEATURE           Location/Qualifiers

```

---

-continued

---

```

misc_feature          1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..130
note = source = /note="SGN000194"
source                1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 534
gcgcacagc attcaggctg gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctg cccattgaaa tgggcttctc 120
cccatattatt                                130

SEQ ID NO: 535      moltype = RNA length = 135
FEATURE
misc_feature          1..135
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..135
note = source = /note="SGN000930"
source                1..135
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 535
gaacaactca aatggaaatg aatatgtcat agttccatgaa aagccaaaag tggcttgat 60
gtttctatga taagggttgc ggcccggtggc gtcggggatc gcctgcccattccgatggc 120
tttccccat ttatt                                135

SEQ ID NO: 536      moltype = RNA length = 130
FEATURE
misc_feature          1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
misc_feature          1..130
note = source = /note="SGN001681"
source                1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 536
ccgtgccttg acctaccctg gtcatagttc cattaaagcc aaaagtggct ttgatgttc 60
tatgataagg gtttcgaccc gtggcgctgg ggatcgctg cccattgaaa tgggcttctc 120
cccatattatt                                130

SEQ ID NO: 537      moltype = DNA length = 20
FEATURE
misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="SGN000139 target sequence"
source                1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 537
aggttttaat ggcccgagctt                                20

SEQ ID NO: 538      moltype = DNA length = 20
FEATURE
misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="SGN000143 target sequence"
source                1..20
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 538
catggcagta cattagagca                                20

SEQ ID NO: 539      moltype = DNA length = 20
FEATURE
misc_feature          1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature          1..20
note = source = /note="SGN000186 target sequence"

```

---

---

-continued

---

```

source          1..20
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 539
ggacagtgcg catctccctg                                         20

SEQ ID NO: 540      moltype = DNA  length = 20
FEATURE
misc_feature        Location/Qualifiers
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..20
note = source = /note="SGN000194 target sequence"
source             1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 540
gccgcacagc attcaggatcg                                         20

SEQ ID NO: 541      moltype = DNA  length = 25
FEATURE
misc_feature        Location/Qualifiers
1..25
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..25
note = source = /note="SGN000930 target sequence"
source             1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 541
gaacaactca aatggaaatg aatat                                         25

SEQ ID NO: 542      moltype = DNA  length = 20
FEATURE
misc_feature        Location/Qualifiers
1..20
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..20
note = source = /note="SGN001681 target sequence"
source             1..20
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 542
ccgtgccttg acctaccctg                                         20

SEQ ID NO: 543      moltype = DNA  length = 53
FEATURE
misc_feature        Location/Qualifiers
1..53
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..53
note = source = /note="SGN000139 forward primer"
source             1..53
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 543
tcgtcgccag cgtcagatgt gtataagaga cagctttagt ctggagggcc atc      53

SEQ ID NO: 544      moltype = DNA  length = 52
FEATURE
misc_feature        Location/Qualifiers
1..52
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
misc_feature        1..52
note = source = /note="SGN000143 forward primer"
source             1..52
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 544
tcgtcgccag cgtcagatgt gtataagaga cagacatttgc acgagcagcg aa      52

SEQ ID NO: 545      moltype = DNA  length = 53
FEATURE
misc_feature        Location/Qualifiers
1..53
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"

```

---

---

-continued

---

```

misc_feature          1..53
source                note = source = /note="SGN000186 forward primer"
                     1..53
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 545
tcgtcggcag cgtcagatgt gtataagaga cagtggcccc tatgtggaga tca      53

SEQ ID NO: 546      moltype = DNA length = 53
FEATURE             Location/Qualifiers
misc_feature         1..53
                     note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature         1..53
                     note = source = /note="SGN000194 forward primer"
source               1..53
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 546
tcgtcggcag cgtcagatgt gtataagaga cagatgacat tcaggccaca gtg      53

SEQ ID NO: 547      moltype = DNA length = 53
FEATURE             Location/Qualifiers
misc_feature         1..53
                     note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature         1..53
                     note = source = /note="SGN000930 forward primer"
source               1..53
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 547
tcgtcggcag cgtcagatgt gtataagaga caggacaccc aagagggttt gcc      53

SEQ ID NO: 548      moltype = DNA length = 53
FEATURE             Location/Qualifiers
misc_feature         1..53
                     note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature         1..53
                     note = source = /note="SGN001681 forward primer"
source               1..53
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 548
tcgtcggcag cgtcagatgt gtataagaga cagtgggttga actggacggg gat      53

SEQ ID NO: 549      moltype = DNA length = 54
FEATURE             Location/Qualifiers
misc_feature         1..54
                     note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature         1..54
                     note = source = /note="SGN000139 reverse primer"
source               1..54
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 549
gtctcgtggg ctggagatgt tgtataagag acagtgtgg caaatctagt ctcg      54

SEQ ID NO: 550      moltype = DNA length = 54
FEATURE             Location/Qualifiers
misc_feature         1..54
                     note = source = /note="Description of Artificial Sequence:
                     Synthetic oligonucleotide"
misc_feature         1..54
                     note = source = /note="SGN000143 reverse primer"
source               1..54
                     mol_type = other DNA
                     organism = synthetic construct
SEQUENCE: 550
gtctcgtggg ctggagatgt tgtataagag acagggcccc tggagaggtt tttaa      54

SEQ ID NO: 551      moltype = DNA length = 54
FEATURE             Location/Qualifiers
misc_feature         1..54

```

---

---

-continued

---

```

        note = source = /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"
misc_feature    1..54
        note = source = /note="SGN000186 reverse primer"
source         1..54
        mol_type = other DNA
        organism = synthetic construct
SEQUENCE: 551
gtctcgtggg ctcggagatg tgtataagag acagggcaga gtcagccctc atag      54

SEQ ID NO: 552          moltype = DNA length = 54
FEATURE
misc_feature    1..54
        note = source = /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"
misc_feature    1..54
        note = source = /note="SGN000194 reverse primer"
source         1..54
        mol_type = other DNA
        organism = synthetic construct
SEQUENCE: 552
gtctcgtggg ctcggagatg tgtataagag acagcttccc ctattcagg ccca      54

SEQ ID NO: 553          moltype = DNA length = 54
FEATURE
misc_feature    1..54
        note = source = /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"
misc_feature    1..54
        note = source = /note="SGN000930 reverse primer"
source         1..54
        mol_type = other DNA
        organism = synthetic construct
SEQUENCE: 553
gtctcgtggg ctcggagatg tgtataagag acagctgtcc cttgcagctt ctgt      54

SEQ ID NO: 554          moltype = DNA length = 54
FEATURE
misc_feature    1..54
        note = source = /note="Description of Artificial Sequence:
        Synthetic oligonucleotide"
misc_feature    1..54
        note = source = /note="SGN001681 reverse primer"
source         1..54
        mol_type = other DNA
        organism = synthetic construct
SEQUENCE: 554
gtctcgtggg ctcggagatg tgtataagag acagcagctt gtggccagg atgt      54

SEQ ID NO: 555          moltype = AA length = 1368
FEATURE
REGION          1..1368
        note = source = /note="Streptococcus pyogenes Cas9"
source
        mol_type = protein
        organism = Streptococcus pyogenes
SEQUENCE: 555
MDKKYSIGLD IGTNSVGWAV ITDDYKVPSK KLKGNGNTDR HGIKKNLIGA LLFDSGETA 60
ATRLKR TARR FYTRRRKNCR YLQEIFSNEM AKVDDDSFFHR LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLADSTDKV LRLIYLALAH MIKFRGHFLI EGDLNPNDSD 180
VDKLFIFIQLVQ TYNQLFEENP INASRVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDLDNLLA QIGDQYADLF LAAKNLSDAT 300
LILSDILRVNS EITKAPLSAS MIKRYDEHHQ DLTLKLALVR QOLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKFIKPIL EKMDGTEELL AKLNRDELLR KQRTFDNGSI PYQIHLGELH 420
AILRRQEDFY PFLKDNREKI ETILFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE 480
VVDKGASAQS PIERMTNFDE NLPNEKVLPK HSLLYELYPTV YNELTKVYVV TEGMRKPAFL 540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI 600
IKDKDFLDNE ENEDILEDIV LTTLTFEDRE MIEERLKTYA HLFDDKVMKQ LKRRYYTGWG 660
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSDL 720
HEHIANLAGS PAIKKGILQT VKVVDDELVKV MGRHKPENIV IEMARENQTT QKGQKNSRER 780
MKRIEEGIKE LGSDILKEYP VENTQLQNPK LYLYYLQNQR DMVYDQELDI NRLSDYDVHD 840
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQQLNAK LITQRKFDSL 900
TKAERGGLSE LDKVGVPIKRQ LVETRQITKH VAQILDLSRMN TKYDENDKLI REVVRVITLKS 960
KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVGTALIKK YPKLESEFVY GDYKVDVRK 1020
MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF 1080
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNNSDKLI ARKKDWDPKK YGGFDSPTVA 1140

```

-continued

---

YSVLLVVAKVE KGKSKKLKV KELLGITIME RSSFEKDPID FLEAKGYKEV RKDLIILPK	1200
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLGSPPE DNEQQLFVE	1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRSK PIREQAENII HLFTLTNLGA	1320
PAAFKYFDTT IDRKYRTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD	1368

SEQ ID NO: 556 moltype = AA length = 1368

FEATURE Location/Qualifiers  
REGION 1..1368  
note = source = /note="Description of Artificial Sequence:  
Synthetic polypeptide"  
REGION 1..1368  
note = source = /note="Streptococcus pyogenes Cas9 D10A  
nickase"  
source 1..1368  
mol\_type = protein  
organism = synthetic construct

SEQUENCE: 556

MDKYSIGLA IGTNSVGWAV ITDDYKVPSK KLKGLGNTDR HGICKNLIGA LLFDSGETA	60
ATRLKRTARR RYTRRKNRIC YLQEIFSNEA AVKDDSFHRR LEESFLVEED KKHERHPIFG	120
NIVDEVAYHE KYPTIYHLRK KLADSTDKVD LRLIYLALAH MIKFRGHFLI EGDLNPNSD	180
VDKLFQIQLVQ TYNQLEFEENP INASRVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN	240
LITALSILGLTP NFKSNPDLAP DAKLQLSKDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAT	300
LLSDILRWN EITKAPLASAS MIKRYDEHH DLTLLKALVR QQLPEKYKEI FFDQSKNGYA	360
GYIDGGASQE EFYKFIKPII EKMDGTEELL AKLNRDELLR KORTFDNGSI PYQIHLGELH	420
AJLRRQEDFY PFLKDNRREKI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPNFEE	480
VVDKGASAQS PIERMTNFDK NLPNEKVLPK HSLLYEFPTV YNELTKVYY TEGMRKP AFL	540
SGEQKKAIVD LLFKTNRKV VQQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI	600
IJKDKDFLDNE ENDEDILEDIV LTTLTFLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRTGWG	660
RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL	720
HEHIANLAGS PAIKKGILQ TKVVKDELVK MGRHKPENI IEMARENQTT QKGQKNSRER	780
MKRIEEGIKE LGSDILKEYP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVH	840
IVPQSFLLKDD SIDNKVLTRS DKNRQKSDNV PSEEVVKKM K NYWRQLLNK LITQRKF DNL	900
TKAERGGLSE LDKVGFPIKRQ LVETRQITKH VAQILDSSRMN TKYDENDKLI REV RVT LKTS	960
KLVSDFRKDF QFYKVREINN YHHADAYLN AVVGTALIKK YPKLESEFVY GDYKVYDV R	1020
MIAKSEQEIG NIMMFNKTIE TLANGEIRKR PLIETNGETG EIVWDKGRDF	1080
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNDSKLI ARKKDWP K YGGFDSP TVA	1140
YSVLLVVAKVE KGKSKKLKV KELLGITIME RSSFEKDPID FLEAKGYKEV RKDLIILPK	1200
YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLGSPPE DNEQQLFVE	1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRSK PIREQAENII HLFTLTNLGA	1320
PAAFKYFDTT IDRKYRTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD	1368

SEQ ID NO: 557 moltype = AA length = 1388

FEATURE Location/Qualifiers  
REGION 1..1388  
note = source = /note="Streptococcus thermophilus Cas9"  
source 1..1388  
mol\_type = protein  
organism = Streptococcus thermophilus

SEQUENCE: 557

MTKPYSIGLD IGTNSVGWAV TTDNYKVPSK KMVKLGNTSK KYIKKNLLGV LLFDSGITAE	60
GRRLKRTARR RYTRRKNRIL YLQEIFSTEM ATLDDDAFFQRL LDDSFLVPDD KRDSKYP IFG	120
NLVEEKAYHD EPFTIYHLRK YLADSTKKAD LRLVYLALAH MIKFRGHFLI EGEFNSKNND	180
IQKNFQDFLD TYNAIFESDL SLENSKQLEE IVKDKISKLE KKDRILKLF P GEKNSGIFSE	240
FLKLIVGNQA DFRKCFNLDE KASLHFSKES YDEDLETLLG YIGDDYSDVF LKAKKLYDAI	300
LLSGFLTVTD NETEAPLSSA MIKRYNEHKE DLALLKEYIR NISLKTYNEV FKDDTKNGYA	360
GYIDGKTNQE DFYVYKLKLL AEFEGADYFL EKIDREDFLR KORTFDNGSI PYQIHLQEMR	420
AJLDKQAKFY PFLAKNKERI EKILTFRIPY YVGPLARGNS RFAWSIRKRN EKITPNFEE	480
VIDKESSAEE FINRMTSFDL YLPEEKVLPK HSLLYEFPTV YNELTKVRFI AESMRDYQFL	540
DSKQKKDIVER LYFKDKRKV DDKDIEYLA IYGYDGIELK GIEKQFNSL STYHDLLNII	600
NDKEFLDDSS NEAIIIEIIH TLTDFREM IKQRLSKFEN IFDKSVLKKL SRRHTGWGK	660
LSAKLINGR DEKSGNTLQ YLIDDGISNR NFQMLIHDAA LSFKKKIQKA QIIGDEDKGN	720
IKEVVKSLPG SPAIKKGILQ SIKIVDELVK VMGGRKPESI VVEMARENQY TNQGKNSSQ	780
RLKRLEKSLK ELGSKILKEN I PAKL SKIDN NALQNDRLYL YLQNGKDMY TGDDLDIDRL	840
SNYDIDHIIP QAFKLKNSID NVKLVSSASN RGKSDDVPSL EVVKKRKTW YQLLKSKLIS	900
QRKF DNLS PEDK AGFIQRLV E TRQITKHW AR LLDEKFNNKK DENNRAVRTV	960
KIITLKSTLV SQFRKDFE KVREINDFH AHDAYLNAV ASALLKKYPK LEPEPVYGDY	1020
PKYNSFRERK SATEKVFYS NIMNIFKKS1 SLADGRVIER PLIEVNEETG ESVWNKESDL	1080
ATVRRVLSYP QVNIVKKVEE QNHGLDRGK P KGLFNANLSS KP KPN SNENL V GAKEYLDPK	1140
KYGGYAGISN SFTV L VKG T EKGAKKKITN VLEFOQGISIL DRINYRKDKL NFLLEKGYKD	1200
IELIIELPKY SLFELSDGR RMLASILSTN NKRGEIHKG N QIFLSQKFV LLYHAKRISN	1260
TINENHRYV ENHKKFEEF FYYI LEFNEN YVGA KNGKL LNSAFQSWQN HSIDE L CSSF	1320
IGPTGSE RKG LFELTSRGSA ADFEFLGVKI PRYRDYTPSS LLKDATLIHQ SVTGLYETRI	1380
DLAKLGE G	1388

SEQ ID NO: 558 moltype = AA length = 1388  
FEATURE Location/Qualifiers

-continued

---

REGION	1..1388	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1388	note = source = /note="Streptococcus thermophilus Cas9 D10A nickase"
source	1..1388	mol_type = protein organism = synthetic construct
SEQUENCE: 558		
MDKPYSIGLA IGTNSVGAWA TTDNYKVPSK KMVKVLGNTSK KYIKKNLLGV LLFDSGITAE 60		
GRRLKRTARR RYTRRRNRL YLQEIFSTEM ATLDDDAFFQR LDDDSFLVPDD KRDSKYPIFG 120		
NLVEEKAYHD EFPTIYHLRK YLADSTKKAD LRLVYLALAH MIKYRGHFLI EGEFNSKNND 180		
IQKNFQDFLD TYNAIFESDL SLENSENQLEE IVKDKISKLE KKDRILKLFV GEKNSGIFSE 240		
FLKLIVGNQA DFRKCFNLDE KASLHSKES YDEDLETLLV YIGDDYSDFV LKAKKLYDAI 300		
LLSGFLTVTD NETEAPLSSA MIKRYNEHKE DLALLKEYIR NISLKTYNEV FKDDTKNGYA 360		
GYIDGKTNQE DFYVYLLKKLL AEFEGADYFL EKIDREDFLR KQRTFDNGSI PYQIHLQEMR 420		
AILDQKAFY PFLAKNKERI EKILTFRIPY YVGPLARGNS DFAWSIRKRN EKITPWNFED 480		
VIDKESSAEA FINRMTSFDL YLPEEKVLPK HSLLYETFVN YNELTKVRFI AESMRDYQFL 540		
DSKQKKDIVER LYFKDKRVT DKDIIYEYLAH IYGYDGIELK GIEKQFNSSL STYHDLLNII 600		
NDKEFLDDSS NEAIIIEIIH TLTIIFEDREM IKQRLSKFEN IFDKSVLK1 SRRHYTGWKG 660		
LSAKLINGIR DEKSGNTILD YLIDDGISNR NFMQLIHDDA LSFKKKIQKA QIIGDEDKGN 720		
IKEVVKSPLPG SPAIKKGILQ SIKIVDELVK VMGGRKPESI VVEMARENQY TNQGKSNSQQ 780		
RLKRLEKSLK ELGSKILKEN IAPLKLSDN NALQNDRLYL YLQLQNGKDMY TGDDLDIDR1 840		
SNYDIDHIIP QAFLKDNSID NKVLVSSASN RGKSDDVPSL EVVKKRKTFW YQLLKSKLIS 900		
QRKFDPNLTKA ERGGGLSPEDK AGFIQRLVE TRQITKHVAR LLDEKFNNKK DENNRAVRTV 960		
KIITLKSTLQ SFQRKDFELY KVREINDFH AHDAYLNAV1 ASALLKYPK LEPEFVYGDY 1020		
PYKNSFRERK SATEKVYFYS NIMNIFKKSI SLADGRVIER PLIEVNEETG ESVWNKESDL 1080		
ATVRRVLSYP QVNVVKKVEE QNHGLDRGKP KGLFNANLSS KPKPNSNENL VGAKEYLDPK 1140		
KYGGYAGISN SFTVLVKGTI EKGAKKKITN VLEFQGISIL DRINYRKDKL NFLLEKGYKD 1200		
IELIIIELSLPKY SFELSDLGSR RMLASILSTN NKRGEIHKGK LLYHAKRISN 1260		
TINENHHRKVY ENHKKEFEEL FYYILEFEN YVGAKKNGKL LNSAFQSWN HSIDDELCSSF 1320		
IGPTGSERKG LFELTSRGSA ADFEFLGVKI PRYRDYTPSS LLKDATLHQ SVTGLYETRI 1380		
DLAKLGEG	1388	
SEQ ID NO: 559		moltype = AA length = 1368
FEATURE		Location/Qualifiers
REGION	1..1368	note = source = /note="Staphylococcus aureus Cas9"
source	1..1368	mol_type = protein organism = Staphylococcus aureus
SEQUENCE: 559		
MDKPYSIGLA IGTNSVGAWA ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE 60		
ATRLKRTARR RYTRRKNRIC YLQEIFSNEM AKVDDSFHRR LEESFLVEED KKHERHPIFG 120		
NIVDEVAYHE KYPTIYHLRK KLVSTDKD LRLIYLALAH MIKFRGHFLI EGDLNPDNSD 180		
VDKLFQIQLVQ TYNQFLFEENP INASQVDAKA ILSARLSKSR RLENLTAQLP GEKKNGLFGN 240		
LIALSGLGLTP NPKSNPDLAE DAQKLQLSDKT YDDDDLNLLA QIGDQYADLF LAAKNLSDAI 300		
LLSDIILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKLALVR QQLPEKYKEI FFDQSKNGYA 360		
GYIDGGASQE EFYKFKIPIL EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH 420		
AILRRQEDFY PFLKDNRKEI EKILTFRIPY YVGPLARGNS RFAWMTRKSE ETITPWNFEE 480		
VVDKGASAQS FIERMTNFDK NLPNEKVLPK HSLLYEFVY YNELTKVKYV TEGMRKP AFL 540		
SGEQOKKAIVD LLFKTNRKV1 VQQLKEDYFK K1ECFDSEVI SGVEDRFNAS LGTYHDLK1 600		
IKDKDFLDNE ENEDILEDIV LTTLTFEDRE MIEERLKYTA HLFDDKVMKQ LKRRRTGWG 660		
RLSRKLINGI RDKQSGKTIL DFLKSDGFFAN RNFQMQLIHD SLTFKEDIQK AQVSQGDSL 720		
HEHIANLAGS PAIKKGILQT KVVKVDELVK MGRHKPENIV IEMARENQTT QKGQKNSRER 780		
MKRIEEGIKE LGSQLILKEHP VENTOLQNEK LYLYYLQNGR DMVYDQELDI NRSLSDYVDH 840		
IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKMKM NYWRQLLNAK LITQRKF DNL 900		
TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAJILD SRMN TKYDENDKLI REVKVITLKS 960		
KLVSDFRKDF QFYKVRREINN YHHAHDAYLN AVVGTALIKV YPKLESEFVY GDYKVYDVRK 1020		
MTAKSEQIG KATAKYFFYS NIMNIFKTEI TLANGEIRKR PLIETNGETG EIVVWDKGDRF 1080		
ATVRKVLSMP QVNIVKKTEV QTGGFSKESI LPKRNNSDKL1 ARKKDWDPKK YGGFDSP TVA 1140		
YSVLLVVAKVE KGKSKKLKS V KELLGITIME RSSFEKPNID FLEAKGYKEV KKDLI1KLPK 1200		
YSLFELENGR KRMLASAGEL QKGNEALPS K VVNFLYLAS HYEKLKGSPE DNEQKQLFVE 1260		
QHKHYLDEII EQISEFSKRV ILADANLDK1 LSAYNKHRDK PIREQAENII HLFTLTNLGA 1320		
PAAFKYFDTT IDRKYRTSTK EVLDATLHQ SITGLYETRI DLSQLGGD	1368	
SEQ ID NO: 560		moltype = AA length = 1368
FEATURE		Location/Qualifiers
REGION	1..1368	note = source = /note="Description of Artificial Sequence: Synthetic polypeptide"
REGION	1..1368	note = source = /note="Staphylococcus aureus Cas9 D10A nickase"
source	1..1368	

-continued

---

```

mol_type = protein
organism = synthetic construct

SEQUENCE: 560
MDKYSIGLA IGTNSVGWAV ITDEYKVPSK KFKVLGNTRD HSIKKNLIGA LLFDSGETAE 60
ATRLKRTARR RYTRRKNRIC YLQEFSNEM AKVDDSFHRR LEESFLVEED KKHERHPIFG 120
NIVDEVAYHE KYPTIYHLRK KLVSTDKDAD LRLIYLALAH MIKFGRGHFLI EGDLNPNDSD 180
VDKLFIFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN 240
LIALSLGLTP NFKSNFDLAP DAKLQLSBDT YDDDLDNLLA QIGDQYADLF LAAKNLSDAI 300
LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLKALVR QQLPEKYKEI FFDQSKNGYA 360
GYIDGGASQE EFYKPIKPL EKMDGTEELL VKLNREDLIR KORTFDNGSI PHQIHLGELH 420
AIIIRRQEDPY PFLKDNRKEI EKILTRIPY VYGPLARGNS RFAWMTRKSE ETITFWNFE 480
VVDKGASAQS PIERMTNFDE NLPNEKVLPK HSLLYEYPTV YNELTKVKVV TEGMRKP AFL 540
SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLKI 600
IKDKDFLDNE ENEDILEDIV LTLLTDRE MIEERLKTYA HLFDDKVMKQ LKRRRTGNG 660
RLSRKLINGI RDQSGKTTIL DFLKSGDFAN RNFMQLIHD SLTFKEDIQK AQVSGQGDSL 720
HEHIANLAGS PAIKKGILQV KVVDDELVKV MGRHKPENIV IEMARENQTT QKGOKNSRER 780
MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNDR DMYVQDQELDI NRSLSDYDWDH 840
IVPQSFLKDD SIDNKVLTRS DKRNQKSDNV PSEEEVVKMMK NYWRQOLLNAK LITQRKFNDL 900
TKAERGLLSE LDKGAGFQKQ LVETRQDFDRE MIEERLKTYA HLFDDKVMKQ LKRRRTGNG 960
KLVSDFRKDF QFYKVRREINN YHHADAYLN AVVGTALIKK YPKLESEFVY GDYKVYDVRK 1020
MIAKSEQEIG KATAKYFFYS NIMNNFKTEI TLANGEIRRK PLIETNGETG EIVWDKGRDF 1080
ATVRKVLMSMP QVNIVRKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPVVA 1140
YSVLVVAKVE GKGSKKLKV S KELLNGP RSSFEKNPID FLEAKGYKEV KKDLIILKLPK 1200
YSLFELENGR KRMLASAGEL QKGNEALAPS KVYNFLYLAS HYEKLGSPPE DNEQKQLFVE 1260
QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHDRK PIREQAENII HLFTLTNLGA 1320
PAAFKYPDTT IDRKRKTSTK EVLDATLHQH SITGLYETRI DLSQLGGD 1368

```

```

SEQ ID NO: 561      moltype = AA length = 1300
FEATURE           Location/Qualifiers
REGION            1..1300
source             note = source = /note="Francisella novicida Cas12a"
                  1..1300
mol_type = protein
note = novicida
organism = Francisella sp.

```

```

SEQUENCE: 561
MSIYQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLLID EKRAKYKKA KQIIDKYHQF 60
FIEBILSSVC ISEDQLQNYS DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK 120
NLFNQNLIDA KKGQESDLIL WLKQNLKDNGI ELFQKNSDIT DIDEALEIJK SFKGWTYFK 180
GFHENRKVNY SSNDIPTII YRIVDDNLPK FLENKAKYES LKDKAPEAIN YEQIKKDLAE 240
ELTFDIDYKT SEVNQRFVSL DEVPEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKG 300
NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA 360
AFKTVEEKS1 KETLSSLFDD LKAKTQLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY 420
ITQQIAPKNL DNPSSKKEQEL IAKKTEKAKY LSLETIKLAL KQCRFEEILA 480
NFAAIPMIFD EIAQNKDNLA QISIKYQNCQ KQDQLLQSAEE DDVKAIDKLL DQTNNLLHKL 540
KIFHISQSED KANILDKDEH FYLVEEFCYF ELANIVPPLYN KIRNYITQKP YSDEKFKLFN 600
ENSTLANGWD NKKEPDNTAA LFIIKDQYKL GVMNKKNNKI FDDKAIKENK GEGYKIVYK 660
LLPGANKMLP KVFFSAKS1K YPNPSEDLR IRNHSTHTKIN GSPQKGYEKF EFNIEDCRKF 720
IDFYKQSIISK HPEWKDFGFF FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSJVNQ 780
GKLYLQFQIYN KDFSAYSKGR PNLLHTLYWKA LFDERNLQDV VYKLNGEAEL FYRKQSPK 840
ITHPAKEAIA NKNKDNPKKE SVFVEYDLKD KRPTEDKPFH HCPITINFKS SGANKFNDI 900
NLLLKEKAND VHILSSTDRE RHLAYLTLD GKGNIIKQDT FNIIGNDRMK TNYHDKLAAI 960
EKDRDSARK WKKINNKKM KEGYLSQVVF EIAKLVIEYN AIVVFPEDLNA GFKRGFRKVE 1020
KQVYQKLEKM LIEKLNLVLF KDNNEFDKTGG VLRAYQLTAP FETFKMGKQ TGIIYYVPAG 1080
FTSKICPVTG FVNQLYPKYF SVSKSQUEFFS KFDKICYNLD KGYFEFSDY KNFGDKAAKG 1140
KWTIASFGSR LINFRNSDKN HNWDTREVYP TKELEKLLKD YSIEYGHGEC IKAACGESD 1200
KKPFAKLTSV LNTILQMRNS KTGTEDLYLI SPVADVNCGF FDSRQAPKNM PQDADANGAY 1260
HIGLKGLMLL GRIKNNQEGK KLNVLVIKNEE YFEFVQNRNMN 1300

```

```

SEQ ID NO: 562      moltype = DNA length = 65
FEATURE           Location/Qualifiers
misc_feature      1..65
note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..65
note = source = /note="SGN001101 genetic locus"
1..65
mol_type = other DNA
organism = synthetic construct

```

```

SEQUENCE: 562
accaaagatg atatttctt taatggtgcc aggataatc caggaaaact gagaacagaa 60
tgaaa 65

```

```

SEQ ID NO: 563      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25

```

-continued

---

```

note = source = /note="Description of Artificial Sequence:
Synthetic oligonucleotide"
1..25
note = source = /note="SGN001101 target sequence"
1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 563
atattttctt taatggtgcc aggca                                25

SEQ ID NO: 564      moltype = RNA length = 130
FEATURE
misc_feature
1..130
note = source = /note="Description of Artificial Sequence:
Synthetic polynucleotide"
1..130
note = source = /note="SGN001101 sgRNA sequence"
1..130
mol_type = other RNA
organism = synthetic construct

SEQUENCE: 564
atattttctt taatggtgcc aggcagtcat agttccattta aagccaaaag tggcttgat 60
gtttctatga taagggttgc gaccctgtggc gtccgggatc gctgcocat tgaaatggc 120
ttctccccat                                         130

```

---

That which is claimed:

1. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein the deaminase polypeptide is encoded by a nucleotide sequence that:
  - a) has at least 90% sequence identity to SEQ ID NO: 449, and
  - b) encodes an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405.
2. The nucleic acid molecule of claim 1, wherein said nucleic acid molecule further comprises a heterologous promoter operably linked to said polynucleotide.
3. The nucleic acid molecule of claim 1, wherein said nucleotide sequence encoding said deaminase polypeptide
  - a) has at least 90% sequence identity to SEQ ID NO: 449, and
  - b) encodes an amino acid sequence having at least 95% sequence identity to SEQ ID NO: 405.
4. The nucleic acid molecule of claim 1, wherein said nucleotide sequence encoding said deaminase polypeptide
  - a) has at least 95% sequence identity to SEQ ID NO: 449, and
  - b) encodes an amino acid sequence having at least 90% sequence identity to SEQ ID NO: 405.
5. A vector comprising the nucleic acid molecule of claim 1.
6. The vector of claim 5, further comprising at least one nucleotide sequence encoding a guide RNA capable of hybridizing to a target nucleic acid.
7. A nucleic acid molecule comprising a polynucleotide encoding a deaminase polypeptide, wherein said deaminase polypeptide is encoded by a nucleotide sequence that
  - a) has the sequence of SEQ ID NO: 449, or
  - b) encodes the amino acid sequence of SEQ ID NO: 405.
8. A vector comprising the nucleic acid molecule of claim 7.
9. The vector of claim 8, further comprising at least one nucleotide sequence encoding a guide RNA capable of hybridizing to a target nucleic acid.

- 10.** A nucleic acid molecule comprising a polynucleotide encoding a fusion protein,  
wherein said fusion protein comprises:  
a Type II CRISPR-Cas protein nickase and a deaminase,  
wherein the deaminase comprises an amino acid sequence  
having at least 90% sequence identity to the amino acid  
sequence of SEQ ID NO: 405,  
and wherein said nickase
  - (a) is a Cas9 nickase; or
  - (b) comprises an amino acid sequence having at least  
95% sequence identity to any one of SEQ ID NO: 42,  
52, 53, 55-59, 61, 397, and 398.
- 11.** The nucleic acid molecule of claim 10, wherein said deaminase comprises an amino acid sequence having at least 95% sequence identity to the amino acid sequence of SEQ ID NO: 405.
- 12.** The nucleic acid molecule of claim 10, wherein said deaminase comprises the amino acid sequence of SEQ ID NO: 405.
- 13.** The nucleic acid molecule of claim 10, wherein the nickase has the amino acid sequence of any one of SEQ ID NO: 42, 52, 53, 55-59, 61, 397, and 398.
- 14.** The nucleic acid molecule of claim 10, wherein the polynucleotide encoding the fusion protein is operably linked at its 5' end to a promoter.
- 15.** The nucleic acid molecule of claim 10, wherein the polynucleotide encoding the fusion protein is operably linked at its 3' end to a terminator.
- 16.** The nucleic acid molecule of claim 10, wherein the fusion protein comprises one or more nuclear localization signals.
- 17.** The nucleic acid molecule of claim 10, wherein the polynucleotide encoding the fusion protein is an mRNA.
- 18.** The nucleic acid molecule of claim 10, wherein said fusion protein comprises the amino acid sequence of SEQ ID NO: 494.
- 19.** The nucleic acid molecule of claim 10, wherein the polynucleotide encoding the fusion protein is codon optimized for expression in a eukaryotic cell.

**20.** A system for modifying a target DNA molecule, said system comprising:

- a) the nucleic acid molecule of claim **10**; and
- b) one or more guide RNAs (gRNAs) capable of hybridizing to said target DNA molecule or one or more nucleic acids encoding the one or more gRNAs.

**21.** The system of claim **20**, wherein the one or more gRNAs are capable of forming a complex with the fusion protein in order to direct said fusion protein to bind to and modify the target DNA molecule.

**22.** The system of claim **20**, wherein at least one of said one or more nucleic acids encoding the one or more gRNAs is operably linked to a promoter.

**23.** The system of claim **20**, wherein the system comprises a vector comprising the nucleic acid molecule of claim **10** and the one or more nucleic acids encoding the one or more gRNAs.

**24.** The system of claim **20**, wherein the target DNA molecule is within a cell.

**25.** The system of claim **24**, wherein the cell is a eukaryotic cell.

**26.** A method for modifying a target DNA molecule comprising a target DNA sequence, said method comprising delivering a system according to claim **20** to said target DNA molecule or a cell comprising the target DNA molecule.

**27.** The method of claim **26**, wherein said modified target DNA molecule comprises an A>N mutation of at least one

nucleotide within the target DNA molecule, wherein N is C, G, or T or an A>G mutation of at least one nucleotide within the target DNA molecule.

**28.** A method for producing a genetically modified cell with a correction in a causal mutation for a genetically inherited disease, the method comprising introducing into the cell:

- a) the nucleic acid molecule of claim **10**, wherein said polynucleotide encoding the fusion protein is operably linked to a promoter to enable expression of the fusion protein in the cell; and
- b) one or more guide RNAs (gRNAs) capable of hybridizing to a target DNA sequence, or one or more nucleic acids encoding the one or more gRNAs, wherein said one or more nucleic acids encoding the one or more gRNAs is operably linked to a promoter to enable expression of the one or more gRNAs in the cell; whereby the fusion protein and one or more gRNAs target to the genomic location of the causal mutation and modify the genomic sequence to correct the causal mutation.

**29.** The method of claim **28**, wherein the correction of the causal mutation comprises introducing an A>G mutation of at least one nucleotide within the target DNA sequence.

**30.** The method of claim **28**, wherein the correction of the causal mutation comprises correcting a nonsense mutation.

**31.** The method of claim **28**, wherein the genetically inherited disease is cystic fibrosis.

\* \* \* \* \*