



(51) International Patent Classification:

Not classified

(21) International Application Number:

PCT/US2019/039325

(22) International Filing Date:

26 June 2019 (26.06.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/690,903	27 June 2018 (27.06.2018)	US
62/716,229	08 August 2018 (08.08.2018)	US

(71) Applicant: **ALTIUS INSTITUTE FOR BIOMEDICAL SCIENCES** [US/US]; 2211 Elliot avenue 4th Floor, Seattle, Washington 98121 (US).

(72) Inventors; and

(71) Applicants: **STAMATOYANNOPOULOS, John A.** [US/US]; 2211 Elliott Avenue, 4th Floor, Seattle, Washington 98121 (US). **URNOV, Fyodor** [US/US]; 2211 Elliott Avenue, 4th Floor, Seattle, Washington 98121 (US).

PEARL, Jocelynn [US/US]; 2211 Elliott Avenue, Suite 410, Seattle, Washington 98121 (US). **THOMAS, Sean** [US/US]; 2211 Elliott Avenue, Suite 410, Seattle, Washington 98121 (US).

(74) Agent: **CHANDRA, Shweta**; Bozicevic, Field & Francis LLP, 201 Redwood Shores Pkwy., Suite 200, Redwood City, California 94065 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: NUCLEASES FOR GENOME EDITING

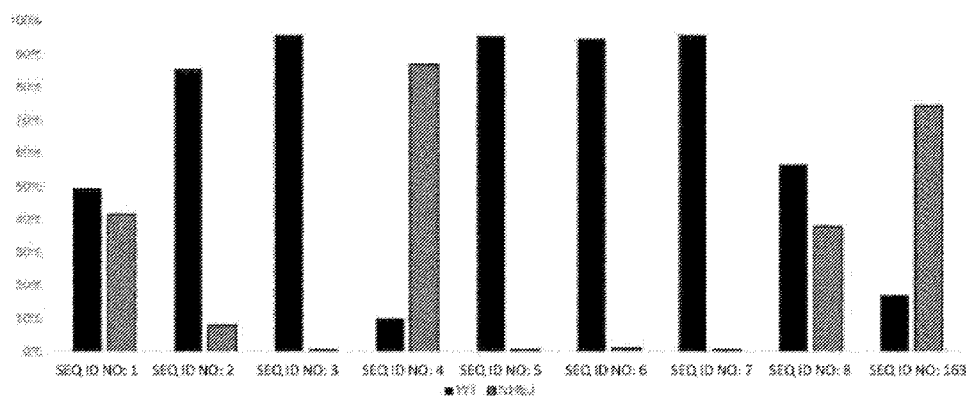


FIG. 1

(57) Abstract: DNA cleaving enzymes are disclosed. The DNA cleaving enzymes are fused to a heterologous DNA binding domain that is designed to bind to a target nucleic acid sequence. Nucleic acids and expression cassettes encoding the DNA cleaving enzymes are also provided. Methods for genome editing using the DNA cleaving enzymes, fusion proteins, and compositions thereof are disclosed herein.



UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

NUCLEASES FOR GENOME EDITING

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority pursuant to 35 U.S.C. §119(e) to U.S. Provisional Application No. 62/690,903, filed June 27, 2018 and U.S. Provisional Application No. 62/716,229, filed August 8, 2018, the disclosures of which are incorporated herein by reference in their entirety.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING PROVIDED AS A TEXT FILE

[0002] A Sequence Listing is provided herewith as a text file, "ALTI-722WO Seq List_ST25.txt," created on June 26, 2019 and having a size of 423 KB. The contents of the text file are incorporated by reference herein in their entirety.

INTRODUCTION

[0003] Type II restriction endonucleases recognize specific DNA sequences, generally four to eight base pairs long, and cut both strands at fixed positions within or close to the recognition site. One of the best known type II restriction endonucleases, FokI, consists of a DNA recognition domain and a non-specific DNA cleavage domain. FokI cleaves DNA within a specified spacer region upon formation of a transient homodimer. The non-specific cleavage domain of FokI has been combined with a variety of DNA-binding domains of other molecules for genome editing purposes, including zinc finger proteins and transcription activator-like effector proteins.

[0004] There still remains a need for cleavage enzymes, which when coupled to DNA binding domains for genome editing, exhibit high cleavage efficiency, small size, and capability to cleave over long spacer regions.

SUMMARY

[0005] A non-naturally occurring fusion protein comprising a nucleic acid binding domain and a cleavage domain, wherein the cleavage domain comprises at least 33.3% divergence from SEQ ID NO: 163 and is immunologically orthogonal to SEQ ID NO: 163 is provided and the nucleic acid binding domain is heterologous to the cleavage domain. In certain aspects, the nucleic acid binding domain binds to a target nucleic acid.

[0006] In some embodiments, the non-naturally occurring fusion protein as described herein comprises one or more of the following characteristics: a) induces greater than 1% indels at a target site; b) the cleavage domain comprises a molecular weight of less than 23 kDa; c) the cleavage domain comprises less than 196 amino acids; and d) capable of cleaving across a spacer region

greater than 24 base pairs. In some embodiments, the non-naturally occurring fusion protein induces greater than 5%, greater than 10%, greater than 20%, greater than 30%, greater than 40%, greater than 50%, greater than 60%, greater than 70%, greater than 80%, or greater than 90% indels at the target site. In some embodiments, the cleavage domain comprises at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% divergence from SEQ ID NO: 163.

[0007] In some embodiments, the cleavage domain comprises a sequence selected from SEQ ID NO: 311 – SEQ ID NO: 314. In some embodiments, the cleavage domain comprises an amino acid sequence having at least 80% or at least 85% sequence identity to the amino acid sequence set forth in one of SEQ ID NO: 1 – SEQ ID NO: 81. In some embodiments, the cleavage domain comprises an amino acid sequence having at least 90% sequence identity to the amino acid sequence set forth in one of SEQ ID NO: 1 – SEQ ID NO: 81. In some embodiments, the cleavage domain comprises an amino acid sequence having at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or a 100% sequence identity to the amino acid sequence set forth in one of SEQ ID NO: 1 – SEQ ID NO: 81.

[0008] Also provided herein are nucleic acid sequences encoding the cleavage domains. In some embodiments, the nucleic acid sequence comprises a sequence having at least 80%, 90%, 95%, 99% or more sequence identity to the nucleic acid sequence set forth in one of SEQ ID NO: 82 – SEQ ID NO: 162. In some embodiments, the nucleic acid sequence is selected from SEQ ID NO: 82 – SEQ ID NO: 162.

[0009] In some embodiments, the nucleic acid binding domain comprises a DNA binding domain (DBD). In some aspects, the DBD binds a first region of genomic DNA. In some embodiments, the cleavage domain fused to the DBD cleaves at a second region of double stranded genomic DNA. In some embodiments, the second region of double stranded genomic DNA is within at most 50 bp of the first region of double stranded genomic DNA. In some embodiments, the second region of double stranded genomic DNA is within at most 15 bp of the first region of double stranded genomic DNA.

[0010] In some embodiments, the nucleic binding domain of the fusion protein comprises a plurality of repeat units. In some embodiments, at least one repeat unit comprises a sequence of $A_{1-11}X_1X_2B_{14-35}$ (SEQ ID NO: 443), wherein each amino acid residue of A_{1-11} comprises any amino acid residue; wherein X_1X_2 comprises base recognition sequence that mediates binding to a nucleotide; wherein each amino acid residue of B_{14-35} comprises any amino acid.

[0011] In some embodiments, the nucleic binding domain of the fusion protein comprises a modular nucleic acid binding domain comprising a potency for a target site greater than 65% and a specificity ratio for the target site of 50:1; and a cleavage domain; wherein the modular nucleic acid binding domain comprises a plurality of repeat units, wherein at least one repeat unit of the plurality comprises a binding region configured to bind to a target nucleic acid base in the target site, wherein the potency comprises indel percentage at the target site, and wherein the specificity ratio comprises indel percentage at the target site over indel percentage at a top-ranked off-target site of the non-naturally occurring fusion protein.

[0012] In some embodiments, the nucleic binding domain comprises a sequence from a zinc finger protein (ZFP). In some embodiments, the cleavage domain is fused to a catalytically inactive Cas9 (dCas9). In some embodiments, the nucleic acid binding domain comprises a guide RNA or a truncated guide RNA.

[0013] In some embodiments, the indel percentage is measured by deep sequencing.

[0014] In some embodiments, the modular nucleic acid binding domain further comprises one or more properties selected from the following: a) binds the target site, wherein the target site comprises a 5' guanine; b) comprises from 7 repeat units to 25 repeat units; and c) upon binding to the target site, the modular nucleic acid binding domain is separated from a second modular nucleic acid binding domain bound to a second target site by from 2 to 50 base pairs. In some embodiments, the plurality of repeat units comprises a *Ralstonia* repeat unit, a *Xanthomonas* repeat unit, a *Legionella* repeat unit, or any combination thereof. In some embodiments, *Ralstonia* repeat unit is a *Ralstonia solanacearum* repeat unit, the *Xanthomonas* repeat unit is a *Xanthomonas spp.* repeat unit, and the *Legionella* repeat unit is a *Legionella quateirensis* repeat unit.

[0015] In some embodiments, the B₁₄₋₃₅ of at least one repeat unit of the plurality of repeat units has at least 92% sequence identity to GGKQALEAVRAQLDLRAAPYG (SEQ ID NO: 280). In some embodiments, the X₁X₂ sequence comprises HD binding to cytosine, NG binding to thymidine, NK binding to guanine, SI binding to adenosine, RS binding to adenosine, HN binding to guanine, or NT binds to adenosine. In some embodiments, the at least one repeat unit comprises any one of SEQ ID NO: 267 – SEQ ID NO: 279. In some embodiments, the at least one repeat unit comprises at least 80% sequence identity with any one of SEQ ID NO: 168 – SEQ ID NO: 263. In some embodiments, the at least one repeat unit comprises at least 80% sequence identity with SEQ ID NO: 209, SEQ ID NO: 197, SEQ ID NO: 233, SEQ ID NO: 253, SEQ ID NO: 203, or SEQ ID NO: 218. In some embodiments, the at least one repeat unit comprises any one of SEQ ID NO: 168 – SEQ ID NO: 263.

In some embodiments, the at least one repeat unit comprises SEQ ID NO: 209, SEQ ID NO: 197, SEQ ID NO: 233, SEQ ID NO: 253, SEQ ID NO: 203, or SEQ ID NO: 218.

[0016] In some embodiments, the target nucleic acid base is cytosine, guanine, thymidine, adenosine, uracil, or a combination thereof.

[0017] In some embodiments, the modular nucleic acid binding domain comprises an N-terminus amino acid sequence, a C-terminus amino acid sequence, or a combination thereof. In some embodiments, the N-terminus amino acid sequence is from *Xanthomonas spp.*, *Legionella quateirensis*, or *Ralstonia solanacearum*. In some embodiments, the N-terminus amino acid sequence comprises at least 80% sequence identity to SEQ ID NO: 264, SEQ ID NO: 300, SEQ ID NO: 331, SEQ ID NO: 303, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 315, SEQ ID NO: 316, or SEQ ID NO: 317. In some embodiments, the N-terminus amino acid sequence comprises SEQ ID NO: 264, SEQ ID NO: 300, SEQ ID NO: 331, SEQ ID NO: 303, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 315, SEQ ID NO: 316, or SEQ ID NO: 317.

[0018] In some embodiments, the C-terminus amino acid sequence is from *Xanthomonas spp.*, *Legionella quateirensis*, or *Ralstonia solanacearum*. In some embodiments, the C-terminus amino acid sequence comprises at least 80% sequence identity to SEQ ID NO: 266, SEQ ID NO: 298, or SEQ ID NO: 306. In some embodiments, the C-terminus amino acid sequence comprises SEQ ID NO: 266, SEQ ID NO: 298, or SEQ ID NO: 306. In some embodiments, the C-terminus amino acid sequence serves as a linker between the modular nucleic acid binding domain and the cleavage domain.

[0019] In some embodiments, the modular nucleic acid binding domain comprises a half repeat. In some embodiments, the half repeat comprises at least 80% sequence identity to one of SEQ ID NO: 265, SEQ ID NO: 322 – SEQ ID NO: 329, or SEQ ID NO: 290.

[0020] In some embodiments, the at least one repeat unit comprises 1-20 additional amino acid residues at the C-terminus. In some embodiments, the at least one repeat unit of the plurality of repeat units is separated from a neighboring repeat unit by a linker. In some embodiments, the linker comprises a recognition site. In some embodiments, the recognition site is for a small molecule, a protease, or a kinase. In some embodiments, the recognition site serves as a localization signal, e.g., nuclear localization signal. In some embodiments, the plurality of repeat units comprises 3 to 60 repeat units.

[0021] In some embodiments, the target site is a nucleic acid sequence within a PDCD1 gene, a CTLA4 gene, a LAG3 gene, a TET2 gene, a BTLA gene, a HAVCR2 gene, a CCR5 gene, a CXCR4 gene, a TRA gene, a TRB gene, a B2M gene, an albumin gene, a HBB gene, a HBA1 gene, a TTR

gene, a NR3C1 gene, a CD52 gene, an erythroid specific enhancer of the BCL11A gene, a CBLB gene, a TGFBR1 gene, a SERPINA1 gene, a HBV genomic DNA in infected cells, a CEP290 gene, a DMD gene, a CFTR gene, or an IL2RG gene. In some embodiments, a nucleic acid sequence encoding a chimeric antigen receptor (CAR), alpha-L iduronidase (IDUA), iduronate-2-sulfatase (IDS), or Factor 9 (F9), is inserted at the target site.

[0022] In some embodiments, the nucleic acid binding domain and the cleavage domain are linked by a synthetic linker comprising 0-15 residues of glycine, methionine, aspartic acid, alanine, lysine, serine, leucine, threonine, tryptophan, or any combination thereof.

[0023] An expression cassette comprising a nucleic acid sequence encoding for a sequence selected from SEQ ID NO: 1 – SEQ ID NO: 81 is disclosed. In some embodiments, the nucleic acid sequence is selected from SEQ ID NO: 82 – SEQ ID NO: 162.

[0024] In some aspects, an expression cassette comprises a nucleic acid sequence encoding the cleavage domain set forth here is provided.

[0025] In some embodiments, the cleavage domain comprises at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% divergence from SEQ ID NO: 163. In some embodiments, the cleavage domain comprises a sequence selected from SEQ ID NO: 311 – SEQ ID NO: 314. In some embodiments, the nucleic acid sequence has at least 80% sequence identity with SEQ ID NO: 82 – SEQ ID NO: 162. In some embodiments, the nucleic acid sequence is selected from SEQ ID NO: 82 – SEQ ID NO: 162. In some embodiments, the nucleic acid binding domain binds a first region of double stranded genomic DNA. In some embodiments, the cleavage domain cleaves at a second region of double stranded genomic DNA. In some embodiments, the second region of double stranded genomic DNA is within at most 50 bp of the first region of double stranded genomic DNA. In some embodiments, the second region of double stranded genomic DNA is within at most 15 bp of the first region of double stranded genomic DNA.

[0026] In some embodiments, the expression cassette is a part of a viral vector.

[0027] In some aspects, a repeat unit of the plurality of repeat units recognizes a target nucleic acid base and wherein the plurality of repeat units has one or more of the following characteristics: (a) at least one repeat unit comprising greater than 39 amino acid residues; (b) at least one repeat unit comprising greater than 35 amino acid residues derived from the genus of *Ralstonia*; (c) at least one repeat unit comprising less than 32 amino acid residues; and (d) each repeat unit of the plurality of repeat units is separated from a neighboring repeat unit by a linker comprising a recognition site. In some aspects, the at least one repeat unit comprises an amino acid selected from glycine, alanine,

threonine or histidine at a position after an amino acid residue at position 35. In some aspects, the at least one repeat unit comprises an amino acid selected from glycine, alanine, threonine or histidine at a position after an amino acid residue at position 39.

[0028] In some aspects, a method of genome editing comprises: administering the non-naturally occurring fusion protein of any embodiment described herein or the expression cassette of any embodiment described herein; and inducing a double stranded break.

[0029] In some aspects, a method of genome editing comprises: administering a first non-naturally occurring fusion protein and a second non-naturally occurring fusion protein; dimerizing a cleavage domain of the first non-naturally occurring fusion protein and a cleavage domain of the second non-naturally occurring fusion protein; and cleaving a double stranded genomic DNA at a target site, wherein the first non-naturally occurring fusion protein comprises a first nucleic acid binding domain and a first cleavage domain and the second non-naturally occurring fusion protein comprises a second nucleic binding domain and a second cleavage domain and wherein the first cleavage domain and the second cleavage domain comprise at least 33.3% divergence from SEQ ID NO: 163 and are immunologically orthogonal to SEQ ID NO: 163.

[0030] In some embodiments, the first nucleic acid binding domain binds to a first region of a top strand of the double stranded genomic DNA at the target site and the first cleavage domain cleaves at a second region of the top strand of the double stranded genomic DNA at the target site. In some embodiments, the second nucleic acid binding domain recognizes a first region of a bottom strand of the double stranded genomic DNA at the target site and the second cleavage domain cleaves at a second region of the bottom strand of the double stranded genomic DNA at the target site. In some embodiments, the cleaving the double stranded genomic DNA at the target site comprises a double strand break. In some embodiments, the method further comprises introducing an ectopic nucleic acid encoding a gene at the double strand break. In some embodiments, the ectopic nucleic acid encoding the gene comprises a chimeric antigen receptor (CAR), alpha-L iduronidase (IDUA), iduronate-2-sulfatase (IDS), or Factor 9 (F9). In some embodiments, the cleaving the double stranded genomic DNA at the target site partially or completely knocks out a target gene. In some embodiments, the target gene is a nucleic acid sequence within a PDCD1 gene, a CTLA4 gene, a LAG3 gene, a TET2 gene, a BTLA gene, a HAVCR2 gene, a CCR5 gene, a CXCR4 gene, a TRA gene, a TRB gene, a B2M gene, an albumin gene, a HBB gene, a HBA1 gene, a TTR gene, a NR3C1 gene, a CD52 gene, an erythroid specific enhancer of the BCL11A gene, a CBLB gene, a TGFBR1 gene, a SERPINA1 gene, a HBV genomic DNA in infected cells, a CEP290 gene, a DMD gene, a CFTR gene, or an IL2RG gene.

[0031] In some embodiments, the administering comprises direct administration to a subject in need thereof. In some embodiments, the administering comprises transfecting a cell ex vivo with the first non-naturally occurring fusion protein and the second non-naturally occurring fusion protein, thereby obtaining a cell comprising modified DNA.

[0032] In some embodiments, the cell comprising modified DNA is administered to a subject in need thereof. In some embodiments, the administering comprises intravenous, subcutaneous, intramuscular, or mucosal administration to a subject in need thereof.

[0033] In some embodiments, the target site is in a cell. In some embodiments, the cell comprises a T-cell, a hematopoietic stem cell (HPSC), or a liver cell. In some embodiments, the subject is a human. In some embodiments, the human has cancer, blood cancer, a CD19 malignancy, a BCMA malignancy, transthyretin amyloidosis, HIV, glioblastoma multiforme, acute lymphoblastic leukemia, acute myeloid leukemia, b-thalassemia, sickle cell disease, MPSI, MPSII, hemophilia B, multiple myeloma, melanoma, sarcoma, Leber congenital amaurosis type 10 (LCA10), Duchenne muscular dystrophy, cystic fibrosis, alpha-1 antitrypsin deficiency (dA1AT def), X-linked severe combined immunodeficiency (X-SCID), or Hepatitis B.

BRIEF DESCRIPTION OF THE DRAWINGS

[0034] **FIG. 1** illustrates the percentage of indels (insertions/deletions), produced via non-homologous end joining (NHEJ), achieved with fusion proteins of DNA binding domains and eight different endonucleases having amino acid sequences of SEQ ID NO: 1 – SEQ ID NO: 8 (corresponding nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 89) as compared to fusion proteins of DNA binding domains and a FokI nuclease (amino acid sequence of SEQ ID NO: 163, nucleic acid sequence of SEQ ID NO: 164).

[0035] **FIG. 2** illustrates the base pair cleavage rate at a human SMARCA4 target site edited with a pair of fusion proteins comprising a TALE DNA binding domain and a FokI endonuclease (amino acid sequence of SEQ ID NO: 163, nucleic acid sequence of SEQ ID NO: 164).

[0036] **FIG. 3** illustrates the base pair cleavage rate at a human SMARCA4 target site edited with a pair of fusion proteins comprising a TALE DNA binding domain and an amino acid sequence of SEQ ID NO: 1 (nucleic acid sequence of SEQ ID NO: 82).

[0037] **FIG. 4** illustrates the base pair cleavage rate at a human SMARCA4 target site edited with a pair of fusion proteins comprising a TALE DNA binding domain and an amino acid sequence of SEQ ID NO: 2 (nucleic acid sequence of SEQ ID NO: 83).

[0038] FIG. 5 illustrates the base pair cleavage rate at a human SMARCA4 target site edited with a pair of fusion proteins comprising a TALE DNA binding domain and an amino acid sequence of SEQ ID NO: 4 (nucleic acid sequence of SEQ ID NO: 85).

[0039] FIG. 6 illustrates the base pair cleavage rate at a human SMARCA4 target site edited with a pair of fusion proteins comprising a TALE DNA binding domain and an amino acid sequence of SEQ ID NO: 8 (nucleic acid sequence of SEQ ID NO: 89).

[0040] FIG. 7 illustrates the base pair cleavage rate at a control AAVS1 target site. This positive control targets an intronic region of the AAVS1/PPP1R12C locus.

[0041] FIGS. 8A-8D provide sequence alignments of nucleases, namely SEQ ID NOs:65; 74; 1; 2; 3; 16; 6; 4; 8; 56; 58; 59; 48; 17; 22; 49; 51; 53; 52; 47; 52; 54. Residues that mediate catalytic function of the nucleases and residues that mediate dimerization of the nuclease are indicated in Fig. 8A.

DETAILED DESCRIPTION

[0042] The present disclosure provides cleavage domains for genome editing and methods of using the same for therapeutic purposes. In some aspects, a cleavage domain can be combined with a DNA-binding domain to allow for greater precision and efficacy in genome editing. DNA-binding domains can be derived from transcription activator-like effector (TALE) systems, *Ralstonia*-derived proteins, *Legionella*-derived proteins, zinc finger proteins (ZFPs), or guide RNAs, or truncated guide RNAs (gRNAs, tru-gRNAs) that can subsequently be fused to fusions of catalytically inactive Cas9 (dCas9) and FokI. Also described herein are genome editing techniques using DNA-binding domains fused to the described cleavage domains.

[0043] In some embodiments, “derived” indicates that a protein is from a particular source (e.g., *Ralstonia*), is a variant of a protein from a particular source (e.g., *Ralstonia*), is a mutated or modified form of the protein from a particular source (e.g., *Ralstonia*), and shares at least 30% sequence identity with, at least 40% sequence identity with, at least 50% sequence identity with, at least 60% sequence identity with, at least 70% sequence identity with, at least 80% sequence identity with, or at least 90% sequence identity with a protein from a particular source (e.g., *Ralstonia*).

[0044] In some embodiments, “modular” indicates that a particular composition such as a nucleic acid binding domain, comprises a plurality of repeat units that can be switched and replaced with other repeat units. For example, any repeat unit in a modular nucleic acid binding domain can be switched with a different repeat unit. In some embodiments, modularity of the nucleic acid binding domains disclosed herein allows for switching the target nucleic acid base for a particular repeat unit

by simply switching it out for another repeat unit. In some embodiments, modularity of the nucleic acid binding domains disclosed herein allows for swapping out a particular repeat unit for another repeat unit to increase the affinity of the repeat unit for a particular target nucleic acid. Overall, the modular nature of the nucleic acid binding domains disclosed herein enables the development of genome editing complexes that can precisely target any nucleic acid sequence of interest.

[0001] The term “heterologous” refers to two components that are defined by structures derived from different sources. For example, in the context of a polypeptide, a “heterologous” polypeptide may include operably linked amino acid sequences that are derived from different polypeptides (e.g., a NBD and a functional domain derived from different sources). Similarly, in the context of a polynucleotide encoding a chimeric polypeptide, a “heterologous” polynucleotide may include operably linked nucleic acid sequences that can be derived from different genes. Other exemplary “heterologous” nucleic acids include expression constructs in which a nucleic acid comprising a coding sequence is operably linked to a regulatory element (e.g., a promoter) that is from a genetic origin different from that of the coding sequence (e.g., to provide for expression in a host cell of interest, which may be of different genetic origin than the promoter, the coding sequence or both). In the context of recombinant cells, “heterologous” can refer to the presence of a nucleic acid (or gene product, such as a polypeptide) that is of a different genetic origin than the host cell in which it is present.

[0002] The term “operably linked” refers to linkage between molecules to provide a desired function. For example, “operably linked” in the context of nucleic acids refers to a functional linkage between nucleic acid sequences. By way of example, a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) may be operably linked to a second polynucleotide, wherein the expression control sequence affects transcription and/or translation of the second polynucleotide. In the context of a polypeptide, “operably linked” refers to a functional linkage between amino acid sequences (e.g., different domains) to provide for a described activity of the polypeptide.

[0003] As used herein, the term “cleavage” refers to the breakage of the covalent backbone of a nucleic acid, e.g., a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, the polypeptides provided herein are used for targeted double-stranded DNA cleavage.

[0004] A “cleavage half-domain” is a polypeptide sequence which, in conjunction with a second polypeptide (either identical or different) forms a complex having cleavage activity (preferably double-strand cleavage activity).

[0005] A “target nucleic acid,” “target sequence,” or “target site” is a nucleic acid sequence that defines a portion of a nucleic acid to which a binding molecule, such as, the NBD disclosed herein will bind. The target nucleic acid may be present in an isolated form or inside a cell. A target nucleic acid may be present in a region of interest. A “region of interest” may be any region of cellular chromatin, such as, for example, a gene or a non-coding sequence within or adjacent to a gene, in which it is desirable to bind an exogenous molecule. Binding can be for the purposes of targeted DNA cleavage and/or targeted recombination, targeted activation or repression. A region of interest can be present in a chromosome, an episome, an organellar genome (e.g., mitochondrial, chloroplast), or an infecting viral genome, for example. A region of interest can be within the coding region of a gene, within transcribed non-coding regions such as, for example, promoter sequences, leader sequences, trailer sequences or introns, or within non-transcribed regions, either upstream or downstream of the coding region. A region of interest can be as small as a single nucleotide pair or up to 2,000 nucleotide pairs in length, or any integral value of nucleotide pairs.

[0006] An “exogenous” molecule is a molecule that is not normally present in a cell but can be introduced into a cell by one or more genetic, biochemical or other methods. An exogenous molecule can comprise, for example, a functioning version of a malfunctioning endogenous molecule, e.g. a gene or a gene segment lacking a mutation present in the endogenous gene. An exogenous nucleic acid can be present in an infecting viral genome, a plasmid or episome introduced into a cell. Methods for the introduction of exogenous molecules into cells are known to those of skill in the art and include, but are not limited to, lipid-mediated transfer (i.e., liposomes, including neutral and cationic lipids), electroporation, direct injection, cell fusion, particle bombardment, calcium phosphate co-precipitation, DEAE-dextran-mediated transfer and viral vector-mediated transfer.

[0007] By contrast, an “endogenous” molecule is one that is normally present in a particular cell at a particular developmental stage under particular environmental conditions. For example, an endogenous nucleic acid can comprise a chromosome, the genome of a mitochondrion, chloroplast or other organelle, or a naturally-occurring episomal nucleic acid. Additional endogenous molecules can include proteins, for example, transcription factors and enzymes.

[0008] A “gene,” for the purposes of the present disclosure, includes a DNA region encoding a gene product, as well as all DNA regions which regulate the production of the gene product,

whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control region.

[0009] “Gene expression” refers to the conversion of the information, contained in a gene, into a gene product. A gene product can be the direct transcriptional product of a gene (e.g., mRNA, tRNA, rRNA, antisense RNA, ribozyme, structural RNA, shRNA, RNAi, miRNA or any other type of RNA) or a protein produced by translation of a mRNA. Gene products also include RNAs which are modified, by processes such as capping, polyadenylation, methylation, and editing, and proteins modified by, for example, methylation, acetylation, phosphorylation, ubiquitination, ADP-ribosylation, myristylation, and glycosylation.

Nucleases for Genome Editing

[0010] Genome editing can include the process of modifying a DNA of a cell in order to introduce or knock out a target gene or a target gene region. In some instances, a subject may have a disease in which a protein is aberrantly expressed or completely lacking. One therapeutic strategy for treating this disease can be introduction of a target gene or a target gene region to correct the aberrant or missing protein. For example, genome editing can be used to modify the DNA of a cell in the subject in order to introduce a functional gene, which gives rise to a functional protein. Introduction of this functional gene and expression of the functional protein can relieve the disease state of the subject.

[0011] In other instances, a subject may have a disease in which protein is overexpressed or is targeted by a virus for infection of a cell. Alternatively, a therapy such as a cell therapy for cancer can be ineffective due to repression of certain processes by tumor cells (e.g., checkpoint inhibition). Still alternatively, it may be desirable to eliminate a particular protein expressed at the surface of a cell in order to generate a universal, off-the-shelf cell therapy for a subject in need thereof (e.g., TCR). In such cases, it can be desirable to partially or completely knock out the gene encoding for such a protein. Genome editing can be used to modify the DNA of a cell in the subject in order to partially or completely knock out the target gene, thus reducing or eliminating expression of the protein of interest.

[0012] Genome editing can include the use of any nuclease as described herein in combination with any DNA binding domain disclosed herein in order to bind to a target gene or target gene region and induce a double strand break, mediated by the nuclease. Genes can be introduced during this process,

or DNA binding domains can be designed to cut at regions of the DNA such that after non-homologous end joining, the target gene or target gene region is removed. Genome editing systems that are further disclosed and described in detail herein can include TALENs (with DNA binding domains derived from *Xanthomonas*), *Ralstonia*-derived modular nucleic acid binding domains (RNBDs) fused to nucleases, *Legionella*-derived modular nucleic acid binding domains (MAP-NBDs) fused to nucleases, ZFNs, or CRISPR-Cas9 systems.

[0013] The specificity and efficiency of genome editing can be dependent on the nuclease responsible for cleavage. More than 3,000 type II restriction endonucleases have been identified. They recognize short, usually palindromic, sequences of 4–8 bp and, in the presence of Mg²⁺, cleave the DNA within or in close proximity to the recognition sequence. Naturally, type II restriction enzymes themselves have a DNA recognition domain that can be separated from the catalytic, or cleavage, domain. As such, since cleavage occurs at a site adjacent to the DNA sequence bound by the recognition domain, these enzymes can be referred to as exhibiting “shifted” cleavage. These type II restriction enzymes having both the recognition domain and the cleavage domain can be 400-600 amino acids. The main criterion for classifying a restriction endonuclease as a type II enzyme is that it cleaves specifically within or close to its recognition site and that it does not require ATP hydrolysis for its nucleolytic activity. An example of a type II restriction endonuclease is FokI, which consists of a DNA recognition domain and a non-specific DNA cleavage domain. FokI cleaves DNA nine and thirteen bases downstream of an asymmetric sequence (recognizing a DNA sequence of GGATG).

[0014] In some embodiments, the DNA cleavage domain at the C-terminus of FokI itself can be combined with a variety of DNA-binding domains (e.g., RNBDs, TALEs, MAP-NBDs) of other molecules for genome editing purposes. This cleavage domain can be 180 amino acids in length and can be directly linked to a DNA binding domain (e.g., RNBDs, TALEs, MAP-NBDs). In some embodiments, the FokI cleavage domain only comprises a single catalytic site. Thus, in order to cleave phosphodiester bonds, these enzymes form transient homodimers, providing two catalytic sites capable of cleaving double stranded DNA. In some embodiments, a single DNA-binding domains (e.g., RNBDs, TALEs, MAP-NBDs) linked to a Type IIS cleaving domain may not nick the double stranded DNA at the targeted site. In some embodiments, cleaving of target DNA only occurs when a pair of DNA-binding domains (e.g., RNBDs, TALEs, MAP-NBDs), each linked to a Type IIS cleaving domain (e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) bind to opposing strands of DNA and allow for formation of a transient homodimer in the spacer region (the base pairs between the C-terminus of the DNA

binding domain on a top strand of DNA and the C-terminus of the DNA binding domain on a bottom strand of DNA). Said spacer region can be greater than 2 base pairs, greater than 5 base pairs, greater than 10 base pairs, greater than 15 base pairs, greater than 24 base pairs, greater than 25 base pairs, greater than 30 base pairs, greater than 35 base pairs, greater than 40 base pairs, greater than 45 base pairs, or greater than 50 base pairs. In some embodiments, the spacer region can be anywhere from 2 to 50 base pairs, 5 to 40 base pairs, 10 to 30 base pairs, 14 to 40 base pairs, 24 to 30 base pairs, 24 to 40 base pairs, or 24 to 50 base pairs. In some embodiments, the nuclease disclosed herein (e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) can be capable of cleaving over a spacer region of greater than 24 base pairs upon formation of a transient homodimer.

[0015] In some instances, endonucleases of the present disclosure can comprise one or more mutations relative to any one of SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162). In some cases, the non-naturally occurring enzymes described herein can comprise about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more, e.g., up to 50 mutations relative to any one of SEQ ID NOs: 1-81. A mutation can be engineered to enhance cleavage efficiency. A mutation can abolish cleavage activity. In some cases, a mutation can enhance homodimerization. For example, FokI can have a mutation at one or more amino acid residue positions 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, and 538 to modulate homodimerization, and similar mutations can be designed based on sequence alignment of the nucleases with FokI sequence.

[0016] **TABLE 1** shows exemplary amino acid sequences (SEQ ID NO: 1 – SEQ ID NO: 81) of cleavage domains for genome editing and the corresponding back-translated nucleic acid sequences (SEQ ID NO: 82 – SEQ ID NO: 162) encoding the cleavage domains, which were obtained using Geneious software and selecting for human codon optimization.

TABLE 1 – Amino Acid Sequences of Endonucleases

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
1	FLVKGAMEIKKSEL RHKL RHVPHEYIELI EIAQDSKQNRLLFEK VVEFFKKIYGYRGK HLGGSRKPDGALFT DGLVLNHGILDTKA YKDG YRLPISQADE MQRYVDENNKR SQ VINPNEWWEIYPTSI TDFKFLFVSGFFQGD	82	TTCCTGGTGAAGGGCGCCATGGAGATCAAGAAGAGCGAGCTGA GGCACAAGCTGAGGCACGTGCCCCACGAGTACATCGAGCTGATC GAGATCGCCCAGGACAGCAAGCAGAACAGGCTGCTGGAGTTCA AGGTGGTGGAGTTCTTCAAGAAGATCTACGGCTACAGGGGCAA GCACCTGGGCGGCAGCAGGAAGCCCCGACGGCGCCCTGTTCACC GACGGCCTGGTGCTGAACACGGCATCATCCTGGACACCAAGGC CTACAAGGACGGCTACAGGCTGCCCATCAGCCAGGCCGACGAG ATGCAGAGGTACGTGGACGAGAACAACAAGAGGAGCCAGGTGA TCAACCCCAACGAGTGGTGGGAGATCTACCCACACGATCACC GACTTCAAGTTCTGTTCGTGAGCGGCTTCTTCCAGGGCGACTAC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	YRKQLERVSHLTKC QGAVMSVEQLLLGG EKIKEGSLTLEEVGK KFKNDEIVF		AGGAAGCAGCTGGAGAGGGTGAGCCACCTGACCAAGTGCCAGG GCGCCGTGATGAGCGTGGAGCAGCTGCTGCTGGGCGGCGAGAA GATCAAGGAGGGCAGCCTGACCCTGGAGGAGGTGGGCAAGAAG TTCAAGAACGACGAGATCGTGTTT
2	QIVKSSIEMSKANM RDNLQMLPHDYIELI EISQDPYQNRIFEMK VMDLFINEYGFSGS HLGGSRKPDGAMY AHGFGVIVDTKAYK DGYNLPISQADEME RYVRENIDRNEHVN SNRWWNIFPEDTNE YKFLFVSGFFKGNFE KQLERISIDTGVQGG ALSVEHLLLGAEYIK RGILTLYDFKNSFLN KEIQF	83	CAGATCGTGAAGAGCAGCATCGAGATGAGCAAGGCCAACATGA GGGACAACCTGCAGATGCTGCCCCACGACTACATCGAGCTGATC GAGATCAGCCAGGACCCCTACCAGAACAGGATCTTCGAGATGA AGGTGATGGACCTGTTCATCAACGAGTACGGCTTCAGCGGCAGC CACCTGGGCGGCAGCAGGAAGCCCCGACGGCGCCATGTACGCC ACGGCTTCGGCGTGATCGTGGACACCAAGGCCCTACAAGGACGG CTACAACCTGCCCATCAGCCAGGCCGACGAGATGGAGAGGTAC GTGAGGGAGAACATCGACAGGAACGAGCACGTGAACAGCAACA GGTGGTGGAAACATCTTCCCCGAGGACACCAACGAGTACAAGTTC CTGTTCTGTGAGCGGCTTCTTCAAGGGCAACTTCGAGAAGCAGCT GGAGAGGATCAGCATCGACACCGGCGTGACGGGCGGCGCCCTG AGCGTGGAGCACCTGCTGCTGGGCGCCGAGTACATCAAGAGGG GCATCCTGACCCTGTACGACTTCAAGAACAGCTTCCTGAACAAG GAGATCCAGTTC
3	QTIKSSIEELKSELRT QLNVISHDYLLQLVDI SQDSQQNRLFEMKV MDLFINEFGYNGSH LGGSRKPDGILYTEG LSKDYGIIVDTKAYK DGYNLPISQADEME RYIRENIDRNEVVNP NRWWEVFPKINDY KFLFVSAYFKGNFK EQLERISINTGILGGA ISVEHLLLGAEYFKR GILSLEDVRDKFCNT EIEF	84	CAGACCATCAAGAGCAGCATCGAGGAGCTGAAGAGCGAGCTGA GGACCCAGCTGAACGTGATCAGCCACGACTACCTGCAGCTGGTG GACATCAGCCAGGACAGCCAGCAGAACAGGCTGTTCGAGATGA AGGTGATGGACCTGTTCATCAACGAGTTCGGCTACAACGGCAGC CACCTGGGCGGCAGCAGGAAGCCCCGACGGCATCCTGTACACCG AGGGCCTGAGCAAGGACTACGGCATCATCGTGGACACCAAGGC CTACAAGGACGGCTACAACCTGCCCATCGCCAGGCCGACGAG ATGGAGAGGTACATCAGGGAGAACATCGACAGGAACGAGGTGG TGAACCCCAACAGGTGGTGGGAGGTGTTCCCCAGCAAGATCAAC GACTACAAGTTCCTGTTCTGTGAGCGCCTACTTCAAGGGCAACT CAAGGAGCAGCTGGAGAGGATCAGCATCAACACCGGCATCCTG GGCGGCGCCATCAGCGTGGAGCACCTGCTGCTGGGCGCCGAGTA CTTCAAGAGGGGCATCCTGAGCCTGGAGGACGTGAGGGACAAG TTCTGCAACACCGAGATCGAGTTC
4	GKSEVETIKEQMRG ELTHLSHEYLGLLDL AYDSKQNRFLFELKT MQLLTEECGFEGHL LGGSRKPDGIVYTK DENEQVGKENYGIH DTKAYSGGYSPLISQ ADEMERYIGENQTR DIRINPNEWWKNFG DGVTEYYLFLVAGH FKGKYQEIDRINC KNIKGAAVSIQQLLR IVNDYKAGKLTHED MKLKIFHY	85	GGCAAGAGCGAGGTGGAGACCATCAAGGAGCAGATGAGGGGCG AGCTGACCCACCTGAGCCACGAGTACCTGGGCCTGCTGGACCTG GCCTACGACAGCAAGCAGAACAGGCTGTTTCGAGCTGAAGACCA TGCAGCTGCTGACCGAGGAGTGCGGCTTCGAGGGCCTGCACCTG GGCGGCAGCAGGAAGCCCCGACGGCATCGTGTACACCAAGGACG AGAACGAGCAGGTGGGCAAGGAGAACTACGGCATCATCATCGA CACCAAGGCCTACAGCGGCGGCTACAGCCTGCCATCAGCCAGG CCGACGAGATGGAGAGGTACATCGGCGAGAACCAGACCAGGGA CATCAGGATCAACCCCAACGAGTGGTGGAGAAGTTCGGCGAC GGCGTGACCGAGTACTACTACCTGTTCTGTGGCCGGCCACTTCAA GGGCAAGTACCAGGAGCAGATCGACAGGATCAACTGCAACAAG AACATCAAGGGCGCCGCCGTGAGCATCCAGCAGCTGCTGAGGA TCGTGAACGACTACAAGGCCGGCAAGCTGACCCACGAGGACAT GAAGCTGAAGATCTTCCACTAC
5	MKILELLINECGYKG LHLGGARKPDGIIYT EKEKYNNGVIIDTK AYSKGYNLPIGQIDE MIRYIIENNERNIKR NTNCWWNNFEKNV NEFYFSFISGEFTGNI EEKLNRIFISTNIKGN AMSVKTLLEYLANEI	86	ATGAAGATCCTGGAGCTGCTGATCAACGAGTGCGGCTACAAGG GCCTGCACCTGGGCGGCGCCAGGAAGCCCCGACGGCATCATCTAC ACCGAGAAGGAGAAGTACAACCTACGGCGTGATCATCGACACCA AGGCCTACAGCAAGGGCTACAACCTGCCATCGGCCAGATCGAC GAGATGATCAGGTACATCATCGAGAACAACGAGAGGAACATCA AGAGGAACACCAACTGCTGGTGGAAACAACCTTCGAGAAGAACGT GAACGAGTTCTACTTCAGCTTCATCAGCGGCGAGTTCACCGGCA ACATCGAGGAGAAGCTGAACAGGATCTTCATCAGCACCAACATC AAGGGCAACGCCATGAGCGTGAAGACCCTGCTGTACCTGGCCA

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	KANRISYIELLNIFYD NKV		ACGAGATCAAGGCCAACAGGATCAGCTACATCGAGCTGCTGAA CTACTTCGACAACAAGGTG
6	AKSSQSETKEKLRE KLRNLPHEYLSLVD LAYDSKQNRLEFEMK VIELLTEECGFQGLH LGGSRPDGVLYTA GLTDNYGIILDTKAY SSGYSPLIAQADEME RYVRENQTRDELVN PNQWWENFENGLG TFYFLFVAGHFNGN VQAQLERISRNTGV LGAAASISQLLLLAD AIRGGRMDRERLRH LMFQNEEFL	87	GCCAAGAGCAGCCAGAGCGAGACCAAGGAGAAGCTGAGGGAG AAGCTGAGGAACCTGCCCCACGAGTACCTGAGCCTGGTGGACCT GGCCTACGACAGCAAGCAGAACAGGCTGTTCGAGATGAAGGTG ATCGAGCTGCTGACCGAGGAGTGCGGCTTCCAGGGCCTGCACCT GGGCGGCAGCAGGAGGCCCCGACGGCGTGCTGTACACCGCCGGC CTGACCGACAACCTACGGCATCATCTGGACACCAAGGCCTACAG CAGCGGCTACAGCCTGCCCATCGCCAGGCCGACGAGATGGAG AGGTACGTGAGGGAGAACCAGACCAGGACGAGCTGGTGAACC CCAACCAGTGGTGGGAGAACTTCGAGAACGGCCTGGGCACCTTC TACTTCCTGTTCGTGGCCGGCCACTTCAACGGCAACGTGCAGGC CCAGCTGGAGAGGATCAGCAGGAACACCGGCGTGCTGGGCGCC GCCGCCAGCATCAGCCAGCTGCTGCTGCTGGCCGACGCCATCAG GGGCGGCAGGATGGACAGGGAGAGGCTGAGGCACCTGATGTTT CAGAACGAGGAGTTCCTG
7	NSEKSEFTQEKDNL REKLDLTSHEYLSLV DLAFDSQQNRLEFEM KTVELLTKECNKYG VHLGGSRKPDGIITY ENSTDNYGVIIDTKA YSNGYNLPISQVDE MVRVVEENNKREK ERNSNEWWKEFGD NINKFYFSFISGKFIG NIEEKLQRITFTNVY GNAMTITLLYLANE IKANRLKTMEVVKY FDNKV	88	AACAGCGAGAAGAGCGAGTTACCCAGGAGAAGGACAACCTGA GGGAGAAGCTGGACACCCTGAGCCACGAGTACCTGAGCCTGGT GGACCTGGCCTTCGACAGCCAGCAGAACAGGCTGTTCGAGATGA AGACCGTGGAGCTGCTGACCAAGGAGTGCAACTACAAGGGCGT GCACCTGGGCGGCAGCAGGAAGCCCCGACGGCATCATCTACACC GAGAACAGCACCAGCAACTACGGCGTGATCATCGACACCAAGG CCTACAGCAACGGCTACAACCTGCCCATCAGCCAGGTGGACGAG ATGGTGAGGTACGTGGAGGAGAACAACAAGAGGGAGAAGGAG AGGAACAGCAACGAGTGGTGGAAAGGAGTTCGGCGACAACATCA ACAAGTTCTACTTCAGCTTCATCAGCGCAAGTTCATCGGCAAC ATCGAGGAGAAGCTGCAGAGGATCACCATCTTCACCAACGTGTA CGGCAACGCCATGACCATCATCACCTGTGTACCTGGCCAACG AGATCAAGGCCAACAGGCTGAAGACCATGGAGGTGGTGAAGTA CTTCGACAACAAGGTG
8	NLTCSDLTEIKEEVR NALTHLSHEYLLALID LAYDSTQNRLEFEMK TLQLLVEECGYQGT HLGGSRKPDGICYSE EAKSEGLEANYGIII DTKSYSGGYGLPISQ ADEMERYIRENQTR DAEVNRNKWWEAF PETIDIFYFMFVAGH FKGNYFNQLERLQR STGIKGAAVDIKTLL LTANRCKTGELDHA GIESCFNNCRLL	89	AACCTGACCTGCAGCGACCTGACCGAGATCAAGGAGGAGGTGA GGAACGCCCTGACCCACCTGAGCCACGAGTACCTGGCCCTGATC GACCTGGCCTACGACAGCACCCAGAACAGGCTGTTCGAGATGA AGACCTGTCAGCTGCTGGTGGAGGAGTGCGGCTACCAGGGCAC CCACCTGGGCGGCAGCAGGAAGCCCCGACGGCATCTGTACAGC GAGGAGGCCAAGAGCGAGGGCCTGGAGGGCAACTACGGCATCA TCATCGACACCAAGAGCTACAGCGGCGGCTACGGCCTGCCCATC AGCCAGGCCGACGAGATGGAGAGGTACATCAGGGAGAACCAGA CCAGGGACGCCGAGGTGAACAGGAACAAGTGGTGGGAGGCCTT CCCCGAGACCATCGACATCTTCTACTTCATGTTCTGTGGCCGCCA CTTCAAGGGCAACTACTTCAACCAGCTGGAGAGGCTGCAGAGG AGCACCGGCATCAAGGGCGCCGCGCTGGACATCAAGACCCTGCT GCTGACCGCCAACAGGTGCAAGACCGGCGAGCTGGACCACGCC GGCATCGAGAGCTGCTTCTTCAACAACCTGCAGGCTG
9	DNVKSNNFQEKDEL REKLDLTSHEYLYL LDLAYDSKQNKLF MKILELLINECGYRG LHLGGVRKPDGIITY EKEKYNNGYVIIDTK AYSKGYNLPIDQIDE MIRYIENNERNIKR NTNCWWNNFEKNV NEFYFSFISGEFTGNI EEKLNRFISTNIKGN	90	GACAACGTGAAGAGCAACTTCAACCAGGAGAAGGACGAGCTGA GGGAGAAGCTGGACACCCTGAGCCACGAGTACCTGTACCTGCTG GACCTGGCCTACGACAGCAAGCAGAACAGTACCTGTTCGAGATGA AGATCCTGGAGCTGCTGATCAACGAGTGCGGCTACAGGGGCTG CACCTGGGCGGCTGAGGAAGCCCCGACGGCATCATCTACACCG AGAAGGAGAAGTACAACCTACGGCGTGATCATCGACACCAAGGC CTACAGCAAGGGCTACAACCTGCCCATCGGCCAGATCGACGAG ATGATCAGGTACATCATCGAGAACAACGAGAGGAACATCAAGA GGAACACCAACTGCTGGTGGAAACAACCTTCGAGAAGAAGCTGAA CGAGTTCTACTTCAGCTTCATCAGCGGCGAGTTCACCGGCAACA TCGAGGAGAAGCTGAACAGGATCTTCATCAGCACCAACATCAA

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	AMSVKTLLEYLANEI KANRISFLEMEKYF DNKV		GGGCAACGCCATGAGCGTGAAGACCCTGCTGTACCTGGCCAACG AGATCAAGGCCAACAGGATCAGCTTCCTGGAGATGGAGAAGTA CTTCGACAACAAGGTG
10	EGIKSNISLLKDEL GQISHISHEYL AFDSKQNRFLFEMKV LELLVNEYGFKGRH LGGSRKPDGIVYSTT LEDNFGIIVDTKAYS EGYSLPISQADEMER YVRENSNRDEEVNP NKWWENFSEEVKK YYFVFISGSFKGKFE EQLRRLSMTTGVNG SAVNVNLLLGAEK IRSGEMTIEELERAM FNNSEFI	91	GAGGGCATCAAGAGCAACATCAGCCTGCTGAAGGACGAGCTGA GGGGCCAGATCAGCCACATCAGCCACGAGTACCTGAGCCTGATC GACCTGGCCTTCGACAGCAAGCAGAACAGGCTGTTTCGAGATGA AGGTGCTGGAGCTGCTGGTGAACGAGTACGGCTTCAAGGGCAG GCACCTGGGCGGCAGCAGGAAGCCCGACGGCATCGTGTACAGC ACCACCCTGGAGGACAACTTCGGCATCATCGTGGACACCAAGGC CTACAGCGAGGGCTACAGCCTGCCCATCAGCCAGGCCGACGAG ATGGAGAGGTACGTGAGGGAGAACAGCAACAGGGACGAGGAG GTGAACCCCAACAAGTGGTGGGAGAACTTCAGCGAGGAGGTGA AGAAGTACTACTTCGTGTTTCATCAGCGGCAGCTTCAAGGGCAAG TTCGAGGAGCAGCTGAGGAGGCTGAGCATGACCACCGGCGTGA ACGGCAGCGCCGTGAACGTGGTGAACCTGCTGCTGGGCGCCGA GAAGATCAGGAGCGGCGAGATGACCATCGAGGAGCTGGAGAGG GCCATGTTCAACAACAGCGAGTTCATC
11	ISKTNVLELKDKVR DKLKYVDNRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKPdGIIISYDI NGVIIDNKAYSSGY NLPINQADEMIRYIE ENQTRDKKINPNKW WESFDDKVKDFNYL FVSSFFKGNFKNNL KHIANRTGVNNGVI NVENLLYFAEELKS GRLSYVDLFKMYDN DEINI	92	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACAACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCGACGGCATCATCAGCTACGACAT CAACGGCGTGATCATCGACAACAAGGCCTACAGCAGCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACAAGAAGATCAACCCCAACAAGTGG TGGGAGAGCTTCGACGACAAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGAACGGCGGCGTGATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGAGCGGCAGG CTGAGCTACGTGGACCTGTTCAAGATGTACGACAACGACGAGAT CAACATC
12	ISKTNVLELKDKVR DKLKYVDHRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKPdGIIISYDI NGVIIDNKAYSTGY NLPINQADEMIRYIE ENQTRDKKINSNKW WESFDDKVKNFNYL FVSSFFKGNFKNNL KHIANRTGVNNGAI NVENLLYFAEELKA GRLSYVDSFTMYDN DEIYV	93	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCGACGGCATCATCAGCTACGACAT CAACGGCGTGATCATCGACAACAAGGCCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACAAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGACAAGGTGAAGAACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGAACGGCGGCGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGCGCAGG CTGAGCTACGTGGACAGCTTCACCATGTACGACAACGACGAGAT CTACGTG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
13	KA EKSEFLIEKDKLR EKLDTLPHDYLSMV DLAYDSKQNRLEFEM KTIELLINECNYKGL HLGGTRKPDGIVYT NNEVENYGIIDTKA YSKGYNLPISQVDE MTRYVEENNKREK KRNPNEWNNFDS NVKKFYFSFISGKFV GNIEEKLQRITLFEI YGNAITVTLLYIAN EIKANRMKKS DIME YFNDKV	94	AAGGCCGAGAAGAGCGAGTTCCTGATCGAGAAGGACAAGCTGA GGGAGAAGCTGGACACCCTGCCCCACGACTACCTGAGCATGGTG GACCTGGCCTACGACAGCAAGCAGAACAGGCTGTTCGAGATGA AGACCATCGAGCTGCTGATCAACGAGTGCAACTACAAGGGCCTG CACCTGGGCGGCACCAGGAAGCCCCGACGGCATCGTGTACACCA ACAACGAGGTGGAGAACTACGGCATCATCATCGACACCAAGGC CTACAGCAAGGGCTACAACCTGCCCATCAGCCAGGTGGACGAG ATGACCAGGTACGTGGAGGAGAACAACAAGAGGGAGAAGAAG AGGAACCCCAACGAGTGGTGGAACTTCGACAGCAACGTGA AGAAGTTCTACTTCAGCTTCATCAGCGCAAGTTCGTGGGCAAC ATCGAGGAGAAGCTGCAGAGGATCACCTGTTCACCGAGATCTA CGGCAACGCCATCACCGTGACCACCCTGCTGTACATCGCCAACG AGATCAAGGCCAACAGGATGAAGAAGAGCGACATCATGGAGTA CTTCAACGACAAGGTG
14	ISKTNVLELKDKVR DKLKYVDHRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKP DGII SYNI NGVIIDNKAYSTGY NLPINQADEMIRYIE ENQTRDEKINSNKW WESFDDEVKDFNYL FVSSFFKGNFKNNL KHIANRTGVNNGAI NVENLLYFAELKA GRLSYVDSFTMYDN DEIYV	95	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCCGACGGCATCATCAGCTACAACAT CAACGGCGTGATCATCGACAACAAGGCCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGACGAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGACGGCGGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGGCAGG CTGAGCTACGTGGACAGCTTCAACATGTACGACAACGACGAGAT CTACGTG
15	ISKTNILELKDKVRD KLKYVDHRYLALID LAYDGTANRDFEIQ TIDLLINELKFKGVR LGESRKP DGII SYNIN GVIIDNKAYSTGYNL PINQADEMIRYIEEN QTRDEKINSNKWWE SFDEKVKDFNYL FV SSFFKGNFKNNLKHI ANRTGVNNGAINVE NLLYFAEELKAGRIS YLDSFKMYNNDEIY L	96	ATCAGCAAGACCAACATCCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCCGACGGCATCATCAGCTACAACAT CAACGGCGTGATCATCGACAACAAGGCCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGAGAAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGACGGCGGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGGCAGG ATCAGCTACCTGGACAGCTTCAAGATGTACAACAACGACGAGAT CTACCTG
16	ISKTNVLELKDKVR DKLKYVDHRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKP DGII SYNI NGVIIDNKAYSTGY NLPINQADEMIRYIE ENQTRDEKINSNKW WESFDDKVKDFNYL FVSSFFKGNFKNNL KHIANRTGVSSGAI NVENLLYFAEELKA	97	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCCGACGGCATCATCAGCTACAACAT CAACGGCGTGATCATCGACAACAAGGCCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGACAAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGAGCGGCCGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGGCAGG GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGGCAGG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	GRLSYVDSFKMYDN DEIYV		CTGAGCTACGTGGACAGCTTCAAGATGTACGACAACGACGAGAT CTACGTG
17	ISKTNVLELKDKVR NKLKYVDHRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKPDGIISYDI NGVIIDNKSYSYTGYN LPINQADEMIRYIEE NQTRDEKINSNKW WESFDEKVKDFNYL FVSSFFKGNFKNNL KHIANRTGVNNGAI NVENLLYFAEELKS GRLSYVDSFTMYDN DEIYV	98	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGA ACAAGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCGACGGCATCATCAGCTACGACAT CAACGGCGTGATCATCGACAACAAGAGCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGAGAAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGACGGCGGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGAGCGGCAGG CTGAGCTACGTGGACAGCTTCACCATGTACGACAACGACGAGAT CTACGTG
18	ISKTNVLELKDKVR DKLKYVDHRYLSLI DLAYDGNANRDFEI QTIDLLINELNFKGV RLGESRKPDGIISYNI NGVIIDNKAYSTGY NLPINQADEMIRYIE ENQTRDEKINSNKW WESFDDKVKDFNYL FVSSFFKGNFKNNL KHIANRTGVSGGAI NVENLLYFAEELKA GRLSYADSFTMYDN DEIYV	99	ATCAGCAAGACCAACGTGCTGGAGCTGAAGGACAAGGTGAGGG ACAAGCTGAAGTACGTGGACCACAGGTACCTGAGCCTGATCGAC CTGGCCTACGACGGCAACGCCAACAGGGACTTCGAGATCCAGA CCATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGG CTGGGCGAGAGCAGGAAGCCCGACGGCATCATCAGCTACAACA TCAACGGCGTGATCATCGACAACAAGGCCTACAGCACCGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACAGCAACAAGTGG TGGGAGAGCTTCGACGACAAGGTGAAGGACTTCAACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGAGCGGCGGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGGCCGCGCAGG CTGAGCTACGCCGACAGCTTCACCATGTACGACAACGACGAGAT CTACGTG
19	IAKTNVLGLKDKVR DRLKYVDHRYLALI DLAYDGTANRDFEI QTIDLLINELKFKGV RLGESRKPDGIISYN VNGVIIDNKAYS YNLPINQADEMIRYI EENQTRDEKINANK WWESFDDKVEEFSY LFVSSFFKGNFKNNL KHIANRTGVNNGAI NVENLLYFAEELKS GRLSYMDSFSLYDN DEICV	100	ATCGCCAAGACCAACGTGCTGGGCCTGAAGGACAAGGTGAGGG ACAGGCTGAAGTACGTGGACCACAGGTACCTGGCCCTGATCGAC CTGGCCTACGACGGCACCGCCAACAGGGACTTCGAGATCCAGAC CATCGACCTGCTGATCAACGAGCTGAAGTTCAAGGGCGTGAGGC TGGGCGAGAGCAGGAAGCCCGACGGCATCATCAGCTACAACGT GAACGGCGTGATCATCGACAACAAGGCCTACAGCAAGGGCTAC AACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCGA GGAGAACCAGACCAGGGACGAGAAGATCAACGCCAACAAAGTGG TGGGAGAGCTTCGACGACAAGGTGGAGGAGTTCACTACCTGTT CGTGAGCAGCTTCTTCAAGGGCAACTTCAAGAACAACCTGAAGC ACATCGCCAACAGGACCGGCGTGACGGCGGCCATCAACGT GGAGAACCTGCTGTACTTCGCCGAGGAGCTGAAGAGCGGCAGG CTGAGCTACATGGACAGCTTCAGCCTGTACGACAACGACGAGAT CTGCGTG
20	ELKDEQSEKRKAKF LKETKLPMKYIELLD IAYDGRNRNDFEIVT MELFREYRLNSKL LGGGRKPDGLIYTD DFGVIVDTKAYGEG YSKSINQADEMIRYI EDNKRREDEKRNPIK WWESFPSSISQNNFY FLWVSSKFVGFQFE QLAYTANETQTKGG	101	GAGCTGAAGGACGAGCAGAGCGAGAAGAGGAAGGCCAAGTTCC TGAAGGAGACCAAGCTGCCCATGAAGTACATCGAGCTGCTGGA CATCGCCTACGACGGCAAGAGGAACAGGGACTTCGAGATCGTG ACCATGGAGCTGTTCAAGGAGGTGTACAGGCTGAACAGCAAGC TGCTGGGCGGCGGCAGGAAGCCCGACGGCCTGATCTACACCGA CGACTTCGGCGTGATCGTGGACACCAAGGCCTACGGCGAGGGCT ACAGCAAGAGCATCAACCAGGCCGACGAGATGATCAGGTACAT CGAGGACAACAAGAGGAGGGACGAGAAGAGGAACCCCATCAA GTGGTGGGAGAGCTTCCCCAGCAGCATCAGCCAGAACAACCTTCT ACTTCTGTGGGTGAGCAGCAAGTTCGTGGGCAAGTTCCAGGAG CAGCTGGCCTACACCGCCAACGAGACCCAGACCAAGGGCGGCG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	AINVEQILIGADLIM QKMLDINTIPSFFEN QEIIIF		CCATCAACGTGGAGCAGATCCTGATCGGCGCCGACCTGATCATG CAGAAGATGCTGGACATCAACACCATCCCCAGCTTCTTCGAGAA CCAGGAGATCATCTTC
21	IFKTNVLELKDSIRE KLDYIDHRYLSLVD LAYDSKANRDFEIQ TIDLLINELDFKGLR LGESRKPDIISYDIN GVIIDNKAYSKGYN LPINQADEMIRYIQE NQSREKINPNKWW ENFEDKVIKFNLYFI SSLFVGGFKKNLQHI ANRTGVNGGAIDVE NLLYFAEEIKSGRLT YKDSFSRYINDEIKM	102	ATCTTCAAGACCAACGTGCTGGAGCTGAAGGACAGCATCAGGG AGAAGCTGGACTACATCGACCACAGGTACCTGAGCCTGGTGGAC CTGGCCTACGACAGCAAGGCCAACAGGGACTTCGAGATCCAGA CCATCGACCTGCTGATCAACGAGCTGGACTTCAAGGGCCTGAGG CTGGGCGAGAGCAGGAAGCCCCGACGGCATCATCAGCTACGACA TCAACGGCGTGATCATCGACAACAAGGCCTACAGCAAGGGCTA CAACCTGCCCATCAACCAGGCCGACGAGATGATCAGGTACATCC AGGAGAACCAGAGCAGGAACGAGAAGATCAACCCCAACAAGTG GTGGGAGAACTTCGAGGACAAGGTGATCAAGTTCAACTACCTGT TCATCAGCAGCCTGTTCTGTTGGGCGGCTTCAAGAAGAACCTGCAG CACATCGCCAACAGGACCGGCGTGAACGGCGGCGCCATCGACG TGGAGAACCTGCTGTACTTCGCCGAGGAGATCAAGAGCGGCAG GCTGACCTACAAGGACAGCTTCAGCAGGTACATCAACGACGAG ATCAAGATG
22	LPVKSEVSFVKDYL RTHLTHVDHRYLIL VDLFGDGSSDRDYE MKTAELFTAELGFM GARLGDTRKPDVCV YHGANGLIIDNKAY GKGYSLPIKQADEIY RYIEENKERDARLNP NQWWKVFEDEVTH FRFAFISGSFTGGFK DRIELISMRSIGCA AVNSVNLMLMAEEL KSGRLDYEEWFQYF DCNDEISF	103	CTGCCCCGTGAAGAGCGAGGTGAGCGTGTTCAAGGACTACCTGAG GACCCACCTGACCCACGTGGACCACAGGTACCTGATCCTGGTGG ACCTGGGCTTCGACGGCAGCAGCGACAGGGACTACGAGATGAA GACCGCCGAGCTGTTACCGCCGAGCTGGGCTTCATGGGCGCCA GGCTGGGCGACACCAGGAAGCCCCGACGTGTGCGTGTACCACGG CGCCAACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGGC TACAGCCTGCCCATCAAGCAGGCCGACGAGATCTACAGGTACAT CGAGGAGAACAAGGAGAGGGACGCCAGGCTGAACCCCAACCAG TGGTGGAAAGGTGTTTCGACGAGAGCGTGACCCACTTCAGGTTCGC CTTCATCAGCGGCAGCTTCACCGGCGGCTTCAAGGACAGGATCG AGCTGATCAGCATGAGGAGCGGCATCTGCGGCGCCGCCGTGAA CAGCGTGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGGC AGGCTGGACTACGAGGAGTGGTTCCAGTACTTCGACTGCAACGA CGAGATCAGCTTC
23	ISVKSMDMAVVKDSV RERLAHVSHEYLLILI DLGFDGTSDRDYEI QTAEFLTRELDLGG RLGDTRKPDVCIYY GKDGMIIIDNKAYGK GYSPLIKQADEMYR YLEENKERNEKINPN RWWKVFDEGVTDY RFAFVSGSFTGGFKD RLENIHMRSGLCGG AIDSVTLLLLAEELK AGRMEYSEFFRLFD CNDEVTF	104	ATCAGCGTGAAGAGCGACATGGCCGTGGTGAAGGACAGCGTGA GGGAGAGGCTGGCCACGTGAGCCACGAGTACCTGATCCTGATC GACCTGGGCTTCGACGGCACCAGCAGCAGGGACTACGAGATCC AGACCGCCGAGCTGTTACCCAGGGAGCTGGACTTCCTGGGCGGC AGGCTGGGCGACACCAGGAAGCCCCGACGTGTGCATCTACTACG GCAAGGACGGCATGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACC TGGAGGAGAACAAGGAGAGGAACGAGAAGATCAACCCCAACA GGTGGTGGAAAGGTGTTTCGACGAGGGCGTGACCGACTACAGGTTC GCCTTCGTGAGCGGCAGCTTCACCGGCGGCTTCAAGGACAGGCT GGAGAACATCCACATGAGGAGCGGCCTGTGCGGCGGCGCCATC GACAGCGTGACCTGCTGCTGCTGGCCGAGGAGCTGAAGGCCG GCAGGATGGAGTACAGCGAGTTCTTCAGGCTGTTTCGACTGCAAC GACGAGGTGACCTTC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
24	ELKDKAADAVKAK FLKLTGLSMKYIELL DIAYDSSRNDRFEIL TADLFKNVYGLDA MHLGGGRKPDIAQ TSHFGIIDTKAYGN GYSKISQEDEMVR YIEDNQQRSITRNSV EWWKNFNSSIPSTAF YFLWVSSKFVGFDF DQLLATYNRTNTCG GALNVEQLLIGAYK VKAGLLGIGQIPSYF KNKEIAW	105	GAGCTGAAGGACAAGGCCGCCGACGCCGTGAAGGCCAAGTTCC TGAAGCTGACCGGCCTGAGCATGAAGTACATCGAGCTGCTGGAC ATCGCCTACGACAGCAGCAGGAACAGGGACTTCGAGATCCTGA CCGCCGACCTGTTCAAGAACGTGTACGGCCTGGACGCCATGCAC CTGGGCGGCGGCAGGAAGCCCGACGCCATCGCCAGACCAGCC ACTTCGGCATCATCATCGACACCAAGGCCTACGGCAACGGCTAC AGCAAGAGCATCAGCCAGGAGGACGAGATGGTGAGGTACATCG AGGACAACCAGCAGAGGAGCATCACCAGGAACAGCGTGGAGTG GTGGAAGAAGTTCAACAGCAGCATCCCCAGCACCCTTCTACT TCCTGTGGGTGAGCAGCAAGTTCGTGGGCAAGTTCGACGACCAG CTGCTGGCCACCTACAACAGGACCAACACCTGCGGCGGCGCCCT GAACGTGGAGCAGCTGCTGATCGGCGCCTACAAGGTGAAGGCC GGCCTGCTGGGCATCGGCCAGATCCCCAGCTACTTCAAGAACAA GGAGATCGCCTGG
25	ISVKSMDMAVVKDSV RERLAHVSHEYLLLI DLGFDGTSRDRYEI QTAELLTRELDLFG GRLGDTRKPDVCIY YGKDGMIIDNKAYG KGYSPLIKQADEMY RYLEENKERNEKINP NRWWKVFDEGVT YRFAFVSGSFTGGFK DRLENIHMRSGLCG GAIDSVTLALLAEEL KAGRMEYSEFFRLF DCNDEVTF	106	ATCAGCGTGAAGAGCGACATGGCCGTGGTGAAGGACAGCGTGA GGGAGAGGCTGGCCACGTGAGCCACGAGTACCTGCTGCTGATC GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCGAGCTGCTGACCAGGGAGCTGGACTTCCTGGGCGGC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCATCTACTACG GCAAGGACGGCATGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACC TGGAGGAGAACAAGGAGAGGAACGAGAAGATCAACCCCAACA GGTGGTGGAAAGGTGTTTCGACGAGGGCGTGACCGACTACAGGTT GCCTTCGTGAGCGGCAGCTTCACCGGCGGCTTCAAGGACAGGCT GGAGAACATCCACATGAGGAGCGGCCTGTGCGGCGGCGCCATC GACAGCGTGACCTGCTGCTGCTGGCCGAGGAGCTGAAGGCCG GCAGGATGGAGTACAGCGAGTTCTTCAGGCTGTTGACTGCAAC GACGAGGTGACCTTC
26	ELKDEQAQEKRAKF LKETNLPKMYIELLD IAYDGKRNDRFEIVT MELFRNVYRLHSLK LGGGRKPDGLLYQD RFGVIVDTKAYGKG YSKSINQADEMIRYI EDNKRDRDENRNPIK WWEAFPDTPQEEF YFMWVSSKFIGKFQ EQLDYTSNETQIKG AALNVEQLLLGADL VLKGQLHISDLPSYF QNKEIEF	107	GAGCTGAAGGACGACGAGGCCGAGAAGAGGAAGGCCAAGTTCC TGAAGGAGACCAACCTGCCCATGAAGTACATCGAGCTGCTGGAC ATCGCCTACGACGGCAAGAGGAACAGGGACTTCGAGATCGTGA CCATGGAGCTGTTCAAGAACGTGTACAGGCTGCACAGCAAGCTG CTGGGCGGCGGCAGGAAGCCCGACGGCCTGCTGTACCAGGACA GGTTCGGCGTGATCGTGGACACCAAGGCCTACGGCAAGGGCTAC AGCAAGAGCATCAACCAGGCCGACGAGATGATCAGGTACATCG AGGACAACAAGAGGAGGGACGAGAACAGGAACCCCATCAAGTG GTGGGAGGCCTTCCCCGACACCATCCCCAGGAGGAGTTCTACT TCATGTGGGTGAGCAGCAAGTTCATCGGCAAGTTCAGGAGCAG CTGGACTACACCAGCAACGAGACCCAGATCAAGGGCGCCGCC TGAACGTGGAGCAGCTGCTGCTGGGCGCCGACCTGGTGCTGAAG GGCCAGCTGCACATCAGCGACCTGCCAGCTACTTCCAGAACAA GGAGATCGAGTTC
27	RNLDNVERDNRKAE FLAKTSLPPRIELLS IAYESKSNRDFEMIT AELFKDVYGLGAVH LGNACKPDALAFND DFGIIIDTKAYSNGY SKNINQEDEMVRYIE DNQIRSPDRNNNEW WLSFPPSIPENDFHF LWVSSYFTGRFEEQ LQETSARTGGTTGG ALDVEQLLIGGSLIQ	108	AGGAACCTGGACAACGTGGAGAGGGACAACAGGAAGGCCGAGT TCCTGGCCAAGACCAGCCTGCCCCCAGGTTTCATCGAGCTGCTG AGCATCGCCTACGAGAGCAAGAGCAACAGGGACTTCGAGATGA TCACCGCCGAGCTGTTCAAGGACGTGTACGGCCTGGGCGCCGTG CACCTGGGCAACGCCAAGAAGCCCGACGCCCTGGCCTTCAACGA CGACTTCGGCATCATCATCGACACCAAGGCCTACAGCAACGGCT ACAGCAAGAACATCAACCAGGAGGACGAGATGGTGAGGTACAT CGAGGACAACCAGATCAGGAGCCCCGACAGGAACAACAACGAG TGGTGGCTGAGCTTCCCCCAGCATCCCCGAGAACGACTTCCA CTTCCTGTGGGTGAGCAGCTACTTCACCGGCAGGTTTCGAGGAGC AGCTGCAGGAGACCAGCGCCAGGACCGGCGGCACCACCGGCGG CGCCCTGGACGTGGAGCAGCTGCTGATCGGCGGCAGCCTGATCC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	EGSLAPHEVPAYMQ NRVIHF		AGGAGGGGCAGCCTGGCCCCCACGAGGTGCCCCGCTACATGCA GAACAGGGTGATCCACTTC
28	SPVKSEVSVFKDYL RTHLTHVDHRYLIL VDLGFDSDDRDYE MKTAELEFTAELGFM GARLGDTRKPDVCV YHGAHGLIIDNKAY GKGYSPLIKQADEIY RYIEENKERAVRLNP NQWWKVFDESVAH FRFAFISGFTGGFK DRIELISMRSIGCGA AVNSVNLMLMAEEL KSGRLNYEEWFQYF DCNDEISL	109	AGCCCCGTGAAGAGCGAGGTGAGCGTGTTCAAGGACTACCTGA GGACCCACCTGACCCACGTGGACCACAGGTACCTGATCCTGGTG GACCTGGGCTTCGACGGCAGCAGCGACAGGGACTACGAGATGA AGACCGCCGAGCTGTTCAACGCCGAGCTGGGCTTCATGGGCGCC AGGCTGGGCGACACCAGGAAGCCCCGACGTGTGCGTGTACCACG GCGCCACGGCCTGATCATCGACAACAAGGCCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATCTACAGGTACA TCGAGGAGAACAAGGAGAGGGCCGTGAGGCTGAACCCCAACCA GTGGTGGAAGGTGTTTCGACGAGAGCGTGCCCACTTCAGGTTTCG CCTTCATCAGCGGCAGCTTCACCGGCGGCTTCAAGGACAGGATC GAGCTGATCAGCATGAGGAGCGGCATCTGCGGCGCCGCGGTGA ACAGCGTGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGAACTACGAGGAGTGGTTCCAGTACTTCGACTGCAACG ACGAGATCAGCCTG
29	TLVDIEKERKKAYFL KETSLSPRYIELLEIA FDPKRNDRFEVITAE LLKAGYGLKAKVLG GGRRPDGIAYTKDY GLIVDTKAYSNGYG KNIGQADEMIRYIED NQKRDNKRNPIEW WREFEVQIPANSYY YLWVSGRFTGRFDE QLVYTSSQTNTRGG ALEVEQLLWGADA VMKGKLNVSDDLK YMNNSIIL	110	ACCCTGGTGGACATCGAGAAGGAGAGGAAGAAGGCCTACTTCC TGAAGGAGACCAGCCTGAGCCCCAGGTACATCGAGCTGCTGGA GATCGCCTTCGACCCCAAGAGGAACAGGGACTTCGAGGTGATCA CCGCCGAGCTGCTGAAGGCCGGCTACGGCCTGAAGGCCAAGGT GCTGGGCGGCGGCAGGAGGCCCGACGGCATCGCCTACACCAAG GACTACGGCCTGATCGTGGACACCAAGGCCTACAGCAACGGCTA CGGCAAGAACATCGGCCAGGCCGACGAGATGATCAGGTACATC GAGGACAACCAGAAGAGGGGACAACAAGAGGAACCCCATCGAGT GGTGGAGGGAGTTCGAGGTGCAGATCCCCGCCAACAGCTACTAC TACCTGTGGGTGAGCGGCAGGTTACCGGCAGGTTTCGACGAGCA GCTGGTGTACACCAGCAGCCAGACCAACACCAGGGGCGGCGCC CTGGAGGTGGAGCAGCTGCTGTGGGGCGCCGACGCCGTGATGA AGGGCAAGCTGAACGTGAGCGACCTGCCCAAGTACATGAACAA CAGCATCATCAAGCTG
30	ELRDKVIEEQKAIFL QKTKLPLSYIELLEIA RDGKRSRDFELITIE LFKNIYKINARILGG ARKPDGVLYMPEFG VIVDTKAYADGYSK SIAQADEMIRYIEDN KRRDPSRNSTKWWE HFPTSIPANNFYFLW VSSVFNKFHEQLS YTAQETQTVGAALS VEQLLLGADSVLKG NLTTEKFIDSFKNQE IVF	111	GAGCTGAGGGACAAGGTGATCGAGGAGCAGAAGGCCATCTTCC TGCAGAAGACCAAGCTGCCCCTGAGCTACATCGAGCTGCTGGAG ATCGCCAGGGACGGCAAGAGGAGCAGGGACTTCGAGCTGATCA CCATCGAGCTGTTCAAGAACATCTACAAGATCAACGCCAGGATC CTGGGCGGCGCCAGGAAGCCCCGACGGCGTGCTGTACATGCCCG AGTTTCGGCGTGATCGTGGACACCAAGGCCTACGCCGACGGCTAC AGCAAGAGCATCGCCAGGCCGACGAGATGATCAGGTACATCG AGGACAACAAGAGGAGGGACCCAGCAGGAACAGCACCAAGTG GTGGGAGCACTTCCCCACCAGCATCCCCGCCAACAACTTCTACT TCCTGTGGGTGAGCAGCGTGTTCTGTGAACAAGTTCCACGAGCAG CTGAGCTACACCGCCCAGGAGACCCAGACCGTGGGCGCCGCCCT GAGCGTGGAGCAGCTGCTGCTGGGCGCCGACAGCGTGCTGAAG GGCAACCTGACCACCGAGAAGTTCATCGACAGCTTCAAGAACCA GGAGATCGTGTTTC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
31	GATKSDLSLLKDDIR KKLNHINHXYLVLI DLGFDGTADRDYEL QTADLLTSELAFKG ARLGDSRKPDPVCVY HDKNGLIIDNKAYG SGYSLPIKQADEML RYIEENQKRDKALN PNEWWTIFDDAVSK FNFAFVSGEFTGGFK DRLENISRRSYTNGA AINSVNLLLLAEEIK SGRISYGDAFTKFEC NDEIII	112	GGCGCCACCAAGAGCGACCTGAGCCTGCTGAAGGACGACATCA GGAAGAAGCTGAACCACATCAACCACAAGTACCTGGTGTGATC GACCTGGGCTTCGACGGCACC GCCGACAGGGACTACGAGCTGC AGACCGCCGACCTGCTGACCAGCGAGCTGGCCTTCAAGGGCGCC AGGCTGGGCGACAGCAGGAAGCCCCGACGTGTGCGTGTACCACG ACAAGAACGGCCTGATCATCGACAACAAGGCCTACGGCAGCGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGCTGAGGTACA TCGAGGAGAACCAGAAGAGGGACAAGGCCCTGAACCCCAACGA GTGGTGGACCATCTTCGACGACGCCGTGAGCAAGTTCAACTTCG CCTTCGTGAGCGGCGAGTTTACCGGGCGGCTTCAAGGACAGGCTG GAGAACATCAGCAGGAGGAGCTACACCAACGGCGCCGCCATCA ACAGCGTGAACCTGCTGCTGCTGGCCGAGGAGATCAAGAGCGG CAGGATCAGCTACGGCGACGCCTTACCAAGTTCGAGTGCAACG ACGAGATCATCATC
32	ELRNAALDKQKVN INKTGLPMKYIELLE IAFDGSRNRDFEMV TADLFKNVYGFNSIL LGGGRKPDGLIFTDR FGVIIDTKAYGNGYS KSIQEDEMVRYIED NQLRDSNRNSVEW WKNFDEKIESENFYF MWISSKFIGQFSDQL QSTSDRTNTKGAAL NVEQLLLGAAAARD GKLDINSLPIYMNNK EILW	113	GAGCTGAGGAACGCCGCCCTGGACAAGCAGAAGGTGAACCTTCA TCAACAAGACCGGCCTGCCCATGAAGTACATCGAGCTGCTGGAG ATCGCCTTCGACGGCAGCAGGAACAGGGACTTCGAGATGGTGA CCGCCGACCTGTTCAAGAACGTGTACGGCTTCAACAGCATCCTG CTGGGCGGCGGCAGGAAGCCCCGACGGCCTGATCTTACCGACA GGTTCGGCGTGATCATCGACACCAAGGCCTACGGCAACGGCTAC AGCAAGAGCATCGGCCAGGAGGACGAGATGGTGAGGTACATCG AGGACAACCAGCTGAGGGACAGCAACAGGAACAGCGTGGAGTG GTGGAAGAAGTTCGACGAGAAGATCGAGAGCGAGAAGTTCTAC TTCATGTGGATCAGCAGCAAGTTCATCGGCCAGTTCAGCGACCA GCTGCAGAGCACCAGCGACAGGACCAACACCAAGGGCGCCGCC CTGAACGTGGAGCAGCTGCTGCTGGGCGCCGCCGCCAGGG ACGGCAAGCTGGACATCAACAGCCTGCCCATCTACATGAACAAC AAGGAGATCCTGTGG
33	ELKDEQSEKRKAYF LKETNLPLKYIELLDI AYDGKRNDRDFEIVT MELFRNVYRLQSKL LGGVRKPDGLLYKH RFGIIVDTKAYGEGY SKSISQADEMIRYIE DNKRDRDENRNSTK WWEHFPDCIPKQSF YFMWVSSKFVGKQF EQLDYTANETKTNG AALNVEQLLWGAD LVAKGKLDISQLPSY FQNKEIEF	114	GAGCTGAAGGACGACGAGCAGAGCGAGAAGAGGAAGGCCTACTTCC TGAAGGAGACCAACCTGCCCCTGAAGTACATCGAGCTGCTGGAC ATCGCCTACGACGGCAAGAGGAACAGGGACTTCGAGATCGTGA CCATGGAGCTGTTCAAGAACGTGTACAGGCTGCAGAGCAAGCTG CTGGGCGGCGTGAGGAAGCCCCGACGGCCTGCTGTACAAGCACA GGTTCGGCATCATCGTGACACCAAGGCCTACGGCGAGGGCTAC AGCAAGAGCATCAGCCAGGCCGACGAGATGATCAGGTACATCG AGGACAACAAGAGGAGGGACGAGAACAGGAACAGCACCAAGT GGTGGGAGCACTTCCCCGACTGCATCCCCAAGCAGAGCTTCTAC TTCATGTGGGTGAGCAGCAAGTTCGTGGGCAAGTTCAGGAGCA GCTGGACTACACCGCCAACGAGACCAAGACCAACGGCGCCGCC CTGAACGTGGAGCAGCTGCTGTGGGGCGCCGACCTGGTGGCCAA GGGCAAGCTGGACATCAGCCAGCTGCCAGCTACTTCCAGAACA AGGAGATCGAGTTC
34	HNNKFKNYLRENSE LSFKFIELIDAYDGN RNRDMEIITAELLKE IYGLNVKLLGGGRK PDILAYTDDIGHIIT KAYKDGYGKQINQ ADEMIRYIEDNQRR DLIRNPNEWWRYPF KSIKEKIYFMWISS YFKNNFYEQVQYTA QETKSIGAAALNVRQ LLLCADAIQKEVLSL DTFLGSFRNEEINL	115	CACAACAACAAGTTCAAGAACTACCTGAGGGAGAACAGCGAGC TGAGCTTCAAGTTCATCGAGCTGATCGACATCGCCTACGACGGC AACAGGAACAGGGACATGGAGATCATCACCGCCGAGCTGCTGA AGGAGATCTACGGCCTGAACGTGAAGCTGCTGGGCGGCGGCAG GAAGCCCAGCATCCTGGCCTACACCGACGACATCGGCATCATCA TCGACACCAAGGCCTACAAGGACGGCTACGGCAAGCAGATCAA CCAGGCCGACGAGATGATCAGGTACATCGAGGACAACCAGAGG AGGGACCTGATCAGGAACCCCAACGAGTGGTGGAGGTACTTCCC CAAGAGCATCAGCAAGGAGAAGATCTACTTCATGTGGATCAGC AGCTACTTCAAGAACAACCTTCTACGAGCAGGTGCAGTACACCGC CCAGGAGACCAAGAGCATCGGCGCCGCCCTGAACGTGAGGCAG CTGCTGCTGTGCGCCGACCCATCCAGAAGGAGGTGCTGAGCCT GGACACCTTCTGGGCGAGCTTCAGGAACGAGGAGATCAACCTG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
35	LPVKSEVSILKDYLRL SHLTHIDHKYLILVD LGYDGTSDRDYEIQ TAQLLTAELSFLGGR LGDTRKPDVCIYYE DNGLIIDNKAYGKG YSLPMKQADEMYR YIEENKERSELLNPN CWWNIFDKDVKTFRH FAFLSGEFTGGFRDR LNHISMRSRGMRGAA VNSANLLIMAEKLK AGTMEYEEFFRLFD TNDEILF	116	CTGCCCCGTGAAGAGCGAGGTGAGCATCCTGAAGGACTACCTGA GGAGCCACCTGACCCACATCGACCACAAGTACCTGATCCTGGTG GACCTGGGCTACGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCCAGCTGCTGACCGCCGAGCTGAGCTTCCTGGGCGGC AGGCTGGGCGACACCAGGAAGCCCCGACGTGTGCATCTACTACG AGGACAACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATGAAGCAGGCCGACGAGATGTACAGGTAC ATCGAGGAGAACAAGGAGAGGAGCGAGCTGCTGAACCCCAACT GCTGGTGGAACATCTTCGACAAGGACGTGAAGACCTTCCACTTC GCCTTCCTGAGCGGCGAGTTCACCGGCGGCTTCAGGGACAGGCT GAACCACATCAGCATGAGGAGCGGCATGAGGGGCGCCGCCGTG AACAGCGCCAACCTGCTGATCATGGCCGAGAAGCTGAAGGCCG GCACCATGGAGTACGAGGAGTTCTTCAGGCTGTTCGACACCAAC GACGAGATCCTGTTC
36	LPVKSQVSILKDYLRL SYLSHVDHKYLILLD LGFDGTSDRDYEIW TAQLLTAELSFLGGR LGDTRKPDVCIYYE DNGLIIDNKAYGKG YSLPIKQADEMYRYI EENKERSDLLNPN WWNIFGEGVKTFRF AFLSGEFTGGFKDRL NHISMRSRGIKGA VNSANLLIMAEQLKS GTMSYEEFFQLFDY NDEIIF	117	CTGCCCCGTGAAGAGCCAGGTGAGCATCCTGAAGGACTACCTGAG GAGCTACCTGAGCCACGTGGACCACAAGTACCTGATCCTGCTGG ACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCTG GACCGCCCAGCTGCTGACCGCCGAGCTGAGCTTCCTGGGCGGCA GGCTGGGCGACACCAGGAAGCCCCGACGTGTGCATCTACTACGA GGACAACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGGC TACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACAT CGAGGAGAACAAGGAGAGGAGCGACCTGCTGAACCCCAACTGC TGCTGGGAACATCTTCGCGGAGGGCGTGAAGACCTTCAGGTTTCG CTTCCTGAGCGGCGAGTTCACCGGCGGCTTCAGGACAGGCTGA ACCACATCAGCATGAGGAGCGGCATCAAGGGCGCCGCCGTGAA CAGCGCCAACCTGCTGATCATGGCCGAGCAGCTGAAGAGCGGC ACCATGAGCTACGAGGAGTTCTTCAGCTGTTCGACTACAACGA CGAGATCATCTTC
37	VSKTNILELKDNTR KLVDLHRYLSLFD LAYDDKASRDFEIQ TIDLLINELQFKGLR LGERRKPDGIISYGV NGVIIDNKAYS KGYNLPIRQADEMIRYIQ ENQSRDEKLNP KNWENFEEETSKFNYL FISSKFISGFKK NLQYIADRTGVN GGAINVENLLCFAE MLKSGKLEYNDF FNQYNND EIIIM	118	GTGAGCAAGACCAACATCCTGGAGCTGAAGGACAACACCAGGG AGAAGCTGGTGATACCTGGACCACAGGTACCTGAGCCTGTTCGAC CTGGCCTACGACGACAAGGCCAGCAGGGACTTCGAGATCCAGA CCATCGACCTGCTGATCAACGAGCTGCAGTTCAAGGGCCTGAGG CTGGGCGAGAGGAGGAAGCCCCGACGCATCATCAGCTACGGCG TGAACGGCGTGATCATCGACAACAAGGCCTACAGCAAGGGCTA CAACCTGCCCATCAGGCAGGCCGACGAGATGATCAGGTACATCC AGGAGAACCAGAGCAGGGACGAGAAGCTGAACCCCAACAAGTG GTGGGAGAACTTCGAGGAGGAGACCAGCAAGTTCAACTACCTG TTCATCAGCAGCAAGTTCATCAGCGGCTTCAAGAAGAACCTGCA GTACATCGCCGACAGGACCGGCGTGAACGGCGGCGCCATCAAC GTGGAGAACCTGCTGTGCTTCGCCGAGATGCTGAAGAGCGGCAA GCTGGAGTACAACGACTTCTTCAACCAGTACAACAACGACGAGA TCATCATG
38	LPVKSQVSILKDYLRL SCLSHVDHKYLILLD LGFDGTSDRDYEIQ TAQLLTAELSFLGGR LGDTRKPDVCIYYE DNGLIIDNKAYGKG YSLPIKQADEMYRYI EENKERSSELLNPN CWVNIFDEGVKTFRFA FLSGEFTGGFKDRLN HISMRSRGIKGA AVNSANLLIMAEQLKSGTM	119	CTGCCCCGTGAAGAGCCAGGTGAGCATCCTGAAGGACTACCTGAG GAGCTGCCTGAGCCACGTGGACCACAAGTACCTGATCCTGCTGG ACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCCA GACCGCCCAGCTGCTGACCGCCGAGCTGAGCTTCCTGGGCGGCA GGCTGGGCGACACCAGGAAGCCCCGACGTGTGCATCTACTACGA GGACAACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGGC TACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACAT CGAGGAGAACAAGGAGAGGAGCGAGCTGCTGAACCCCAACTGC TGCTGGGAACATCTTCGACGAGGGCGTGAAGACCTTCAGGTTTCG CTTCCTGAGCGGCGAGTTCACCGGCGGCTTCAGGACAGGCTGA ACCACATCAGCATGAGGAGCGGCATCAAGGGCGCCGCCGTGAA CAGCGCCAACCTGCTGATCATCGCCGAGCAGCTGAAGAGCGGC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	SYEEFFQLFDQNDEI TV		ACCATGAGCTACGAGGAGTTCTTCCAGCTGTTTCGACCAGAACGA CGAGATCACCGTG
39	MSSKSEISVIKDNIR KRLNHINHKYLVLD LGFDGTADRDYELQ TADLLTSELSFKGAR LGDTRKPDVVCVYHG TNGLIIDNKAYGKG YSLPIKQADEMLRYI EENQKRDKSLNPNE WWTIFDDAVSKFNF AFVSGEFTGGFKDR LENISRRSSVNGAAI NSVNLALLAEEIKSG RMSYSDAFKNFDCN KEITI	120	ATGAGCAGCAAGAGCGAGATCAGCGTGATCAAGGACAACATCA GGAAGAGGCTGAACCACATCAACCACAAGTACCTGGTGCTGATC GACCTGGGCTTCGACGGCACCAGCCGACAGGGACTACGAGCTGC AGACCGCCGACCTGCTGACCAGCGAGCTGAGCTTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACCACG GCACCAACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGCTGAGGTACA TCGAGGAGAACCAGAAAGAGGGACAAGAGCCTGAACCCCAACGA GTGGTGGACCATCTTCGACGACGCCGTGAGCAAGTTCAACTTCG CCTTCGTGAGCGGCGAGTTTACCGGCGGCTTCAAGGACAGGCTG GAGAACATCAGCAGGAGGAGCAGCGTGAACGGCGCCGCCATCA ACAGCGTGAACCTGCTGCTGCTGGCCGAGGAGATCAAGAGCGG CAGGATGAGCTACAGCGACGCCTTCAAGAACTTCGACTGCAACA AGGAGATCACCATC
40	RNLDKVERDSRKA FLAKTSLPPRIELLS IAYESKSNRDFEMIT AEFFKDVYGLGAVH LGNARKPDALAFTD NFGIVIDTKAYSNGY SKNINQEDEMVR DNQIRSPERNKNEW WLSFPPSIPENNHF LWVSSYFTGYFEEQ LQETSDRAGGMTGG ALDIEQLLIGGSLVQ EGKLAPHDIPEYMQ NRVIHF	121	AGGAACCTGGACAAGGTGGAGAGGGACAGCAGGAAGGCCGAGT TCCTGGCCAAGACCAGCCTGCCCCCAGGTTTCATCGAGCTGCTG AGCATCGCCTACGAGAGCAAGAGCAACAGGGACTTCGAGATGA TCACCGCCGAGTTCTTCAAGGACGTGTACGGCCTGGGCGCCGTG CACCTGGGCAACGCCAGGAAGCCCGACGCCCTGGCCTTACCCGA CAACTTCGGCATCGTGATCGACACCAAGGCCTACAGCAACGGCT ACAGCAAGAACATCAACCAGGAGGACGAGATGGTGAGGTACAT CGAGGACAACCAGATCAGGAGCCCCGAGAGGAACAAGAACGAG TGCTGGCTGAGCTTCCCCCAGCATCCCCGAGAACAACTTCCA CTTCCTGTGGGTGAGCAGCTACTTACCGGCTACTTCGAGGAGC AGCTGCAGGAGACCAGCGACAGGGCCGCGGCATGACCGGCGG CGCCCTGGACATCGAGCAGCTGCTGATCGGCGGCAGCCTGGTGC AGGAGGGCAAGCTGGCCCCCACGACATCCCCGAGTACATGCA GAACAGGGGTGATCCACTC
41	APVKSEVSLCKDILR SHLTHVDHKYLILL DLGFDGTSRDRYEI QTAQLLTAELDFKG ARLGDTRKPDVVCVY YGEDGLILDNKAYG KGYSPLIKQADEMY RYIEENKERNERLNP NKWWEIFDKDVVR YHFAFVSGTFTGGF KERLDNIRMRSIGCG AAVNSMNLALLMAE ELKSGRLGYKECFA LFDCNDEIAF	122	GCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACATCCTGA GGAGCCACCTGACCCACGTGGACCACAAGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCAGCTGCTGACCGCCGAGCTGGACTTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCCTGGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAGGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAAGGACGTGGTGAGGTACCACTTCG CCTTCGTGAGCGGCACCTTACCGGCGGCTTCAAGGAGAGGCTG GACAACATCAGGATGAGGAGCGGCATCTGCGGCGCCGCCGTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACAAGGAGTGCTTCGCCCTGTTTCGACTGCAACG ACGAGATCGCCTTC
42	SCVKDEVNDIVDRV RVKLNIDHKYLILI SLAYSDETERTKKN SDARDFEIQTAEFLT KELGFNGIRLGESNK PDVLISFGANGTIIDN KSYKDGFNIPRVTS QMIYINENNQRRT QLNPNEWKWNFDSS VSNTFLFVTSFLKG SFKNQIEYISNATNG	123	AGCTGCGTGAAGGACGAGGTGAACGACATCGTGGACAGGGTGA GGGTGAAGCTGAAGAACATCGACCACAAGTACCTGATCCTGATC AGCCTGGCCTACAGCGACGAGACCGAGAGGACCAAGAAGAACA GCGACGCCAGGGACTTCGAGATCCAGACCGCCGAGCTGTTACC AAGGAGCTGGGCTTCAACGGCATCAGGCTGGGCGAGAGCAACA AGCCCGACGTGCTGATCAGCTTCGGCGCCAACGGCACCATCATC GACAACAAGAGCTACAAGGACGGCTTCAACATCCCCAGGGTGA CCAGCGACCAGATGATCAGGTACATCAACGAGAACAACCAGAG GACCACCCAGCTGAACCCCAACGAGTGGTGGAAGAACTTCGAC AGCAGCGTGAGCAACTACACCTTCTGTTCGTGACCAGCTTCTC GAAGGGCAGCTTCAAGAACCAGATCGAGTACATCAGCAACGCC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	TRGAAINVESLLYIS EDIKSGKIKQSDFY EFKNDEIVY		ACCAACGGCACCAGGGGCGCCGCCATCAACGTGGAGAGCCTGC TGTACATCAGCGAGGACATCAAGAGCGGCAAGATCAAGCAGAG CGACTTCTACAGCGAGTTCAAGAACGACGAGATCGTGTAC
43	SQGDKAREQLKAKF LAKTNLLPRYVELL DIAYDSKRNRDFEM VTAELEFNAYLLPA VHLGGVRKPDALVA TKKFGIIVDTKAYAN GYSRNANQADEMA RYITENQKRDPKTNP NRWWDNFDARIPP N A Y Y F L W V S S F T G Q F D D Q L S Y T A H R T N T H G G A L N V E Q L L I G A N M I Q T G Q L D R N K L P E Y M Q D K E I T F	124	AGCCAGGGCGACAAGGCCAGGGAGCAGCTGAAGGCCAAGTTCC TGGCCAAGACCAACCTGCTGCCAGGTACGTGGAGCTGCTGGAC ATCGCCTACGACAGCAAGAGGAACAGGGACTTCGAGATGGTGA CCGCCGAGCTGTTCAACTTCGCCTACCTGCTGCCCGCCGTGCACC TGGGCGGCGTGAGGAAGCCCGACGCCCTGGTGGCCACCAAGAA GTTCGGCATCATCGTGGACACCAAGGCCTACGCCAACGGCTACA GCAGGAACGCCAACCAGGCCGACGAGATGGCCAGGTACATCAC CGAGAACCAGAAGAGGGACCCCAAGACCAACCCCAACAGGTGG TGGGACAACTTCGACGCCAGGATCCCCCAACGCCTACTACTT CCTGTGGGTGAGCAGCTTCTTCACCGGCCAGTTCGACGACCAGC TGAGCTACACCGCCACAGGACCAACACCCACGGCGGCGCCCTG AACGTGGAGCAGCTGCTGATCGGCGCCAACATGATCCAGACCG GCCAGCTGGACAGGAACAAGCTGCCCGAGTACATGCAGGACAA GGAGATCACCTTC
44	KVQKSNILDVIEKCR EKINNIPHEYLALIP MSFDENESTMFEIKT IELL TEHCKFDGLHC GGASKPDGLIYSED YGVIIIDTKSYKDGFN IQTPERDKMKRYIEE NQNRNPQHNKTRW WDEFPHNISNLFRLF VSGKFGGNFKEQLRI LSEQTNNTLGGALSS YVLLNIAEQIAINKID HCDFKTRISCLDEVA	125	AAGGTGCAGAAGAGCAACATCCTGGACGTGATCGAGAAGTGCA GGGAGAAGATCAACAACATCCCCACGAGTACCTGGCCCTGATC CCCATGAGCTTCGACGAGAACGAGAGCACCATGTTTCGAGATCAA GACCATCGAGCTGCTGACCGAGCACTGCAAGTTCGACGGCCTGC ACTGCGGCGGCGCCAGCAAGCCCGACGGCCTGATCTACAGCGA GGACTACGGCGTGATCATCGACACCAAGAGCTACAAGGACGGC TTCAACATCCAGACCCCCGAGAGGGACAAGATGAAGAGGTACA TCGAGGAGAACCAGAACAGGAACCCCAAGCACAACAAGACCAG GTGGTGGGACGAGTTCCCCACAACATCAGCAACTTCCTGTTC TGTTCTGTGAGCGGCAAGTTCGGCGGCAACTTCAAGGAGCAGCTG AGGATCCTGAGCGAGCAGACCAACAACACCCTGGGCGGCGCCC TGAGCAGCTACGTGCTGCTGAACATCGCCGAGCAGATCGCCATC AACAAGATCGACCACTGCGACTTCAAGACCAGGATCAGCTGCCT GGACGAGGTGGCC
45	VPVKSEVSLCKDYL RSYLTHVDHKYLILL DLGFDGTSRDRYEI QTAQLLTAELDFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPLIKQADEIYR YIEENKKRDEKLN NKWWEIFDKGVVR YHFAFVSGAFTGGF KERLDNIRMRSIGC A A I N S M N L L M A E E L K S G R L G Y E E C F A L F D C N D E I T F	126	GTGCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGAGCTACCTGACCCACGTGGACCACAAGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCCAGCTGCTGACCGCCGAGCTGGACTTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCATCAAGCAGGCCGACGAGATCTACAGGTACA TCGAGGAGAACAAGAAGAGGGACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAAGGGCGTGGTGAAGTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGCTG GACAACATCAGGATGAGGAGCGGCATCTGCGGCGCCGCCATCA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACGAGGAGTGCTTCGCCCTGTTTCGACTGCAACG ACGAGATCACCTTC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
46	VPVKSEVSLCKDYL RSHLNHVDHRYLIL LDLGFDTSDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPLIKQADEMY RYIEENKERNEKLN NKWWEIFDKDVIHY HFAFVSGAFTGGFK ERLENIRMRSIYGA AVNSMNNLLMAEEL KSGRLDYKECFKLF DCNDEIVL	127	GTGCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGAGCCACCTGAACCACGTGGACCACAGGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAAGGACGTGATCCACTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGCTG GAGAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCCGTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGACTACAAGGAGTGCTTCAAGCTGTTCGACTGCAACG ACGAGATCGTGCTG
47	VPVKSEVSLCKDYL RSHLVHVDHKYLV LDLGFDTSDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPLIKQADEMY RYIEENKERNEKLN NKWWEIFGNDVIHY HFAFVSGAFTGGFK ERLDNIRMRSIYGA AVNSMNNLLLAEEEL KSGRLGYKECFKLF DCNDEIVL	128	GTGCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGAGCCACCTGGTGCACGTGGACCACAAGTACCTGGTGTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGGCAACGACGTGATCCACTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGCTG GACAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCCGTGA ACAGCATGAACCTGCTGCTGCTGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACAAGGAGTGCTTCAAGCTGTTCGACTGCAACG ACGAGATCGTGCTG
48	ECVKDNVVDIKDRV RNKLIHLHDHXYLALI DLAYSDAASRAKKN ADAREFEIQTADLFT KELSFNGQRLGDSR KPDVHSYGLDGTIV DNKSYKDGFNISRT CADEMSRYINENNL RQKSLNPNEWWKN FDSTITAYTFLFITSY LKGQFEDQLEYVSN ANGGIKGAAGVESL LYLSEGIKAGRISHA DFYSNFFNNKEMIY	129	GAGTGCGTGAAGGACAACGTGGTGGACATCAAGGACAGGGTGA GGAACAAGCTGATCCACCTGGACCACAAGTACCTGGCCCTGATC GACCTGGCCTACAGCGACGCCGCCAGCAGGGCCAAGAAGAAGC CCGACGCCAGGGAGTTCGAGATCCAGACCGCCGACCTGTTCACC AAGGAGCTGAGCTTCAACGGCCAGAGGCTGGGCGACAGCAGGA AGCCCGACGTGATCATCAGCTACGGCCTGGACGGCACCATCGTG GACAACAAGAGCTACAAGGACGGCTTCAACATCAGCAGGACCT GCGCCGACGAGATGAGCAGGTACATCAACGAGAACAACCTGAG GCAGAAGAGCCTGAACCCCAACGAGTGGTGGAAGAAGTTCGAC AGCACCATACCGCCTACACCTTCCTGTTTCATCACCAGCTACCTG AAGGGCCAGTTCGAGGACCAGCTGGAGTACGTGAGCAACGCCA ACGGCGGCATCAAGGGCGCCGCCATCGGCGTGGAGAGCCTGCT GTACCTGAGCGAGGGCATCAAGGCCGGCAGGATCAGCCACGCC GACTTCTACAGCAACTTCAACAACAAGGAGATGATCTAC
49	IAKSDFSIIKDNIRRK LOYYNHKYLILLIDL GFDSDSNRDYEIQTA ELLTTELAFKGARL GDTRKPDVCVYYGE NGLIIDNKAYS KGYS LPMSQADEMVRYIE ENKARQSSINPNQW WKIFEDTVCFNFYA FVSGEFTGGFKDRL NNICERTRVSGGAIN TINLLLLAEELKSGR	130	ATCGCCAAGAGCGACTTCAGCATCATCAAGGACAACATCAGGA GGAAGCTGCAGTACGTGAACCACAAGTACCTGCTGCTGATCGAC CTGGGCTTCGACAGCGACAGCAACAGGGACTACGAGATCCAGA CCGCCGAGCTGCTGACCACCGAGCTGGCCTTCAAGGGCGCCAGG CTGGGCGACACCAGGAAGCCCCGACGTGTGCGTGTACTACGGCG AGAACGGCCTGATCATCGACAACAAGGCCTACAGCAAGGGCTA CAGCCTGCCCATGAGCCAGGCCGACGAGATGGTGAGGTACATC GAGGAGAACAAGGCCAGGCAGAGCAGCATCAACCCCAACCAAGT GGTGGAAGATCTTCGAGGACACCGTGTGCAACTTCAACTACGCC TTCGTGAGCGGCGAGTTACCGGCGGCTTCAAGGACAGGCTGAA CAACATCTGCGAGAGGACCAGGGTGAGCGGCGGCGCCATCAAC ACCATCAACCTGCTGCTGCTGGCCGAGGAGCTGAAGAGCGGCA

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	MSYPKCFSYFDTND EVHI		GGATGAGCTACCCCAAGTGCTTCAGCTACTTCGACACCAACGAC GAGGTGCACATC
50	LKYLGIKKQNRAFEI ITAELFNTSYKLSAT HLGGRRPDVLYN DNFGIIVDTKAYKD GYGRNVNQEDMV RYITENNIRKQDINK NDWWKYFSKSIPST SYYHLWISSQFVGM FSDQLRETSSRTGEN GGAMNVEQLLIGAN QVLNNVLDPNCLPK YMENKEIIF	131	CTGAAGTACCTGGGCATCAAGAAGCAGAACAGGGCCTTCGAGA TCATCACCGCCGAGCTGTTCAACACCAGCTACAAGCTGAGCGCC ACCCACCTGGGCGGCGGCAGGAGGCCCGACGTGCTGGTGTACA ACGACAACCTCGGCATCATCGTGGACACCAAGGCCTACAAGGAC GGCTACGGCAGGAACGTGAACCAGGAGGACGAGATGGTGAGGT ACATCACCGAGAACAACATCAGGAAGCAGGACATCAACAAGAA CGACTGGTGGAAAGTACTTCAGCAAGAGCATCCCCAGCACCAGCT ACTACCACCTGTGGATCAGCAGCCAGTTCGTGGGCATGTTTCAGC GACCAGCTGAGGGAGACCAGCAGCAGGACCGGCGAGAACGGCG GCGCCATGAACGTGGAGCAGCTGCTGATCGGCGCCAACCAGGT GCTGAACAACGTGCTGGACCCCAACTGCCTGCCCAAGTACATGG AGAACAAGGAGATCATCTTC
51	VPVKSEVSLCKDYL RSHLNHVDHLYLIL LDLGFDTSDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPIKQADEMY RYIEENKERNEKLN NKWWEIFDKDVIHY HFAFVSGAFTGGFR ERLENIRMRSIYGA AVNSMNNLLMAEEL KSGRLGYKECFKLF DCNDEIVL	132	GTGCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGAGCCACCTGAACCACGTGGACCACAAGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAAGGACGTGATCCACTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAGGGAGAGGGCTG GAGAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCGGTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACAAGGAGTGCTTCAAGCTGTTTCGACTGCAACG ACGAGATCGTGCTG
52	VPVKSEVSLCKDYL RTHLLHVDHRYLILL DLGFDGTSDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPIKQADEMY RYIEENKERNEKLN NKWWEIFDNDVIHY HFAFISGAFTGGFKE RLDNIRMRSIYGA AVNSMNNLLMAEEL KSGRLGYKECFKLF DCNDEIVL	133	GTGCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGACCCACCTGCTGCACGTGGACCACAGGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAACGACGTGATCCACTACCACTTCG CCTTCATCAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGGCTG GACAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCGGTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACAAGGAGTGCTTCAAGCTGTTTCGACTGCAACG ACGAGATCGTGCTG
53	VPVKSEVSLCKDYL RSHLNHVDHLYLIL LDLGFDTSDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPIKQADEMY RYIEENKERNEKLN NKWWEIFDNDVIHY HFAFVSGAFTGGFR ERLENIRMRSIYGA AVNSMNNLLMAEEL	134	GTGCCCCGTGAAGAGCGAGGTGAGCCTGTGCAAGGACTACCTGA GGAGCCACCTGAACCACGTGGACCACAAGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCACCAGCGACAGGGACTACGAGATCC AGACCGCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAACGACGTGATCCACTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGGCTG GAGAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCGGTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	KSGRLGYKECFKLF DCNDEIVL		CAGGCTGGGCTACAAGGAGTGCTTCAAGCTGTTCGACTGCAACG ACGAGATCGTGCTG
54	VPVKSEMSLLKDYL RTHLLHVDHRYLILL DLGFDGASDRDYEI QTAQLLTGELNFKG ARLGDTRKPDVCVY YGEDGLIIDNKAYG KGYSPIKQADEMY RYIEENKERNEKLN NKWWEIFDNDVIHY HFAFVSGAFTGGFK ERLDNIRMRSIGYGA AVNSMNNLLMAEEL KSGRLGYKECFKLF DCNDEIVL	135	GTGCCCCGTGAAGAGCGAGATGAGCCTGCTGAAGGACTACCTGA GGACCCACCTGCTGCACGTGGACCACAGGTACCTGATCCTGCTG GACCTGGGCTTCGACGGCGCCAGCGACAGGGACTACGAGATCC AGACCGCCCAGCTGCTGACCGGCGAGCTGAACCTCAAGGGCGCC AGGCTGGGCGACACCAGGAAGCCCGACGTGTGCGTGTACTACG GCGAGGACGGCCTGATCATCGACAACAAGGCCTACGGCAAGGG CTACAGCCTGCCCATCAAGCAGGCCGACGAGATGTACAGGTACA TCGAGGAGAACAAGGAGAGGAACGAGAAGCTGAACCCCAACAA GTGGTGGGAGATCTTCGACAACGACGTGATCCACTACCACTTCG CCTTCGTGAGCGGCGCCTTCACCGGCGGCTTCAAGGAGAGGGCTG GACAACATCAGGATGAGGAGCGGCATCTACGGCGCCGCGCTGA ACAGCATGAACCTGCTGCTGATGGCCGAGGAGCTGAAGAGCGG CAGGCTGGGCTACAAGGAGTGCTTCAAGCTGTTCGACTGCAACG ACGAGATCGTGCTG
55	ILVDKEREMRKAKF LKETVLDSKFISLLD LAADATKSRDFEIVT AELFKEAYNLNSVL LGGSNKPDLGVFTD DFGILLDTKAYKNG FSIYAKDRDQMIRY VDDNNKRDKIRNPN EWWKSFSPILPNDKF YYLWVSNFFKGQFK NQIEYVNRETNTYG AVLNVEQLLYGADA VIKGIINPNKLHEYFS NDEIKF	136	ATCCTGGTGGACAAGGAGAGGGAGATGAGGAAGGCCAAGTTCC TGAAGGAGACCGTGCTGGACAGCAAGTTCATCAGCCTGCTGGAC CTGGCCCGCGACGCCACCAAGAGCAGGGACTTCGAGATCGTGA CCGCCGAGCTGTTCAAGGAGGCCTACAACCTGAACAGCGTGCTG CTGGGCGGCAGCAACAAGCCCGACGGCCTGGTGTTCACCGACG ACTTCGGCATCCTGCTGGACACCAAGGCCTACAAGAACGGCTTC AGCATCTACGCCAAGGACAGGGACCAGATGATCAGGTACGTGG ACGACAACAACAAGAGGGACAAGATCAGGAACCCCAACGAGTG GTGGAAGAGCTTCAGCCCCCTGATCCCCAACGACAAGTTCTACT ACCTGTGGGTGAGCAACTTCTTCAAGGGCCAGTTCAAGAACCAG ATCGAGTACGTGAACAGGGAGACCAACACCTACGGCGCCGCTGC TGAACGTGGAGCAGCTGCTGTACGGCGCCGACGCCGTGATCAAG GGCATCATCAACCCCAACAAGCTGCACGAGTACTTCAGCAACGA CGAGATCAAGTTC
56	TVDEKERLELKEYFI SNTRIPSKYITLLDLA YDGNANRDFEIVTA ELFKDIFKLQSKHM GGTRKPDILIWTDKF GVIADTKAYSKGYK KNISEADKMVRYVN ENTNRNKVDNTNE WWNSFDSRIPKDAY YFLWISSEFVGKFDE QLTETSSRTGRNGAS INVYQLLRGADLVQ KSKFNIHDLPLNMQ NNEIKF	137	ACCGTGGACGAGAAGGAGAGGCTGGAGCTGAAGGAGTACTTCA TCAGCAACACCAGGATCCCCAGCAAGTACATCACCTGCTGGAC CTGGCCTACGACGGCAACGCCAACAGGGACTTCGAGATCGTGAC CGCCGAGCTGTTCAAGGACATCTTCAAGCTGCAGAGCAAGCACA TGGGCGGCACCAGGAAGCCCGACATCCTGATCTGGACCGACAA GTTTCGGCGTGATCGCCGACACCAAGGCCTACAGCAAGGGCTACA AGAAGAACATCAGCGAGGCCGACAAGATGGTGAGGTACGTGAA CGAGAACACCAACAGGAACAAGGTGGACAACACCAACGAGTGG TGGAACAGCTTCGACAGCAGGATCCCCAAGGACGCCTACTACTT CCTGTGGATCAGCAGCGAGTTCGTGGGCAAGTTCGACGAGCAGC TGACCGAGACCAGCAGCAGGACCGGCAGGAACGGCGCCGAGCAT CAACGTGTACCAGCTGCTGAGGGGCGCCGACCTGGTGCAGAAG AGCAAGTTCAACATCCACGACCTGCCCAACCTGATGCAGAACAA CGAGATCAAGTTC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
57	TLQKSDIEKFKNQLR TELTNIDHSYLKGIDI ASKKTTTNVENTEF EAISTKVFTDELGFF GEHLGGSNKPDGLI WDNDCAIILDSKAY SEGFLPTASHTDAM GRYLROFKERKEEIK PTWWDIAPDNLANT YFAYVSGSFSGNYK AQLQKFRQDTNHM GGALEFVKLLLLAN NYKAHKMSINEVKE SILDYNISY	138	ACCCTGCAGAAGAGCGACATCGAGAAGTTCAAGAACCAGCTGA GGACCGAGCTGACCAACATCGACCACAGCTACCTGAAGGGCAT CGACATCGCCAGCAAGAAGACCACCACCAACGTGGAGAACACC GAGTTCGAGGCCATCAGCACCAAGGTGTTACCGACGAGCTGGG CTTCTTCGGCGAGCACCTGGGCGGCAGCAACAAGCCGACGGCC TGATCTGGGACAACGACTGCGCCATCATCTGGACAGCAAGGCC TACAGCGAGGGCTTCCCCCTGACCGCCAGCCACACCGACGCCAT GGGCAGGTACCTGAGGCAGTTCAAGGAGAGGAAGGAGGAGATC AAGCCCACCTGGTGGGACATCGCCCCGACAACCTGGCCAACAC CTACTTCGCCTACGTGAGCGGCAGCTTCAGCGGCAACTACAAGG CCCAGCTGCAGAAGTTCAGGCAGGACACCAACCACATGGGCGG CGCCCTGGAGTTCGTGAAGCTGCTGCTGCTGGCCAACAACACTACA AGGCCACAAGATGAGCATCAACGAGGTGAAGGAGAGCATCCT GGACTACAACATCAGCTAC
58	VKEKTDAAALVKERV RLQLHNINHKYLALI DYAFSGKNNSRDFE VYTIDLLVNELTFGG LHLGGTRKPDGIFY HGSNGIIIDNKAYAK GFVITRNMADDEMIR YVQENNDRNPERNP NCWWKGFPHDVTR YNYVFISSMFKGEV EHMLDNIRQSTGIDG CVLTIENTLLYYADAI KGGTLSKATFINGFN ANKEMVF	139	GTGAAGGAGAAGACCGACGCCGCCCTGGTGAAGGAGAGGGTGA GGCTGCAGCTGCACAACATCAACCACAAGTACCTGGCCCTGATC GACTACGCCTTCAGCGGCAAGAACAACAGCAGGGACTTCGAGG TGTACACCATCGACCTGCTGGTGAACGAGCTGACCTTCGGCGGC CTGCACCTGGGCGGCACCAGGAAGCCCGACGGCATCTTCTACCA CGGCAGCAACGGGCATCATCATCGACAACAAGGCCTACGCCAAG GGCTTCGTGATCACCAGGAACATGGCCGACGAGATGATCAGGTA CGTGCAGGAGAACAACGACAGGAACCCCGAGAGGAACCCCAAC TGCTGGTGGAAGGGCTTCCCCACGACGTGACCAGGTACAACATA CGTGTTTCATCAGCAGCATGTTCAAGGGCGAGGTGGAGCACATGC TGGACAACATCAGGCAGAGCACCGGCATCGACGGCTGCGTGCT GACCATCGAGAACCTGCTGTACTACGCCGACGCCATCAAGGGCG GCACCCTGAGCAAGGCCACCTTCATCAACGGCTTCAACGCCAAC AAGGAGATGGTGTTC
59	VKETDSDVIKDRVR LKLHHVNHKYLTLI DYAFSGKNNCMDFE VYTIDLLVNELAFN GVHLGGTRKPDGIF YHNRNGIIIDNKAYS HGFTLSRAMADEMI RYIQENNDRNPERN PNKWENFDKGVN QFNVFISSLFKGEIE HMLTNIKQSTDGVE GCVLSAENLLYFAE AMKSGVMPKTEFIS YFGAGKEIQF	140	GTGAAGGAGACCACCGACAGCGTGATCATCAAGGACAGGGTGA GGCTGAAGCTGCACCACGTGAACCACAAGTACCTGACCCTGATC GACTACGCCTTCAGCGGCAAGAACAACACTGCATGGACTTCGAGGT GTACACCATCGACCTGCTGGTGAACGAGCTGGCCTTCAACGGCG TGCACCTGGGCGGCACCAGGAAGCCCGACGGCATCTTCTACCAC AACAGGAACGGGCATCATCATCGACAACAAGGCCTACAGCCACG GCTTCACCCTGAGCAGGGCCATGGCCGACGAGATGATCAGGTAC ATCCAGGAGAACAACGACAGGAACCCCGAGAGGAACCCCAACA AGTGGTGGGAGAACTTCGACAAGGGCGTGAACAGTTCAACTTC GTGTTTCATCAGCAGCCTGTTCAAGGGCGAGATCGAGCACATGCT GACCAACATCAAGCAGAGCACCGACGGCGTGGAGGGCTGCGTG CTGAGCGCCGAGAACCTGCTGTACTTCGCCGAGGCCATGAAGAG CGGCGTGATGCCCAAGACCGAGTTCATCAGCTACTTCGGCGCCG GCAAGGAGATCCAGTTC
60	SACKADITELKDKIR KSLKVLDPKYLVLV DLAYSDASTKSKKN SDAREFEIQTADLFT KELKFDGMRGDSN RPDVIIISHDNFGTIID NKSYPDGFNIDKKC ADEMSRYINENQRI PELPKNEWKWNFD VNVDIFTFLFITSYLK GNFKDQLEYISKSQS DIKGAAISVEHLLYI	141	AGCGCTGCAAGGCCGACATCACCGAGCTGAAGGACAAGATCA GGAAGAGCCTGAAGGTGCTGGACCACAAGTACCTGGTGTGCTGGT GGACCTGGCCTACAGCGACGCCAGCACCAAGAGCAAGAAGAAC AGCGACGCCAGGGAGTTCGAGATCCAGACCGCCGACCTGTTTAC CAAGGAGCTGAAGTTCGACGGCATGAGGCTGGGCGACAGCAAC AGGCCCCGACGTGATCATCAGCCACGACAACCTTCGGCACCATCAT CGACAACAAGAGCTACAAGGACGGCTTCAACATCGACAAGAAG TGCGCCGACGAGATGAGCAGGTACATCAACGAGAACCAGAGGA GGATCCCCGAGCTGCCCAAGAACGAGTGGTGAAGAAGTTCGA CGTGAACGTGGACATCTTCACTTCTGTTTCATCACCAGCTACCT GAAGGGCAACTTCAAGGACCAGCTGGAGTACATCAGCAAGAGC CAGAGCGACATCAAGGGCGCCGCCATCAGCGTGGAGCACCTGC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	SEKVKNQSMKADF FKLFNNDIEIRV		TGTACATCAGCGAGAAGGTGAAGAACGGCAGCATGGACAAGGC CGACTTCTTCAAGCTGTTCAACAACGACGAGATCAGGGTG
61	VLKDKHLEKIKEKF LENTSLDPRFISLIEIS RDKKQNRAFEIITAE LFNTSYNLSAHLGG GRRPDVLAAYNDNFG IIVDTKAYKNGYGR NVNQEDDEMVRYTE NKIRKQDISKNNWW KYFSKSIPSTSYHL WISSEFVGMFSDQL RETSSRTGENGGAM NVEQLLIGANQVLN NVLDPNRLPEYME KEIIF	142	GTGCTGAAGGACAAGCACCTGGAGAAGATCAAGGAGAAGTTCC TGGAGAACACCAGCCTGGACCCCAGGTTTCATCAGCCTGATCGAG ATCAGCAGGGACAAGAAGCAGAACAGGGCCTTCGAGATCATCA CCGCCGAGCTGTTCAACACCAGCTACAACCTGAGCGCCATCCAC CTGGGCGGCGGCAGGAGGCCCCGACGTGCTGGCCTACAACGACA ACTTCGGCATCATCGTGGACACCAAGGCCTACAAGAACGGCTAC GGCAGGAACGTGAACCAGGAGGACGAGATGGTGAGGTACATCA CCGAGAACAAGATCAGGAAGCAGGACATCAGCAAGAACAACCTG GTGGAAGTACTTCAGCAAGAGCATCCCCAGCACCAGCTACTACC ACCTGTGGATCAGCAGCGAGTTCGTGGGCATGTTTCAGCGACCAG CTGAGGGAGACCAGCAGCAGGACCGGCGAGAACGGCGGCGCCA TGAACGTGGAGCAGCTGCTGATCGGCGCCAACCAAGGTGCTGAAC AACGTGCTGGACCCCAACAGGCTGCCCCGAGTACATGGAGAACA AGGAGATCATCTTC
62	ALKDKHLEKIKEKF LENTSLDPRFISLIEIS RDKKQNRAFEIITAE LFNTSYKLSATHLG GGRRPDVLVYNDNF GIIIVDTKAYKDG RNVNQEDDEMVRIT ENNIRKQDINKNDW WKYFSKSIPSTSYH LWISSQFVGMFSDQ LRETSSRTGENGGA MNVEQLLIGANQVL NNVLDPNCLPKYME NKEIIF	143	GCCCTGAAGGACAAGCACCTGGAGAAGATCAAGGAGAAGTTCC TGGAGAACACCAGCCTGGACCCCAGGTTTCATCAGCCTGATCGAG ATCAGCAGGGACAAGAAGCAGAACAGGGCCTTCGAGATCATCA CCGCCGAGCTGTTCAACACCAGCTACAAGCTGAGCGCCACCCAC CTGGGCGGCGGCAGGAGGCCCCGACGTGCTGGTGTACAACGACA ACTTCGGCATCATCGTGGACACCAAGGCCTACAAGGACGGCTAC GGCAGGAACGTGAACCAGGAGGACGAGATGGTGAGGTACATCA CCGAGAACAACATCAGGAAGCAGGACATCAACAAGAACGACTG GTGGAAGTACTTCAGCAAGAGCATCCCCAGCACCAGCTACTACC ACCTGTGGATCAGCAGCCAGTTCGTGGGCATGTTTCAGCGACCAG CTGAGGGAGACCAGCAGCAGGACCGGCGAGAACGGCGGCGCCA TGAACGTGGAGCAGCTGCTGATCGGCGCCAACCAAGGTGCTGAAC AACGTGCTGGACCCCAACTGCCTGCCCCAAGTACATGGAGAACAA GGAGATCATCTTC
63	VLEKSDIEKFKNQLR TELTNIDHSYLKGIDI ASKKTSNVENTEF EAISTKIFTDELGFSG KHLGGSNKPDLW DDDCAILDSKAYSE GFPLTASHTDAMGR YLRQFTERKEEIKPT WWDIAPEHLDNITYF AYVSGSFGNYKEQ LQKFRQDTNHLGGA LEFVKLLLLANNYK TQKMSKKEVKKSIL DYNISY	144	GTGCTGGAGAAGAGCGACATCGAGAAGTTCAAGAACCAGCTGA GGACCGAGCTGACCAACATCGACCACAGCTACCTGAAGGGCAT CGACATCGCCAGCAAGAAGAAGACCAGCAACGTGGAGAACCACC GAGTTCGAGGCCATCAGCACCAAGATCTTCACCGACGAGCTGGG CTTCAGCGGCAAGCACCTGGGCGGCAGCAACAAGCCCCGACGGC CTGCTGTGGGACGACGACTGCGCCATCATCCTGGACAGCAAGGC CTACAGCGAGGGCTTCCCCCTGACCGCCAGCCACACCGACGCCA TGGGCAGGTACCTGAGGACGTTACCGAGAGGAAGGAGGAGAT CAAGCCCACCTGGTGGGACATCGCCCCGAGCAGCTGGACAACA CCTACTTCGCCTACGTGAGCGGCAGCTTCAGCGGCAACTACAAG GAGCAGCTGCAGAAGTTCAGGCAGGACACCAACCACCTGGGCG GCGCCCTGGAGTTCGTGAAGCTGCTGCTGCTGGCCAACAACCTAC AAGACCCAGAAGATGAGCAAGAAGGAGGTGAAGAAGAGCATCC TGGACTACAACATCAGCTAC
64	AEADVTSEKIKNHF RRVTELPERYLELLD IAFDHKRNRDFEMV TAGLFKDVYGLSEV HLGGANKPDGVVY NDNFGIILDTKAYEN GYGKHISQIDEMVR YIDNRLRDTTRNP NKWWENFDADIPSD QFYYLWVSGKFLPN FAEQLKQTNYSRSHA	145	GCCGAGGCCGACGTGACCAGCGAGAAGATCAAGAACCACCTTCA GGAGGGTGACCGAGCTGCCCCGAGAGGTACCTGGAGCTGCTGGA CATCGCCTTCGACCACAAGAGGAACAGGGACTTCGAGATGGTG ACCGCCGGCCTGTTCAAGGACGTGTACGGCCTGGAGAGCGTGCA CCTGGGCGGCGCCAACAAGCCCCGACGGCGTGGTGTACAACGAC AACTTCGGCATCATCCTGGACACCAAGGCCTACGAGAACGGCTA CGGCAAGCACATCAGCCAGATCGACGAGATGGTGAGGTACATC GACGACAACAGGCTGAGGGACACCACCAGGAACCCCAACAAGT GGTGGGAGAACTTCGACGCCGACATCCCCAGCGACCAAGTTCTAC TACCTGTGGGTGAGCGGCAAGTTCCTGCCCAACTTCGCCGAGCA GCTGAAGCAGACCAACTACAGGAGCCACGCCAACGGCGGCGGC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	NGGGLEVVQQLLLGA DAVKRRKLDVNTIP NYMKNEVITL		CTGGAGGTGCAGCAGCTGCTGCTGGGCGCCGACGCCGTGAAGA GGAGGAAGCTGGACGTGAACACCATCCCCAACTACATGAAGAA CGAGGTGATCACCCCTG
65	AEADLNSEKIKNH RKITNLPEKYIELLDI AFDHRRHQDFEIVT AGLFKDCYGLSSIHL GGQNKPDGVVFN KFGIILDTKAYEKG GMHIGQIDEMCRYI DDNKKRDIVRQPN WWKNFGDNIPKDQF YYLWISGKFLPRFNE QLKQTHYRTSINGG GLEVSQQLLGANAA MKGKLDVNTLPKH MNNQVIKL	146	GCCGAGGCCGACCTGAACAGCGAGAAGATCAAGAACCCTACA GGAAGATCACCAACCTGCCCCGAGAAGTACATCGAGCTGCTGGA CATCGCCTTCGACCACAGGAGGCACCAGGACTTCGAGATCGTGA CCGCCGGCCTGTTCAAGGACTGCTACGGCCTGAGCAGCATCCAC CTGGGCGGCCAGAACAAGCCCCGACGGCGTGGTGTTCACAACA AGTTCGGCATCATCTGGACACCAAGGCCTACGAGAAGGGCTAC GGCATGCACATCGGCCAGATCGACGAGATGTGCAGGTACATCG ACGACAACAAGAAGAGGGGACATCGTGAGGCAGCCCAACGAGTG GTGGAAGAAGTTCGGCGACAACATCCCCAAGGACAGTTCTACT ACCTGTGGATCAGCGGCAAGTTCCTGCCAGGTTCAACGAGCAG CTGAAGCAGACCCACTACAGGACCAGCATCAACGGCGGCGGCC TGGAGGTGAGCCAGCTGCTGCTGGGCGCCAACGCCGCCATGAA GGGCAAGCTGGACGTGAACACCTGCCCAAGCATGAACAAC CAGGTGATCAAGCTG
66	VLKDAALQKTKNTL LNELTEIDPADIEVIE MSWKKATTRSQNTL EATLFEVKVVEIFKK YFELNGEHLGGQNR PDGAVYYNSTYGIIL DTKAYSNGYNIPVD QQREMVDYITDVID KNQNVTPNRWWEA FPATLLKNNIYYLW VAGGFTGKYLDQLT RTHNQTNMDGGAM TTEVLLRLANKVSS GNLKTDDIPKLMTN KLILS	147	GTGCTGAAGGACGCCGCCCTGCAGAAGACCAAGAACACCCTGC TGAACGAGCTGACCGAGATCGACCCCGCCGACATCGAGGTGATC GAGATGAGCTGGAAGAAGGCCACCACCAGGAGCCAGAACACCC TGGAGGCCACCCTGTTTCGAGGTGAAGGTGGTGGAGATCTTCAAG AAGTACTTCGAGCTGAACGGCGAGCACCTGGGCGGCCAGAACA GGCCCGACGGCGCCGTGTACTACAACAGCACCTACGGCATCATC CTGGACACCAAGGCCTACAGCAACGGCTACAACATCCCCGTGGA CCAGCAGAGGGAGATGGTGGACTACATCACCGACGTGATCGAC AAGAACCAGAACGTGACCCCAACAGGTGGTGGGAGGCCTTCC CCGCCACCCTGCTGAAGAACAACATCTACTACCTGTGGGTGGCC GGCGGCTTACCGGCAAGTACCTGGACCAGCTGACCAGGACCCA CAACCAGACCAACATGGACGGCGGCCATGACCACCGAGGTG CTGCTGAGGCTGGCCAACAAGGTGAGCAGCGGCAACCTGAAGA CCACCGACATCCCCAAGCTGATGACCAACAAGCTGATCCTGAGC
67	AEADLDSEKIKNH RKITNLPEKYIELLDI AFDHRRHQDFEITA GLFKDCYGLSSIHLG GQNKPDGVVFNKGF GIILDTKAYEKG MHINQIDEMCRYIED NKQRDKIRQPN WNNFGDNIPENKFY YLWVSGKFLPKFNE QLKQTHYRTGINGG GLEVSQQLLGADAV MKGALNVNLPYTM HNNVIQ	148	GCCGAGGCCGACCTGGACAGCGAGAGGATCAAGAACCCTACA GGAAGATCACCAACCTGCCCCGAGAAGTACATCGAGCTGCTGGA CATCGCCTTCGACCACACAGGACACCAGGACTTCGAGATCATCA CCGCCGGCCTGTTCAAGGACTGCTACGGCCTGAGCAGCATCCAC CTGGGCGGCCAGAACAAGCCCCGACGGCGTGGTGTTCACGGCA AGTTCGGCATCATCTGGACACCAAGGCCTACGAGAAGGGCTAC GGCATGCACATCAACCAGATCGACGAGATGTGCAGGTACATCG AGGACAACAAGCAGAGGGACAAGATCAGGCAGCCCAACGAGTG GTGGAACAAGTTCGGCGACAACATCCCCGAGAACAAGTTCTACT ACCTGTGGGTGAGCGGCAAGTTCCTGCCAAGTTCAACGAGCAG CTGAAGCAGACCCACTACAGGACCGGCATCAACGGCGGCGGCC TGGAGGTGAGCCAGCTGCTGCTGGGCGCCGACGCCGTGATGAA GGGCGCCCTGAACGTGAACATCTGCCCACCTACATGCACAACA ACGTGATCCAG

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
68	EISDIALQKEKAYFY KNTALSKRHISILEIA FDGSKNRDLEILSAE VFKDYYQLESIHG GGLKPDGIAFNQNF GIIVDTKAYKGVYS RSRAEADKMFYIE DNKKRDPKRNQSL WWRSFNEHIPANNF YFLWISGKFQRNFD TQINQLNYETGYRG GALSARQFLIGADAI QKGKIDINDLPSYFN NSVISF	149	GAGATCAGCGACATCGCCCTGCAGAAGGAGAAGGCCTACTTCTA CAAGAACACCGCCCTGAGCAAGAGGCACATCAGCATCCTGGAG ATCGCCTTCGACGGCAGCAAGAACAGGGACCTGGAGATCCTGA GCGCCGAGGTGTTCAAGGACTACTACCAGCTGGAGAGCATCCAC CTGGGCGGCGGCCTGAAGCCCCGACGGCATCGCCTTCAACCAGAA CTTCGGCATCATCGTGGACACCAAGGCCTACAAGGGCGTGTACA GCAGGAGCAGGGCCGAGGCCGACAAGATGTTCAAGGTACATCGA GGACAACAAGAAGAGGGACCCCAAGAGGAACCAGAGCCTGTGG TGGAGGAGCTTCAACGAGCACATCCCCGCCAACAACTTCTACTT CCTGTGGATCAGCGGCAAGTTCCAGAGGAACCTTCGACACCCAGA TCAACCAGCTGAACTACGAGACCGGCTACAGGGGCGGCGCCCT GAGCGCCAGGCAGTTCTGTATCGGCGCCGACGCCATCCAGAAG GGCAAGATCGACATCAACGACCTGCCAGCTACTTCAACAACAG CGTGATCAGCTTC
69	TSREKSRLNLKEYFV SNTNLPNKFITLLDL AYDGKANRDFELIT SELFREIYKLNTRHL GGTRKPDILIWNENF GIIADTKAYSKGYK KNISEEDKMVRYIDE NIKRSKDYNPNEW KVFDNEISSNNYFYL WISSEFIGKFEEQLQ ETAQRTNVKGASIN VYQLLMGAHKVQT KELNVNSIPKYMNN TEIKF	150	ACCAGCAGGGAGAAGAGCAGGCTGAACCTGAAGGAGTACTTCG TGAGCAACACCAACCTGCCCAACAAGTTCATCACCTGCTGGAC CTGGCCTACGACGGCAAGGCCAACAGGGACTTCGAGCTGATCAC CAGCGAGCTGTTTCAGGGAGATCTACAAGCTGAACACCAGGCAC CTGGGCGGCACCAGGAAGCCCCGACATCCTGATCTGGAACGAGA ACTTCGGCATCATCGCCGACACCAAGGCCTACAGCAAGGGCTAC AAGAAGAACATCAGCGAGGAGGACAAGATGGTGAGGTACATCG ACGAGAACATCAAGAGGAGCAAGGACTACAACCCCAACGAGTG GTGGAAGGTGTTTCGACAACGAGATCAGCAGCAACAACCTACTTCT ACCTGTGGATCAGCAGCGAGTTCATCGGCAAGTTCGAGGAGCAG CTGCAGGAGACCGCCAGAGGACCAACGTGAAGGGCGCCAGCA TCAACGTGTACCAGCTGCTGATGGGCGCCACAAAGGTGCAGACC AAGGAGCTGAACGTGAACAGCATCCCCAAGTACATGAACAACA CCGAGATCAAGTTC
70	NCIKDSIIDIKDRVRT KLVLHDHLYLALID LAFSDADTRTKKNS DAREFEIQTADLFTK ELSFNGQRLGDSRK PDIIISFDKIGTIIDNK SYKDGFNISRPCADE MIRYINENNLRKKS NANEWWNKFDPTIT AYSFLFITSYLKGQF QEQLYISNANGGIK GAAIGIENLLYLSEA LKSGKISHKDFYQNF NNKEITY	151	AACTGCATCAAGGACAGCATCATCGACATCAAGGACAGGGTGA GGACCAAGCTGGTGCACCTGGACCACAAGTACCTGGCCCTGATC GACCTGGCCTTCAGCGACGCCGACACCAGGACCAAGAAGAACA GCGACGCCAGGGAGTTCGAGATCCAGACCGCCGACCTGTTTACC AAGGAGCTGAGCTTCAACGGCCAGAGGCTGGGCGACAGCAGGA AGCCCGACATCATCATCAGCTTCGACAAGATCGGCACCATCATC GACAACAAGAGCTACAAGGACGGCTTCAACATCAGCAGGCCCT GCGCCGACGAGATGATCAGGTACATCAACGAGAACAACCTGAG GAAGAAGAGCCTGAACGCCAACGAGTGGTGAACAAGTTCGAC CCCACCATCACCGCCTACAGCTTCCTGTTTATCACCAGCTACCTG AAGGGCCAGTTCAGGAGCAGCTGGAGTACATCAGCAACGCCA ACGGCGGCATCAAGGGCGCCGCCATCGGCATCGAGAACCTGCT GTACCTGAGCGAGGCCCTGAAGAGCGGCAAGATCAGCCACAAG GACTTCTACCAGAACTTCAACAACAAGGAGATCACCTAC
71	LPQKDQVQQQDEL RPMLKNVDHRYLQL VELALDSDQNSEYS QFEQLTMELVLKHL DFDGKPLGGSNKP GIAWDNDGNFIIFDT KAYNKGYSLAGNT DKVKRYIDDVRDRD TSRTSTWWQLVPKS IDVHLLRFVYVSG NFTGNMMLLLDSL SWSNAQGLASVEK LLLTSELYLRNMY	152	CTGCCCCAGAAGGACCAGGTGCAGCAGCAGGACGAGCTGA GGCCCATGCTGAAGAACGTGGACCACAGGTACCTGCAGCTGGTG GAGCTGGCCCTGGACAGCGACCAGAACAGCGAGTACAGCCAGT TCGAGCAGCTGACCATGGAGCTGGTGCTGAAGCACCTGGACTTC GACGGCAAGCCCCCTGGGCGGCAGCAACAAGCCCCGACGGCATCG CCTGGGACAACGACGGCAACTTCATCATCTTCGACACCAAGGCC TACAACAAGGGCTACAGCCTGGCCGGCAACACCGACAAGGTGA AGAGGTACATCGACGACGTGAGGGACAGGGACACCAGCAGGAC CAGCACCTGGTGGCAGCTGGTGCCCAAGAGCATCGACGTGCACA ACCTGCTGAGGTTCGTGTACGTGAGCGGCAACTTCACCGGCAAC TACATGAAGCTGCTGGACAGCCTGAGGAGCTGGAGCAACGCC AGGGCGGCCTGGCCAGCGTGGAGAAGCTGCTGCTGACCAGCGA

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
	HQELIDSWTDNNVKH		GCTGTACCTGAGGAACATGTACAGCCACCAGGAGCTGATCGACA GCTGGACCGACAACAACGTGAAGCAC
72	TTDAVVVKDRARV RLHNINHKYLTLDY AFSGKNNCTEFEIYT IDLLVNELAFNGIHL GGTRKPDGIFDYNQ QGIIDNKAYSKGFTI TRSMADDEMVRVYVQ ENNDRNPERNKTQ WWLNFQDNVNHFN FVFISMFKEVRH MLNNIKQSTGVDGC VLTAENLLYFADAI KGGTVKRTDFINLF GKNDEL	153	ACCACCGACGCCGTGGTGGTGAAGGACAGGGCCAGGGTGAGGC TGCACAACATCAACCACAAGTACCTGACCCTGATCGACTACGCC TTCAGCGGCAAGAACAACCTGCACCGAGTTCGAGATCTACACCAT CGACCTGCTGGTGAACGAGCTGGCCTTCAACGGCATCCACCTGG GCGGCACCAGGAAGCCCGACGGCATCTTCGACTACAACCAGCA GGGCATCATCATCGACAACAAGGCCTACAGCAAGGGCTTCACCA TCACCAGGAGCATGGCCGACGAGATGGTGAGGTACGTGCAGGA GAACAACGACAGGAACCCCGAGAGGAACAAGACCCAGTGGTGG CTGAACTTCGGCGACAACGTGAACCACTTCAACTTCGTGTTTAT CAGCAGCATGTTCAAGGGCGAGGTGAGGCACATGCTGAACAAC ATCAAGCAGAGCACCGGCGTGGACGGCTGCGTGCTGACCGCCG AGAACCTGCTGTACTTCGCCGACGCCATCAAGGGCGGCACCGTG AAGAGGACCGACTTCATCAACCTGTTTCGGCAAGAACGACGAGCT G
73	LPKKDNVQRQQDEL RPLLKHVDHRYLQL VELALDSSQNSEYS MLESMTMELLTHL DFDGASLGGASKPD GIAWDKLDGNFLIVD TKAYDNGYSLAGNT DKVARYIDDVRAKD PNRASTWWTQVPES LNVDDNLSFMYVSG SFTGNYQRLLKDLR ARTNARGGLTTVEK LLLTSEAYLAKSGY GHTQLLNDWTDNDI DH	154	CTGCCCCAAGAAGGACAACGTGCAGAGGCAGCAGGACGAGCTGA GGCCCCCTGCTGAAGCACGTGGACCACAGGTACCTGCAGCTGGTG GAGCTGGCCCTGGACAGCAGCCAGAACAGCGAGTACAGCATGC TGGAGAGCATGACCATGGAGCTGCTGCTGACCCACCTGGACTTC GACGGCGCCAGCCTGGGCGGCGCCAGCAAGCCCCGACGGCATCG CCTGGGACAAGGACGGCAACTTCCTGATCGTGGACACCAAGGCC TACGACAACGGCTACAGCCTGGCCGGCAACACCGACAAGGTGG CCAGGTACATCGACGACGTGAGGGCCAAGGACCCCCAACAGGGC CAGCACCTGGTGGACCCAGGTGCCCCGAGAGCCTGAACGTGGAC GACAACCTGAGCTTCATGTACGTGAGCGGCAGCTTCACCGGCAA CTACCAGAGGCTGCTGAAGGACCTGAGGGCCAGGACCAACGCC AGGGGCGGCCTGACCACCGTGGAGAAGCTGCTGCTGACCAGCG AGGCCTACCTGGCCAAGAGCGGCTACGGCCACACCCAGCTGCTG AACGACTGGACCGACGACAACATCGACCAC
74	QIKDKYLEDLKLEL YKKTNLNPKYYEM VDIAYDGKRNREFEI YTSDLMQEIYGFKT TLLGGTRKPDVVSY SDAHGYIIDTKAYA NGYRKEIKQEDEMV RYIEDNQLKDVLRN PNKWWEFCDDAEH KKEYYFLWISSKFV GEFSSQLQDTSRRTG IKGGAVNIVQLLLG AHLVYSGEISKDQF AAYMNNTEINF	155	CAGATCAAGGACAAGTACCTGGAGGACCTGAAGCTGGAGCTGT ACAAGAAGACCAACCTGCCCAACAAGTACTACGAGATGGTGGA CATCGCCTACGACGGCAAGAGGAACAGGGAGTTCGAGATCTAC ACCAGCGACCTGATGCAGGAGATCTACGGCTTCAAGACCACCCT GCTGGGCGGCACCAGGAAGCCCGACGTGGTGAGCTACAGCGAC GCCACGGCTACATCATCGACACCAAGGCCTACGCCAACGGCTA CAGGAAGGAGATCAAGCAGGAGGACGAGATGGTGAGGTACATC GAGGACAACCAGCTGAAGGACGTGCTGAGGAACCCCAACAAGT GGTGGGAGTGCTTCGACGACGCCGAGCACAAGAAGGAGTACTA CTTCCTGTGGATCAGCAGCAAGTTCGTGGGCGAGTTCAGCAGCC AGCTGCAGGACACCAGCAGGAGGACCGGCATCAAGGGCGGGCGC CGTGAACATCGTGACGTGCTGCTGGGCGCCACCTGGTGTACA GCGGCGAGATCAGCAAGGACCAAGTTCGCCGCCTACATGAACAA CACCGAGATCAACTTC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
75	MNPRNEIVIAKHLSG GNRPEIVCYHPEDKP DHGLILDSKAYKSG FTIPSGERDKMVRYI EEYITKNQLQNPNE WWKNLKGAEYPGI VGFGFISNSFLGHYR KQLDYIMRRTKIKG SSITTEHLLKTVEDV LSEKGNVIDFFKYFL E	156	ATGAACCCCAGGAACGAGATCGTGATCGCCAAGCACCTGAGCG GCGGCAACAGGCCCCGAGATCGTGTGCTACCACCCCAGGACAA GCCCCACCACGGCCTGATCCTGGACAGCAAGGCCTACAAGAGC GGCTTCACCATCCCCAGCGGCGAGAGGGACAAGATGGTGAGGT ACATCGAGGAGTACATCACCAAGAACCAGCTGCAGAACCCCAA CGAGTGGTGGAAGAACCTGAAGGGCGCCGAGTACCCCGGCATC GTGGGCTTCGGCTTCATCAGCAACAGCTTCCTGGGCCACTACAG GAAGCAGCTGGACTACATCATGAGGAGGACCAAGATCAAGGGC AGCAGCATCACACCGAGCACCTGCTGAAGACCGTGAGGAGCG TGCTGAGCGAGAAGGGCAACGTGATCGACTTCTTCAAGTACTTC CTGGAG
76	EIKNQEIIEELKQIALN KYTALPSEWVELIEI SRDKDQSTIFEMKV AELFKTCYRIKSLHL GGASKPDCLLWDDS FSVIVDAKAYKDF PFQASEKDKMVRYL RECERKDKAENATE WWNNFPPELNSNQL FFMFASSFFSSTAEEK HLESVSIASKFSGCA WDVDNLLSGANFFL QNPQATLQYHLIRV FSNKVVD	157	GAGATCAAGAACCAGGAGATCGAGGAGCTGAAGCAGATCGCCC TGAACAAGTACACCGCCCTGCCAGCGAGTGGGTGGAGCTGATC GAGATCAGCAGGGACAAGGACCAGAGCACCATCTTCGAGATGA AGGTGGCCGAGCTGTTCAAGACCTGCTACAGGATCAAGAGCCTG CACCTGGGCGGCGCCAGCAAGCCCGACTGCCTGCTGTGGGACGA CAGCTTCAGCGTGATCGTGGACGCCAAGGCCTACAAGGACGGCT TCCCCTTCAGGCCAGCGAGAAGGACAAGATGGTGAGGTACCTG AGGGAGTGCGAGAGGAAGGACAAGGCCGAGAACGCCACCGAGT GGTGGAACAACCTCCCCCCCCGAGCTGAACAGCAACCAGCTGTTT TTCATGTTTCGCCAGCAGCTTCTTCAGCAGCACCGCCGAGAAGCA CCTGGAGAGCGTGAGCATCGCCAGCAAGTTCAGCGGCTGCGCCT GGGACGTGGACAACCTGCTGAGCGGCGCCAACCTTCTTCTGCAG AACCCCCAGGCCACCCTGCAGTACCACCTGATCAGGGTGTTTCA CAACAAGGTGGTGGAC
77	LPHKDNVIKQDEL RPMKLVNHKYLQ LVELAFESSRNSEYS QFETLTMLVVKYL DFSGKSLGGAKPD GIAWDPLGNFLIFT KAYKHGYTLNNTD RVARYINDVRDKDI QRISRWWQSIPTYID VKNKLQFVYISGSFT GHYLRLLNDRSRT RAKGGLVTVEKLLL TTERYLAEADYTHK ELFDDWMDDNIEH	158	CTGCCCCACAAGGACAACGTGATCAAGCAGCAGGACGAGCTGA GGCCCATGCTGAAGCACGTGAACCACAAGTACCTGCAGCTGGTG GAGCTGGCCTTCGAGAGCAGCAGGAACAGCGAGTACAGCCAGT TCGAGACCCTGACCATGGAGCTGGTGCTGAAGTACCTGGACTTC AGCGGCAAGAGCCTGGGCGGCGCCAACAAGCCCGACGGCATCG CCTGGGACCCCCCTGGGCAACTTCCTGATCTTCGACACCAAGGCC TACAAGCACGGCTACACCCTGAGCAACAACACCGACAGGGTGG CCAGGTACATCAACGACGTGAGGGACAAGGACATCCAGAGGAT CAGCAGGTGGTGGCAGAGCATCCCCACCTACATCGACGTGAAG AACAAGCTGCAGTTCTGTGTACATCAGCGGCAGCTTACCCGGCCA CTACCTGAGGCTGCTGAACGACCTGAGGAGCAGGACCAGGGCC AAGGGCGGCCTGGTGACCGTGAGAGAAGCTGCTGCTGACCACCG AGAGGTACCTGGCCGAGGCCGACTACACCCACAAGGAGCTGTTC GACGACTGGATGGACGACAACATCGAGCAC
78	RISPSNLEQTKQQLR EELINLDHQYLDILD FSIAGNVGARQFEV RIVELLNEIIIAKHL GGNRPEIIGFNPKEN PEDCIIMDSKAYKEG FNIPANERDKMIRYV EEYNAKDNTLNNK WWKNFESPNYPTNQ VKFSFVSSFIGQFT NQLTYINNRTNVNG SAITAETLLRKVENV MNVNTEYNLNNFFE ELGSNTLVA	159	AGGATCAGCCCCAGCAACCTGGAGCAGACCAAGCAGCAGCTGA GGGAGGAGCTGATCAACCTGGACCACCAGTACCTGGACATCCTG GACTTCAGCATCGCCGGCAACGTGGGCGCCAGGCAGTTCGAGGT GAGGATCGTGAGGCTGCTGAACGAGATCATCATCGCCAAGCACC TGAGCGGCGGCAACAGGCCCGAGATCATCGGCTTCAACCCCAA GGAGAACCCCGAGGACTGCATCATCATGGACAGCAAGGCCTAC AAGGAGGGCTTCAACATCCCCGCCAACGAGAGGGACAAGATGA TCAGGTACGTGGAGGAGTACAACGCCAAGGACAACACCCTGAA CAACAACAAGTGGTGGAGAAGTTCGAGAGCCCCAACTACCCC ACCAACCAGGTGAAGTTCAGCTTCGTGAGCAGCAGCTTCATCGG CCAGTTCACCAACCAGCTGACCTACATCAACAACAGGACCAACG TGAACGGCAGCGCCATCACCGCCGAGACCCTGCTGAGGAAGGT GGAGAACGTGATGAACGTGAACACCGAGTACAACCTGAACAAC TTCTTCGAGGAGCTGGGCGAGCAACACCCTGGTGGCC

SEQ ID NO	Amino Acid Sequence	SEQ ID NO	Back Translated Nucleic Acid Sequences
79	TFDSTVADNLKLNIL PKLKELDHKYLQAI DIAYKRSNTTNHEN TLLEVLSDLFTKE MDYHGKHLGGANK PDGFVYDEETGWIL DSKAYRDGFAVTAH TTDAMGRYIDQYRD RDDKSTWWEDFPK DLPQTYFAYVSGFYI GKYQEQLQDFENRK HMKGGIEVAKLILL AEKYKENKITHDQIT LQILNDHISQ	160	ACCTTCGACAGCACCGTGGCCGACAACCTGAAGAACCTGATCCT GCCCAAGCTGAAGGAGCTGGACCACAAGTACCTGCAGGCCATC GACATCGCCTACAAGAGGAGCAACACCACCAACCACGAGAACA CCCTGCTGGAGGTGCTGAGCGCCGACCTGTTACCAAGGAGATG GACTACCACGGCAAGCACCTGGGCGCGCCAACAAGCCCCGACG GCTTCGTGTACGACGAGGAGACCGGCTGGATCCTGGACAGCAA GGCCTACAGGGACGGCTTCGCCGTGACCGCCCACACCACCGACG CCATGGGCAGGTACATCGACCAGTACAGGGACAGGGACGACAA GAGCACCTGGTGGGAGGACTTCCCCAAGGACCTGCCCCAGACCT ACTTCGCTACGTGAGCGGCTTCTACATCGGCAAGTACCAGGAG CAGCTGCAGGACTTCGAGAACAGGAAGCACATGAAGGGCGGCC TGATCGAGGTGGCCAAGCTGATCCTGCTGGCCGAGAAGTACAAG GAGAACAAGATACCCACGACCAGATCACCTGCAGATCCTGA ACGACCACATCAGCCAG
80	PLDVVEQMKAEALRP LLNHVNRLLAIIDF SYNMSRGDDKRLED YTAQIYKLISHDTHL LAGPSRPDVSVIND LGIIDS KAYKQGFNI PQAEEDKMVRYLDE SIRRDPAINPTKWE YLGASTEYVFQFVSS SFSSGASAKLRQIHR RSSIEGSIITAKNLLL LAENFLCTNTINIDL FRQNNEI	161	CCCCTGGACGTGGTGGAGCAGATGAAGGCCGAGCTGAGGCCCC TGCTGAACCACGTGAACCACAGGCTGCTGGCCATCATCGACTTC AGCTACAACATGAGCAGGGGCGACGACAAGAGGCTGGAGGACT ACACCGCCCAGATCTACAAGCTGATCAGCCACGACACCCACCTG CTGGCCGCCCCAGCAGGCCCCGACGTGGTGAGCGTGATCAACG ACCTGGGCATCATCATCGACAGCAAGGCCTACAAGCAGGGCTTC AACATCCCCCAGGCCGAGGAGGACAAGATGGTGAGGTACCTGG ACGAGAGCATCAGGAGGGACCCCGCCATCAACCCACCAAGTG GTGGGAGTACCTGGGCGCCAGCACCGAGTACGTGTTCCAGTTCTG TGAGCAGCAGCTTCAGCAGCGGCGCCAGCGCCAAGCTGAGGCA GATCCACAGGAGGAGCAGCATCGAGGGCAGCATCATCACCGCC AAGAACCTGCTGCTGCTGGCCGAGAACTTCCTGTGCACCAACAC CATCAACATCGACCTGTTCAGGCAGAACAAACGAGATC
81	QLVPSYITQTKLRLS GLINYIDHSYFDLID LGFDGRQNRLYELRI VELLNLI NSL KALHL SGGNRPEIIAYSPDV NPINGVIMDSKSYRG GFNIPNSERDKMIRY INEYNQKNPTLNSN RWWENFRAPDYPQS PLKYSFVSGNFIGQF LNQIQYILTQTGING GAITSEKLIKVNNAV LNPNISYTINNFFND LGCNRLVQ	162	CAGCTGGTGCCAGCTACATCACCCAGACCAAGCTGAGGCTGAG CGCGCTGATCAACTACATCGACCACAGCTACTTCGACCTGATCG ACCTGGGCTTCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG GATCGTGGAGGTGCTGAACCTGATCAACAGCCTGAAGGCCCTGC ACCTGAGCGGCGGCAACAGGCCCGAGATCATCGCCTACAGCCCC GACGTGAACCCCATCAACGGCGTGATCATGGACAGCAAGAGCT ACAGGGGCGGCTTCAACATCCCCAACAGCGAGAGGGACAAGAT GATCAGGTACATCAACGAGTACAACCAGAAAGAACCCACCCCTG AACAGCAACAGGTGGTGGGAGAACTTCAGGGCCCCGACTACC CCCAGAGCCCCCTGAAGTACAGCTTCGTGAGCGGCAACTTCATC GGCCAGTTCTGAACCAGATCCAGTACATCCTGACCCAGACCGG CATCAACGGCGGCGCCATCACCAGCGAGAAGCTGATCGAGAAG GTGAACGCCGTGCTGAACCCCAACATCAGCTACACCATCAACAA CTTCTTCAACGACCTGGGCTGCAACAGGCTGGTGACG

[0017] In some embodiments, an endonuclease of the present disclosure can have a sequence of
 $X_1X_2X_3X_4X_5X_6X_7X_8X_9X_{10}X_{11}X_{12}X_{13}X_{14}X_{15}X_{16}X_{17}X_{18}X_{19}X_{20}X_{21}X_{22}X_{23}X_{24}X_{25}X_{26}X_{27}X_{28}X_{29}X_{30}X_{31}$
 $X_{32}X_{33}X_{34}X_{35}X_{36}X_{37}X_{38}X_{39}X_{40}X_{41}X_{42}X_{43}KX_{44}X_{45}X_{46}X_{47}X_{48}X_{49}X_{50}X_{51}X_{52}X_{53}X_{54}X_{55}GX_{56}HLGGX_{57}$
 $RX_{58}PDGX_{59}X_{60}X_{61}X_{62}X_{63}X_{64}X_{65}X_{66}X_{67}X_{68}X_{69}X_{70}X_{71}X_{72}X_{73}X_{74}GX_{75}IX_{76}DTKX_{77}YX_{78}X_{79}GYX_{80}L$
 $PIX_{81}QX_{82}DEM_{83}RYX_{84}X_{85}ENX_{86}X_{87}RX_{88}X_{89}X_{90}X_{91}NX_{92}NX_{93}WWX_{94}X_{95}X_{96}X_{97}X_{98}X_{99}X_{100}X_{101}$
 $X_{102}X_{103}X_{104}X_{105}X_{106}FX_{107}X_{108}X_{109}X_{110}FX_{111}GX_{112}X_{113}X_{114}X_{115}X_{116}X_{117}X_{118}RX_{119}X_{120}X_{121}X_{122}X_{123}$
 $X_{124}X_{125}X_{126}GX_{127}X_{128}X_{129}X_{130}X_{131}X_{132}X_{133}LLX_{134}X_{135}X_{136}X_{137}X_{138}X_{139}X_{140}X_{141}X_{142}X_{143}X_{144}X_{145}$

X₁₄₆X₁₄₇X₁₄₈X₁₄₉X₁₅₀X₁₅₁X₁₅₂X₁₅₃FX₁₅₄X₁₅₅X₁₅₆X₁₅₇X₁₅₈X₁₅₉X₁₆₀ (SEQ ID NO: 311), wherein X₁ is F, Q, N, D, or absent, X₂ is L, I, T, S, N, or absent, X₃ is V, I, G, A, E, T, or absent, X₄ is K, C, or absent, X₅ is G, S, or absent, X₆ is A, S, E, D, N, or absent, X₇ is M, I, V, Q, F, L, or absent, X₈ is E, S, T, N, or absent, X₉ is I, M, E, T, Q, or absent, X₁₀ is K, S, L, I, T, E, or absent, X₁₁ is K or absent, X₁₂ is S, A, E, D, or absent, X₁₃ is E, N, Q, K, or absent, X₁₄ is L, M, V, or absent, X₁₅ is R or absent, X₁₆ is H, D, T, G, E, N, or absent, X₁₇ is K, N, Q, E, A, or absent, X₁₈ is L or absent, X₁₉ is R, Q, N, T, D, or absent, X₂₀ is H, M, V, N, T, or absent, X₂₁ is V, L, I, or absent, X₂₂ is P, S, or absent, X₂₃ is H or absent, X₂₄ is E, D, or absent, X₂₅ is Y or absent, X₂₆ is I, L, or absent, X₂₇ is E, Q, G, S, A, Y, or absent, X₂₈ is L or absent, X₂₉ is I, V, L, or absent, X₃₀ is E, D, or absent, X₃₁ is I, L, or absent, X₃₂ is A, S, or absent, X₃₃ is Q, Y, F, or absent, X₃₄ is D or absent, X₃₅ is S, P, or absent, X₃₆ is K, Y, Q, T, or absent, X₃₇ is Q or absent, X₃₈ is N or absent, X₃₉ is R, K, or absent, X₄₀ is L, I, or absent, X₄₁ is L, F, or absent, X₄₂ is E or absent, X₄₃ is F, M, L, or absent, X₄₄ is V, T, or I, X₄₅ is V, M, L, or I, X₄₆ is E, D, or Q, X₄₇ is F or L, X₄₈ is F or L, X₄₉ is K, I, T, or V, X₅₀ is K, N, or E, X₅₁ is I or E, X₅₂ is Y, F, or C, X₅₃ is G, or N, X₅₄ is Y, or F, X₅₅ is R, S, N, E, K, or Q, X₅₆ is K, S, L, V, or T, X₅₇ is S, A, or V, X₅₈ is K or R, X₅₉ is A, I, or V, X₆₀ is L, M, V, I, or C, X₆₁ is F or Y, X₆₂ is T, A, or S, X₆₃ is K, E, or absent, X₆₄ is D, E, or absent, X₆₅ is E, A, or absent, X₆₆ is N, K, or absent, X₆₇ is E, S, or absent, X₆₈ is D, E, Q, A, or absent, X₆₉ is G, V, K, N, or absent, X₇₀ is L, G, E, S, or absent, X₇₁ is V, S, K, T, E, or absent, X₇₂ is L, H, K, E, Y, D, or A, X₇₃ is N, G, or D, X₇₄ is H, F, or Y, X₇₅ is I, or V, X₇₆ is L, V, or I, X₇₇ is A or S, X₇₈ is K or S, X₇₉ is D, G, K, S, or N, X₈₀ is R, N, S, or G, X₈₁ is S, A, or G, X₈₂ is A, I, or V, X₈₃ is Q, E, I, or V, X₈₄ is V or I, X₈₅ is D, R, G, I, or E, X₈₆ is N, I, or Q, X₈₇ is K, D, T, E, or K, X₈₈ is S, N, D, or E, X₈₉ is Q, E, I, K, or A, X₉₀ is V, H, R, K, L, or E, X₉₁ is I, V, or R, X₉₂ is P, S, T, or R, X₉₃ is E, R, C, Q, or K, X₉₄ is E, N, or K, X₉₅ is I, V, N, E, or A, X₉₆ is Y or F, X₉₇ is P, G, or E, X₉₈ is T, E, S, D, K, or N, X₉₉ is S, D, K, G, N, or T, X₁₀₀ is I, T, V, or L, X₁₀₁ is T, N, G, or D, X₁₀₂ is D, E, T, K, or I, X₁₀₃ is F or Y, X₁₀₄ is K or Y, X₁₀₅ is F or Y, X₁₀₆ is L, S, or M, X₁₀₇ is V or I, X₁₀₈ is S or A, X₁₀₉ is G or A, X₁₁₀ is F, Y, H, E, or K, X₁₁₁ is Q, K, T, N, or I, X₁₁₂ is D, N, or K, X₁₁₃ is Y, F, I, or V, X₁₁₄ is R, E, K, Q, or F, X₁₁₅ is K, E, A, or N, X₁₁₆ is Q or K, X₁₁₇ is L or I, X₁₁₈ is E, D, N, or Q, X₁₁₉ is V, I, or L, X₁₂₀ is S, N, F, T, or Q, X₁₂₁ is H, I, C, or R, X₁₂₂ is L, D, N, S, or F, X₁₂₃ is T or K, X₁₂₄ is K, G, or N, X₁₂₅ is C, V, or I, X₁₂₆ is Q, L, K, or Y, X₁₂₇ is A, G, or N, X₁₂₈ is V or A, X₁₂₉ is M, L, I, V, or A, X₁₃₀ is S, T, or D, X₁₃₁ is V or I, X₁₃₂ is E, Q, K, S, or I, X₁₃₃ is Q, H, or T, X₁₃₄ is L, R, or Y, X₁₃₅ is G, I, L, or T, X₁₃₆ is G, A, or V, X₁₃₇ is E, N, or D, X₁₃₈ is K, Y, D, E, A, or R, X₁₃₉ is I, F, Y, or C, X₁₄₀ is K or R, X₁₄₁ is E, R, A, G, or T, X₁₄₂ is G or N, X₁₄₃ is S, I, K, R, or E, X₁₄₄ is L, I, or M, X₁₄₅ is T, S, D, or K, X₁₄₆ is L, H, Y, R, T, or F, X₁₄₇ is E, Y, I, M, A, or L, X₁₄₈ is E, D, R, or G, X₁₄₉ is V, F, M, L, or I,

X₁₅₀ is G, K, R, L, V, or E, X₁₅₁ is K, N, D, L, H, or S, X₁₅₂ is K, L, C, or absent, X₁₅₃ is K, S, I, Y, M, or F, X₁₅₄ is K, L, C, H, D, Q, or N, X₁₅₅ is N or Y, X₁₅₆ is D, K, T, E, C, or absent, X₁₅₇ is E, V, R, or absent, X₁₅₈ is I, F, L, or absent, X₁₅₉ is V, Q, E, L, or absent, and X₁₆₀ is F or absent.

[0018] In some embodiments, an endonuclease of the present disclosure can have a sequence of X₁X₂X₃X₄X₅X₆X₇X₈X₉X₁₀X₁₁X₁₂X₁₃X₁₄X₁₅X₁₆X₁₇X₁₈X₁₉X₂₀X₂₁X₂₂X₂₃X₂₄X₂₅X₂₆X₂₇X₂₈X₂₉X₃₀X₃₁X₃₂X₃₃X₃₄X₃₅X₃₆X₃₇X₃₈X₃₉X₄₀X₄₁X₄₂X₄₃KX₄₄X₄₅X₄₆X₄₇X₄₈X₄₉X₅₀X₅₁X₅₂X₅₃X₅₄X₅₅GX₅₆HLGGX₅₇RX₅₈PDGX₅₉X₆₀X₆₁X₆₂X₆₃X₆₄X₆₅X₆₆X₆₇X₆₈X₆₉X₇₀X₇₁X₇₂X₇₃X₇₄GX₇₅IX₇₆DTKX₇₇YX₇₈X₇₉GYX₈₀LPIX₈₁QX₈₂DEMX₈₃RYX₈₄X₈₅ENX₈₆X₈₇RX₈₈X₈₉X₉₀X₉₁NX₉₂NX₉₃WWX₉₄X₉₅X₉₆X₉₇X₉₈X₉₉X₁₀₀X₁₀₁X₁₀₂X₁₀₃X₁₀₄X₁₀₅X₁₀₆FX₁₀₇X₁₀₈X₁₀₉X₁₁₀FX₁₁₁GX₁₁₂X₁₁₃X₁₁₄X₁₁₅X₁₁₆X₁₁₇X₁₁₈RX₁₁₉X₁₂₀X₁₂₁X₁₂₂X₁₂₃X₁₂₄X₁₂₅X₁₂₆GX₁₂₇X₁₂₈X₁₂₉X₁₃₀X₁₃₁X₁₃₂X₁₃₃LLX₁₃₄X₁₃₅X₁₃₆X₁₃₇X₁₃₈X₁₃₉X₁₄₀X₁₄₁X₁₄₂X₁₄₃X₁₄₄X₁₄₅X₁₄₆X₁₄₇X₁₄₈X₁₄₉X₁₅₀X₁₅₁X₁₅₂X₁₅₃FX₁₅₄X₁₅₅X₁₅₆X₁₅₇X₁₅₈X₁₅₉X₁₆₀ (SEQ ID NO: 312), wherein X₁ is F, Q, N, or absent, X₂ is L, I, T, S, or absent, X₃ is V, I, G, A, E, T, or absent, X₄ is K, C, or absent, X₅ is G, S, or absent, X₆ is A, S, E, D, or absent, X₇ is M, I, V, Q, F, L, or absent, X₈ is E, S, T, or absent, X₉ is I, M, E, T, Q, or absent, X₁₀ is K, S, L, I, T, E, or absent, X₁₁ is K or absent, X₁₂ is S, A, E, D, or absent, X₁₃ is E, N, Q, K, or absent, X₁₄ is L, M, V, or absent, X₁₅ is R or absent, X₁₆ is H, D, T, G, E, N, or absent, X₁₇ is K, N, Q, E, A, or absent, X₁₈ is L or absent, X₁₉ is R, Q, N, T, D, or absent, X₂₀ is H, M, V, N, T, or absent, X₂₁ is V, L, I, or absent, X₂₂ is P, S, or absent, X₂₃ is H or absent, X₂₄ is E, D, or absent, X₂₅ is Y or absent, X₂₆ is I, L, or absent, X₂₇ is E, Q, G, S, A, or absent, X₂₈ is L or absent, X₂₉ is I, V, L, or absent, X₃₀ is E, D, or absent, X₃₁ is I, L, or absent, X₃₂ is A, S, or absent, X₃₃ is Q, Y, F, or absent, X₃₄ is D or absent, X₃₅ is S, P, or absent, X₃₆ is K, Y, Q, T, or absent, X₃₇ is Q or absent, X₃₈ is N or absent, X₃₉ is R or absent, X₄₀ is L, I, or absent, X₄₁ is L, F, or absent, X₄₂ is E or absent, X₄₃ is F, M, L, or absent, X₄₄ is V, T, or I, X₄₅ is V, M, L, or I, X₄₆ is E, D, or Q, X₄₇ is F or L, X₄₈ is F or L, X₄₉ is K, I, T, or V, X₅₀ is K, N, or E, X₅₁ is I or E, X₅₂ is Y, F, or C, X₅₃ is G, or N, X₅₄ is Y, or F, X₅₅ is R, S, N, E, K, or Q, X₅₆ is K, S, L, V, or T, X₅₇ is S or A, X₅₈ is K or R, X₅₉ is A, I, or V, X₆₀ is L, M, V, I, or C, X₆₁ is F or Y, X₆₂ is T, A, or S, X₆₃ is K, E, or absent, X₆₄ is D, E, or absent, X₆₅ is E, A, or absent, X₆₆ is N, K, or absent, X₆₇ is E, S, or absent, X₆₈ is D, E, Q, A, or absent, X₆₉ is G, V, K, N, or absent, X₇₀ is L, G, E, S, or absent, X₇₁ is V, S, K, T, E, or absent, X₇₂ is L, H, K, E, Y, D, or A, X₇₃ is N, G, or D, X₇₄ is H, F, or Y, X₇₅ is I, or V, X₇₆ is L, V, or I, X₇₇ is A or S, X₇₈ is K or S, X₇₉ is D, G, K, S, or N, X₈₀ is R, N, S, or G, X₈₁ is S, A, or G, X₈₂ is A, I, or V, X₈₃ is Q, E, I, or V, X₈₄ is V or I, X₈₅ is D, R, G, I, or E, X₈₆ is N, I, or Q, X₈₇ is K, D, T, E, or K, X₈₈ is S, N, D, or E, X₈₉ is Q, E, I, K, or A, X₉₀ is V, H, R, K, L, or E, X₉₁ is I, V, or R, X₉₂ is P, S, T, or R, X₉₃ is E, R, C, Q, or K, X₉₄ is E, N, or K, X₉₅ is I, V, N, E, or A, X₉₆ is Y or F, X₉₇ is P, G, or E, X₉₈ is T, E, S, D, K, or N, X₉₉ is S, D, K, G, N, or T, X₁₀₀ is I, T, V, or L, X₁₀₁ is

T, N, G, or D, X₁₀₂ is D, E, T, K, or I, X₁₀₃ is F or Y, X₁₀₄ is K or Y, X₁₀₅ is F or Y, X₁₀₆ is L, S, or M, X₁₀₇ is V or I, X₁₀₈ is S or A, X₁₀₉ is G or A, X₁₁₀ is F, Y, H, E, or K, X₁₁₁ is Q, K, T, N, or I, X₁₁₂ is D, N, or K, X₁₁₃ is Y, F, I, or V, X₁₁₄ is R, E, K, Q, or F, X₁₁₅ is K, E, A, or N, X₁₁₆ is Q or K, X₁₁₇ is L or I, X₁₁₈ is E, D, N, or Q, X₁₁₉ is V, I, or L, X₁₂₀ is S, N, F, T, or Q, X₁₂₁ is H, I, C, or R, X₁₂₂ is L, D, N, S, or F, X₁₂₃ is T or K, X₁₂₄ is K, G, or N, X₁₂₅ is C, V, or I, X₁₂₆ is Q, L, K, or Y, X₁₂₇ is A, G, or N, X₁₂₈ is V or A, X₁₂₉ is M, L, I, V, or A, X₁₃₀ is S, T, or D, X₁₃₁ is V or I, X₁₃₂ is E, Q, K, S, or I, X₁₃₃ is Q, H, or T, X₁₃₄ is L, R, or Y, X₁₃₅ is G, I, L, or T, X₁₃₆ is G, A, or V, X₁₃₇ is E, N, or D, X₁₃₈ is K, Y, D, E, A, or R, X₁₃₉ is I, F, Y, or C, X₁₄₀ is K or R, X₁₄₁ is E, R, A, G, or T, X₁₄₂ is G or N, X₁₄₃ is S, I, K, R, or E, X₁₄₄ is L, I, or M, X₁₄₅ is T, S, D, or K, X₁₄₆ is L, H, Y, R, or T, X₁₄₇ is E, Y, I, M, or A, X₁₄₈ is E, D, R, or G, X₁₄₉ is V, F, M, L, or I, X₁₅₀ is G, K, R, L, V, or E, X₁₅₁ is K, N, D, L, H, or S, X₁₅₂ is K, L, C, or absent, X₁₅₃ is K, S, I, Y, M, or F, X₁₅₄ is K, L, C, H, D, Q, or N, X₁₅₅ is N or Y, X₁₅₆ is D, K, T, E, C, or absent, X₁₅₇ is E, V, R, or absent, X₁₅₈ is I, F, L, or absent, X₁₅₉ is V, Q, E, L, or absent, and X₁₆₀ is F or absent.

[0019] In some embodiments, an endonuclease of the present disclosure can have a sequence of X₁LVKSSX₂EEX₃KEELREKLX₄HLSEYLYX₅LX₆DLAYDSKQNRLFEMKVX₇ELLINECGYX₈GLHLGGSRKPDGIX₉YTEGLKX₁₀NYGHIIDTKAYSDGYNLPISQADEMERYIRENNTRNX₁₁X₁₂VNPNEWWENFPX₁₃NINEFYFLFVSGHFKGNX₁₄EEQLERISIX₁₅TX₁₆IKGAAMSVX₁₇TLLLLAN EIKAGRLX₁₈LEEVX₁₉KYFDNKEIX₂₀F (SEQ ID NO: 313), wherein X₁ is F, Q, N, D, or absent, X₂ is M, I, V, Q, F, L, or absent, X₃ is K, S, L, I, T, E, or absent, X₄ is R, Q, N, T, D, or absent, X₅ is E, Q, G, S, A, Y, or absent, X₆ is I, V, L, or absent, X₇ is V, M, L, or I, X₈ is R, S, N, E, K, or Q, X₉ is L, M, V, I, or C, X₁₀ is L, H, K, E, Y, D, or A, X₁₁ is Q, E, I, K, or A, X₁₂ is V, H, R, K, L, or E, X₁₃ is T, E, S, D, K, or N, X₁₄ is Y, F, I, or V, X₁₅ is L, D, N, S, or F, X₁₆ is K, G, or N, X₁₇ is E, Q, K, S, or I, X₁₈ is T, S, D, or K, X₁₉ is G, K, R, L, V, or E, and X₂₀ is V, Q, E, L, or absent.

[0020] In some embodiments, an endonuclease of the present disclosure can have a sequence of X₁LVKSSX₂EEX₃KEELREKLX₄HLSEYLYX₅LX₆DLAYDSKQNRLFEMKVX₇ELLINECGYX₈GLHLGGSRKPDGIX₉YTEGLKX₁₀NYGHIIDTKAYSDGYNLPISQADEMERYIRENNTRNX₁₁X₁₂VNPNEWWENFPX₁₃NINEFYFLFVSGHFKGNX₁₄EEQLERISIX₁₅TX₁₆IKGAAMSVX₁₇TLLLLAN EIKAGRLX₁₈LEEVX₁₉KYFDNKEIX₂₀F (SEQ ID NO: 314), wherein X₁ is F, Q, N, or absent, X₂ is M, I, V, Q, F, L, or absent, X₃ is K, S, L, I, T, E, or absent, X₄ is R, Q, N, T, D, or absent, X₅ is E, Q, G, S, A, or absent, X₆ is I, V, L, or absent, X₇ is V, M, L, or I, X₈ is R, S, N, E, K, or Q, X₉ is L, M, V, I, or C, X₁₀ is L, H, K, E, Y, D, or A, X₁₁ is Q, E, I, K, or A, X₁₂ is V, H, R, K, L, or E, X₁₃ is T, E, S, D, K, or N, X₁₄ is Y, F, I, or V, X₁₅ is L, D, N, S, or F, X₁₆ is K, G, or N, X₁₇ is E, Q, K, S, or I, X₁₈ is T, S, D, or K, X₁₉ is G, K, R, L, V, or E, and X₂₀ is V, Q, E, L, or absent.

[0021] In some embodiments, an endonuclease of the present disclosure can have conserved amino acid residues at position 76 (D or E), position 98 (D), and position 100 (K), which together preserve catalytic function. In some embodiments, an endonuclease of the present disclosure can have conserved amino acid residues at position 114 (D) and position 118 (R), which together preserve dimerization of two cleavage domains.

[0022] In some embodiments, endonucleases disclosed herein (e.g., SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) can have at least 33.3% divergence from SEQ ID NO: 163 (FokI) and, is immunologically orthogonal to SEQ ID NO: 163 (FokI). In some embodiments, an immunologically orthogonal endonuclease (e.g., SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) can be administered to a patient that has already received, and is thus can have an adverse immune reaction to, FokI. In some embodiments, endonucleases disclosed herein (e.g., SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) can have at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% divergence from SEQ ID NO: 163 (FokI).

[0023] In some embodiments, an endonuclease disclosed herein (e.g., SEQ ID NO: 1 – SEQ ID NO: 81 (nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) can be fused to any nucleic acid binding domain disclosed herein to form a non-naturally occurring fusion protein. This fusion protein can have one or more of the following characteristics: (a) induces greater than 1% indels (insertions/deletions) at a target site; (b) the cleavage domain comprises a molecular weight of less than 23 kDa; (c) the cleavage domain comprises less than 196 amino acids; and (d) capable of cleaving across a spacer region greater than 24 base pairs. In some embodiments, the non-naturally occurring fusion protein can induce greater than 5%, greater than 10%, greater than 20%, greater than 30%, greater than 40%, greater than 50%, greater than 60%, greater than 70%, greater than 80%, or greater than 90% indels at the target site. In some embodiments, indels are generated via the non-homologous end joining (NHEJ) pathway upon administration of a genome editing complex disclosed herein to a subject. Indels can be measured using deep sequencing,

DNA Binding Domains Fused to Nucleases of SEQ ID NO: 1 – SEQ ID NO: 81 (Nucleic Acid Sequences of SEQ ID NO: 82 – SEQ ID NO: 162)

[0024] The present disclosure provides for novel compositions of endonucleases with modular nucleic acid binding domains (e.g., TALEs, RNBDs, or MAP-NBDs) described herein. In some instances the novel endonucleases can be fused to a DNA binding domain from *Xanthomonas spp.*

(TALE), *Ralstonia* (RNBD), or *Legionella* (MAP-NBD) resulting in genome editing complexes. The genome editing complexes described herein can be used to selectively bind and cleave to a target gene sequence for genome editing purposes. For example, a DNA binding domain from *Xanthomonas*, *Ralstonia*, or *Legionella* of the present disclosure can be used to direct the binding of a genome editing complex to a desired genomic sequence.

[0025] The genome editing complexes described herein, comprising a DNA binding domain fused to an endonuclease, can be used to edit genomic loci of interest by binding to a target nucleic acid sequence via the DNA binding domain and cleaving phosphodiester bonds of target double stranded DNA via the endonuclease.

[0026] In some aspects, DNA binding domains fused to nucleases can create a site-specific double-stranded DNA break when fused to a nuclease. Such breaks can then be subsequently repaired by cellular machinery, through either homology-dependent repair or non-homologous end joining (NHEJ). Genome editing, using DNA binding domains fused to nucleases described herein,, can thus be used to delete a sequence of interest (e.g., an aberrantly expressed or mutated gene) or to introduce a nucleic acid sequence of interest (e.g., a functional gene). DNA binding domains of the present disclosure can be programmed to delivery virtually any nuclease, including those disclosed herein, to any target site for therapeutic purposes, including ex vivo engineered cell therapies obtained using the compositions disclosed herein or gene therapy by direct in vivo administration of the compositions disclosed herein. In addition, the DNA binding domain can bind to specific DNA sequences and in some cases they can activate the expression of host genes. In some instances, the disclosure provides for enzymes, e.g., SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) that can be fused to the DNA binding domains of TALEs, RNBDs, and MAP-NBDs. In some instances, enzymes of the disclosure, including SEQ ID NO: 1 (nucleic acid sequence of SEQ ID NO: 82), SEQ ID NO: 4 (nucleic acid sequence of SEQ ID NO: 85), and SEQ ID NO: 8 (nucleic acid sequence of SEQ ID NO: 89), can achieve greater than 30% indels via the NHEJ pathway on a target gene when fused to a DNA binding domain of a TALE, RNBD, and MAP-NBD.

[0027] A non-naturally occurring fusion protein of the disclosure, e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) fused to a DNA binding domain, can comprise a repeat unit. A repeat unit can be from wild-type *Xanthomonas*-derived protein, *Ralstonia*-derived protein, or *Legionella*-derived protein or a modified repeat unit enhanced for specific recognition of a target nucleic acid base. A modified repeat unit can comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 or more mutations that can enhance

the repeat module for specific recognition of a target nucleic acid base. In some embodiments, a modified repeat unit is modified at amino acid position 2, 3, 4, 11, 12, 13, 21, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, or 35. In some embodiments, a modified repeat unit is modified at amino acid positions 12 or 13.

[0028] For purposes of gene editing, a first DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain and a second DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain can be provided. The first DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain can recognize a top strand of double stranded DNA and bind to said region of double stranded DNA. The second DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain can recognize a separate, non-overlapping bottom strand of double stranded DNA and bind to said region of double stranded DNA. The target nucleic acid sequence on the bottom strand can have its complementary nucleic acid sequence in the top strand positioned 10 to 20 nucleotides towards the 3' end from the first region. In some embodiments this stretch of 10 to 20 nucleotides can be referred to as the spacer region. In some embodiments, this first DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain and the second DNA binding domain (e.g., of a TALE, RNBD, or MAP-NBD) linked to a cleavage domain both bind at a target site, allowing for dimerization of the two cleavage domains in the spacer region and allowing for catalytic activity and cleaving of the target DNA.

[0029] As described in further detail below, a non-naturally occurring fusion protein of the disclosure, e.g., anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) fused to a plurality of repeat units (e.g., derived from *Ralstonia solanacearum*, *Xanthomonas spp.*, or *Legionella quateirensis*), can further comprise a C-terminal truncation, which can served as a linker between the DNA binding domain and the nuclease.

[0030] A non-naturally occurring fusion protein of the disclosure, e.g., anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) fused to a DNA binding domain, can further comprise an N-terminal cap as described in further detail below. An N-terminal cap can be a polypeptide portion flanking the DNA-binding repeat unit. An N-terminal cap can be any length and can comprise from 0 to 136 amino acid residues in length. An N-terminal cap can be 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, or 130 amino acid residues in length. In some embodiments, an N-terminal cap can modulate structural stability of the DNA-binding repeat units. In some embodiments, an N-terminal cap can modulate nonspecific interactions. In some cases, an N-terminal cap can decrease nonspecific interaction. In some cases,

an N-terminal cap can reduce off-target effect. As used here, off-target effect refers to the interaction of a genome editing complex with a sequence that is not the target binding site of interest. An N-terminal cap can further comprise a wild-type N-terminal cap sequence of a protein from *Ralstonia solanacearum*, *Xanthomonas spp.*, or *Legionella quateirensis* or can comprise a modified N-terminal cap sequence.

[0031] In some embodiments, a DNA binding domain comprises at least one repeat unit having a repeat variable diresidue (RVD), which contacts a target nucleic acid base. In some embodiments, a DNA binding domain comprises more than one repeat unit, each having an RVD, which contacts a target nucleic acid base. In some embodiments, the DNA binding domain comprises 1 to 50 RVDs. In some embodiments, the DNA binding domain components of the fusion proteins can be at least 14 RVDs, at least 15 RVDs, at least 16 RVDs, at least 17 RVDs, at least 18 RVDs, at least 19 RVDs, at least 20 RVDs in length, or at least 21 RVDs in length. In some embodiments, the DNA binding domains can be 16 to 21 RVDs in length.

[0032] In some embodiments, any one of the DNA binding domains described herein can bind to a region of interest of any gene. For example, the DNA binding domains described herein can bind upstream of the promoter region, upstream of the gene transcription start site, or downstream of the transcription start site. In certain embodiments, the DNA binding domain binding region is no farther than 50 base pairs downstream of the transcription start site. In some embodiments, the DNA binding domain is designed to bind in proximity to the transcription start site (TSS). In other embodiments, the TALE can be designed to bind in the 5' UTR region.

[0033] A DNA binding domain described herein can comprise between 1 to 50 repeat units. A DNA binding domain described herein can comprise between 5 and 45, between 8 to 45, between 10 to 40, between 12 to 35, between 15 to 30, between 20 to 30, between 8 to 40, between 8 to 35, between 8 to 30, between 10 to 35, between 10 to 30, between 10 to 25, between 10 to 20, or between 15 to 25 repeat units.

[0034] A DNA binding domain described herein can comprise at least 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, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, 50, or more repeat units. A DNA binding domain described herein can comprise 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, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 45, or 50 repeat units. A DNA binding domain described herein can comprise 5 repeat units. A DNA binding domain described herein can comprise 10 repeat units. A DNA binding domain described herein can comprise 11 repeat units. A DNA binding domain described herein can comprise 12 repeat units, or another suitable number.

[0035] A repeat unit of a DNA binding domain can be 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 or 40 residues in length. A repeat unit of a DNA binding domain can be 18 residues in length. A repeat unit of a DNA binding domain can be 19 residues in length. A repeat unit of a DNA binding domain can be 20 residues in length. A repeat unit of a DNA binding domain can be 21 residues in length. A repeat unit of a DNA binding domain can be 22 residues in length. A repeat unit of a DNA binding domain can be 23 residues in length. A repeat unit of a DNA binding domain can be 24 residues in length. A repeat unit of a DNA binding domain can be 25 residues in length. A repeat unit of a DNA binding domain can be 26 residues in length. A repeat unit of a DNA binding domain can be 27 residues in length. A repeat unit of a DNA binding domain can be 28 residues in length. A repeat unit of a DNA binding domain can be 29 residues in length. A repeat unit of a DNA binding domain can be 30 residues in length. A repeat unit of a DNA binding domain can be 31 residues in length. A repeat unit of a DNA binding domain can be 32 residues in length. A repeat unit of a DNA binding domain can be 33 residues in length. A repeat unit of a DNA binding domain can be 34 residues in length. A repeat unit of a DNA binding domain can be 35 residues in length. A repeat unit of a DNA binding domain can be 36 residues in length. A repeat unit of a DNA binding domain can be 37 residues in length. A repeat unit of a DNA binding domain can be 38 residues in length. A repeat unit of a DNA binding domain can be 39 residues in length. A repeat unit of a DNA binding domain can be 40 residues in length.

[0036] In some embodiments, the effector can be a protein secreted from *Xanthomonas* or *Ralstonia* bacteria upon plant infection. In some embodiments, the effector can be a protein that is a mutated form of, or otherwise derived from, a protein secreted from *Xanthomonas* or *Ralstonia* bacteria. The effector can further comprise a DNA-binding module which includes a variable number of about 33-35 amino acid residue repeat units. Each amino acid repeat unit recognizes one base pair through two adjacent amino acids (*e.g.*, at amino acid positions 12 and 13 of the repeat unit). As such, amino acid positions 12 and 13 of the repeat unit can also be referred to as repeat variable diresidue (RVD).

[0037] Also provided herein is a nucleic acid molecule encoding a fusion protein comprising a DNA-binding domain and

[0038] (I) a polypeptide having the activity of an endonuclease, wherein the nucleic acid molecule is selected from the group consisting of:

- (a) a nucleic acid molecule encoding a polypeptide comprising or consisting of the amino acid sequence of any one of SEQ ID NOs: 65; 74; 1; 2; 3; 16; 6; 4; 8; 56; 58; 59; 48; 17; 22; 49; 51; 53; 52; 47; 52; 54;

[0039] (b) a nucleic acid molecule comprising or consisting of the nucleotide sequence of any one of SEQ ID NOs: 146; 155; 82; 83; 84; 85; 97; 87; 89; 137; 139; 140; 129; 98; 103; 130; 132; 134; 128; 133 or 135;

[0040] (c) a nucleic acid molecule encoding an endonuclease, where the amino acid sequence of the endonuclease is at least 80% identical to the amino acid sequence of any one of SEQ ID NOs: 65; 74; 1; 2; 3; 16; 6; 4; 8; 56; 58; 59; 48; 17; 22; 49; 51; 53; 52; 47; 52; 54;

[0041] (d) a nucleic acid molecule comprising or consisting of a nucleotide sequence which is at least 70% identical to the nucleotide sequence of any one of SEQ ID NOs: 146; 155; 82; 83; 84; 85; 97; 87; 89; 137; 139; 140; 129; 98; 103; 130; 132; 134; 128; 133 or 135;

[0042] (e) a nucleic acid molecule which is degenerate with respect to the nucleic acid molecule of (d); and

[0043] (f) a nucleic acid molecule corresponding to the nucleic acid molecule of any one of (a) to (e) wherein T is replaced by U; or

[0044] (II) a fragment of the polypeptide of (I) having the activity of an endonuclease.

[0045] Fragments according to the present disclosures are polypeptides having the activity of an endonuclease as defined herein above and comprise at least 90 amino acids. In certain aspects, the fragments of the endonucleases are polypeptides of at least 100, at least 125, at least 150, or at least 190 amino acids. Fragments of the polypeptide of the disclosure, which substantially retain endonuclease activity, include N-terminal truncations, C-terminal truncations, amino acid substitutions, internal deletions and addition of amino acids (either internally or at either terminus of the protein). For example, conservative amino acid substitutions are known in the art and may be introduced into the endonuclease of the disclosure without substantially affecting endonuclease activity, i.e. reducing said activity. In certain aspects, the deletions and/or substitutions are not introduced at the amino acid residues involved in catalytic function and/or dimerization of the endonuclease.

[0046] Also provided herein are vectors comprising a nucleic acid sequence encoding the cleavage domains disclosed herein. The nucleic acid may be operably linked to a promoter for expression of the cleavage domain. The terms cleavage domain and nucleases are used herein interchangeably. In another embodiment the invention relates to a host cell comprising, e.g., as a result of transformation, transduction, microinjection or transfection, the nucleic acid molecule or the vector of the invention.

[0047] A variety of host-expression systems may be used to express the endonuclease coding sequence in a host cell using a suitable vector.

[0048] The “host cell” in accordance with the present disclosure may be produced by introducing the nucleic acid molecule or vector(s) of the disclosure into the host cell which upon its/their presence preferably mediates the expression of the nucleic acid molecule of the present disclosure encoding the endonuclease of the disclosure. The host from which the host cell is derived may be any prokaryote or eukaryotic cell.

[0049] A suitable eukaryotic host cell may be a mammalian cell, a vertebrate cell, an amphibian cell, a fish cell, an insect cell, a fungal/yeast cell, a nematode cell or a plant cell.

Potency and Specificity of Genome Editing

[0050] In some embodiments, the efficiency of genome editing with a genome editing complex of the present disclosure (e.g., any one of an RNBD, MAP-NBD, or TALE linked to any nuclease disclosed herein such as any one of SEQ ID NO: 1 – SEQ ID NO: 81) can be determined.

Specifically, the potency and specificity of the genome editing complex can indicate whether a particular modular nucleic acid binding domain fused to a nuclease provides efficient editing.

Potency can be defined as the percent indels (insertions/deletions) that are generated via the non-homologous end joining (NHEJ) pathway at a target site after administering a modular nucleic acid binding domain fused to a nuclease to a subject. A modular nucleic acid binding domain can have a potency of greater than 50%, greater than 55%, greater than 60%, greater than 65%, greater than 70%, greater than 75%, greater than 80%, greater than 85%, greater than 90%, greater than 95%, greater than 92%, greater than 95%, greater than 97%, or greater than 99%. A modular nucleic acid binding domain can have a potency of from 50% to 100%, 50% to 60%, 60% to 70%, 70% to 80%, 80% to 90%, or 90% to 100%.

[0051] Specificity can be defined as a specificity ratio, wherein the ratio is the percent indels at a target site of interest over the percent indels at the top-ranked off-target site for a particular genome editing complex (e.g., any DNA binding domain linked to a nuclease described herein linked to any nuclease disclosed herein such as any one of SEQ ID NO: 1 – SEQ ID NO: 81) of interest. A high specificity ratio would indicate that a modular nucleic acid binding domain fused to a nuclease edits primarily at the desired target site and exhibits fewer instances of undesirable, off-target editing. A low specificity ratio would indicate that a modular nucleic acid binding domain fused to a nuclease does not edit efficiently at the desired target site and/or can indicate that the modular nucleic acid binding domain fused to a nuclease exhibits high off-target activity. A modular nucleic acid binding domain can have a specificity ratio for the target site of at least 50:1, 55:1, 60:1, 65:1, 70:1, 75:1, 80:1, 85:1, 90:1, 92:1, 95:1, 97:1, 99:1, 50:2, 55:2, 60:2, 65:2, 70:2, 75:2, 80:2, 85:2, 90:2, 92:2,

95:2, 97:2, 99:2, 50:3, 55:3, 60:3, 65:3, 70:3, 75:3, 80:3, 85:3, 90:3, 92:3, 95:3, 97:3, 99:3, 50:4, 55:4, 60:4, 65:4, 70:4, 75:4, 80:4, 85:4, 90:4, 92:4, 95:4, 97:4, 99:4, 50:5, 55:5, 60:5, 65:5, 70:5, 75:5, 80:5, 85:5, 90:5, 92:5, 95:5, 97:5, or 99:5. Percent indels can be measured via deep sequencing techniques.

[0052] The top-ranked off-target site for a composition (e.g., a modular nucleic acid binding domain linked to a cleavage domain) can be determined using the predicted report of genome-wide nuclease off-target sites (PROGNOS) ranking algorithms as described in Fine et al. (Nucleic Acids Res. 2014 Apr;42(6):e42. doi: 10.1093/nar/gkt1326. Epub 2013 Dec 30.). As described in Fine et al, the PROGNOS algorithm TALEN v2.0 can use the DNA target sequence as input; prior construction and experimental characterization of the specific nucleases are not necessary. Based on the differences between the sequence of a potential off-target site in the genome and the intended target sequence, the algorithm can generate a score that is used to rank potential off-target sites. If two (or more) potential off-target sites have equal scores, they can be further ranked by the type of genomic region annotated for each site with the following order: Exon > Promoter > Intron > Intergenic. A final ranking by chromosomal location can be employed as a tie-breaker to ensure consistency in the ranking order. Thus, a score can be generated for each potential off-target site.

[0053] Any of the nucleases of SEQ ID NO: 1 – SEQ ID NO: 81 can be fused to RNBDs, MAP-NBDs, and TALEs. Further details about RNBDs, MAP-NBDs, and TALEs are provided below.

A. Ralstonia-Derived DNA Binding Domains (RNBDs)

[0054] The present disclosure provides non-naturally occurring fusion proteins of a nuclease (e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81) to a modular Ralstonia-derived nucleic acid binding domain (RNBD).

[0055] The present disclosure provides modular nucleic acid binding domains (NBDs) derived from the genus of bacteria. For example, in some embodiments, the present disclosure provides NBDs derived from bacteria that serve as plant pathogens, such as from the genus of *Xanthomonas spp.* and *Ralstonia*. In particular embodiments, the present disclosure provides NBDs from the genus of *Ralstonia*. Provided herein are sequences of repeat units derived from the genus of *Ralstonia*, which can be linked together to form non-naturally occurring modular nucleic acid binding domains (NBDs), capable of targeting and binding any target nucleic acid sequence (e.g., DNA sequence).

[0056] In particular embodiments, modular nucleic acid binding domains (NBDs), also referred to herein as “DNA binding polypeptides,” are provided herein from the genus of *Ralstonia solanacearum*. In some embodiments, modular nucleic acid binding domains derived from *Ralstonia*

(RNBDs) can be engineered to bind to a target gene of interest for purposes of gene editing or gene repression. An RNBD can be engineered to target and bind a specific nucleic acid sequence. The nucleic acid sequence can be DNA or RNA.

[0057] In some embodiments, the RNBD can comprise a plurality of repeat units, wherein each repeat unit recognizes and binds to a single nucleotide (in DNA or RNA) or base pair. Each repeat unit in the plurality of repeat units can be specifically selected to target and bind to a specific nucleic acid sequence, thus contributing to the modular nature of the DNA binding polypeptide. A non-naturally occurring *Ralstonia*-derived modular nucleic acid binding domain can comprise a plurality of repeat units, wherein each repeat unit of the plurality of repeat units recognizes a single target nucleotide, base pair, or both.

[0058] In some embodiments, the repeat unit of a modular nucleic acid binding domain can be derived from a bacterial protein. For example, the bacterial protein can be a transcription activator like effector-like protein (TALE-like protein). The bacterial protein can be derived from *Ralstonia solanacearum*. Repeat units derived from *Ralstonia solanacearum* can be 33-35 amino acid residues in length. In some embodiments, the repeat unit can be derived from the naturally occurring *Ralstonia solanacearum* TALE-like protein.

[0059] **TABLE 2** below shows exemplary repeat units derived from the genus of *Ralstonia*, which are capable of binding a target nucleic acid.

TABLE 2 – Exemplary *Ralstonia*-derived Repeat Units

SEQ ID NO	Sequence
SEQ ID NO: 168	LDTEQVVAIASHNGGKQALEAVKADLLDLGAPYV
SEQ ID NO: 169	LDTEQVVAIASHNGGKQALEAVKADLLDLRGAPYA
SEQ ID NO: 170	LDTEQVVAIASHNGGKQALEAVKADLLELRGAPYA
SEQ ID NO: 171	LDTEQVVAIASHNGGKQALEAVKAHLLDLRGAPYA
SEQ ID NO: 172	LNTEQVVAIASHNGGKQALEAVKADLLDLRGAPYA
SEQ ID NO: 173	LNTEQVVAIASNNGGKQALEAVKTHLLDLRGARYA
SEQ ID NO: 174	LNTEQVVAIASNPGGKQALEAVRALFPDLRAAPYA
SEQ ID NO: 175	LNTEQVVAIASSHGGKQALEAVRALFPDLRAAPYA
SEQ ID NO: 176	LNTEQVVAVASNKGGKQALEAVGAQLLALRAVPYA
SEQ ID NO: 177	LNTEQVVAVASNKGGKQALEAVGAQLLALRAVPYE
SEQ ID NO: 178	LSAAQVVAIASHDGGKQALEAVGTQLVALRAAPYA
SEQ ID NO: 179	LSIAQVVAVASRSGGKQALEAVRAQLLALRAAPYG
SEQ ID NO: 180	LSPEQVVAIASNHGGKQALEAVRALFRGLRAAPYG
SEQ ID NO: 181	LSPEQVVAIASNNGGKQALEAVKAQLLELRAAPYE
SEQ ID NO: 182	LSTAQLVAIASNPGGKQALEAIRALFRELRAAPYA

SEQ ID NO	Sequence
SEQ ID NO: 183	LSTAQLVAIASNPGGKQALEAVRALFRELRAAPYA
SEQ ID NO: 184	LSTAQLVAIASNPGGKQALEAVRAPFREVRAPYA
SEQ ID NO: 185	LSTAQLVSIASNPGGKQALEAVRALFRELRAAPYA
SEQ ID NO: 186	LSTAQVAIASHDGGKQALEAVGTQLVVLRAAPYA
SEQ ID NO: 187	LSTAQVATIASSIGGRQALEALKVQLPVLRAAPYG
SEQ ID NO: 188	LSTAQVATIASSIGGRQALEAVKVQLPVLRAAPYG
SEQ ID NO: 189	LSTAQVVAIAANNGGKQALEAVRALLPVLRVAPYE
SEQ ID NO: 190	LSTAQVVAIAGNGGGKQALEGIGEQLLKLRTAPYG
SEQ ID NO: 191	LSTAQVVAIASHDGGKQALEAAGTQLVALRAAPYA
SEQ ID NO: 192	LSTAQVVAIASHDGGKQALEAVGAQLVELRAAPYA
SEQ ID NO: 193	LSTAQVVAIASHDGGKQALEAVGTQLVALRAAPYA
SEQ ID NO: 194	LSTAQVVAIASHDGGNQALEAVGTQLVALRAAPYA
SEQ ID NO: 195	LSTAQVVAIASHNGGKQALEAVKAQLDLRGAPYA
SEQ ID NO: 196	LSTAQVVAIASNDGGKQALEEVEAQLLALRAAPYE
SEQ ID NO: 197	LSTAQVVAIASNGGGKQALEGIGEQLLKLRTAPYG
SEQ ID NO: 198	LSTAQVVAIASNGGGKQALEGIGELRLKLRTAPYG
SEQ ID NO: 199	LSTAQVVAIASNPGGKQALEAVRALFRELRAAPYA
SEQ ID NO: 200	LSTAQVVAIASQNGGKQALEAVKAQLDLRGAPYA
SEQ ID NO: 201	LSTAQVVAIASSHGGKQALEAVRALFRELRAAPYG
SEQ ID NO: 202	LSTAQVVAIASSNGGKQALEAVWALLPVLRTAPYD
SEQ ID NO: 203	LSTAQVVAIATRSNGGKQALEAVRAQLDLRAAPYG
SEQ ID NO: 204	LSTAQVVAIVAGRNGGKQALEAVRAQLPALRAAPYG
SEQ ID NO: 205	LSTAQVVAIVASSNGGKQALEAVWALLPVLRTAPYD
SEQ ID NO: 206	LSTAQVVTIASSNGGKQALEAVWALLPVLRTAPYD
SEQ ID NO: 207	LSTEQVVAIAGHDGGKQALEAVGAQLVALRAAPYA
SEQ ID NO: 208	LSTEQVVAIASHDGGKQALEAVGAQLVALLAAPYA
SEQ ID NO: 209	LSTEQVVAIASHDGGKQALEAVGAQLVALRAAPYA
SEQ ID NO: 210	LSTEQVVAIASHDGGKQALEAVGGQLVALRAAPYA
SEQ ID NO: 211	LSTEQVVAIASHDGGKQALEAVGTQLVALRAAPYA
SEQ ID NO: 212	LSTEQVVAIASHDGGKQALEAVGVQLVALRAAPYA
SEQ ID NO: 213	LSTEQVVAIASHDGGKQALEAVVAQLVALRAAPYA
SEQ ID NO: 214	LSTEQVVAIASHDGGKQPLEAVGAQLVALRAAPYA
SEQ ID NO: 215	LSTEQVVAIASHGGGKQVLEGIGEQLLKLRAAPYG
SEQ ID NO: 216	LSTEQVVAIASHKGKQALEGIGEQLLKLRAAPYG
SEQ ID NO: 217	LSTEQVVAIASHNGGKQALEAVKADLLDLRGAPYA
SEQ ID NO: 218	LSTEQVVAIASHNGGKQALEAVKADLLELRGAPYA
SEQ ID NO: 219	LSTEQVVAIASHNGGKQALEAVKAHLLDLRGAPYA
SEQ ID NO: 220	LSTEQVVAIASHNGGKQALEAVKAHLLDLRGVPYA
SEQ ID NO: 221	LSTEQVVAIASHNGGKQALEAVKAHLELRGAPYA
SEQ ID NO: 222	LSTEQVVAIASHNGGKQALEAVKAQLDLRGAPYA

SEQ ID NO	Sequence
SEQ ID NO: 223	LSTEQVVAIASHNGGKQALEAVKAQLLELRGAPYA
SEQ ID NO: 224	LSTEQVVAIASHNGGKQALEAVKAQLPVLRRAPYG
SEQ ID NO: 225	LSTEQVVAIASHNGGKQALEAVKTQLLELRGAPYA
SEQ ID NO: 226	LSTEQVVAIASHNGGKQALEAVRAQLPALRAAPYG
SEQ ID NO: 227	LSTEQVVAIASHNGSKQALEAVKAQLLDLRGAPYA
SEQ ID NO: 228	LSTEQVVAIASNNGGKQALEGIGKQLQELRAAPHG
SEQ ID NO: 229	LSTEQVVAIASNNGGKQALEGIGKQLQELRAAPYG
SEQ ID NO: 230	LSTEQVVAIASNHGGKQALEAVRALFRELRAAPYA
SEQ ID NO: 231	LSTEQVVAIASNHGGKQALEAVRALFRGLRAAPYG
SEQ ID NO: 232	LSTEQVVAIASNKGKQALEAVKADLLDLRGAPYV
SEQ ID NO: 233	LSTEQVVAIASNKGKQALEAVKAHLLDLLGAPYV
SEQ ID NO: 234	LSTEQVVAIASNKGKQALEAVKAQLLALRAAPYA
SEQ ID NO: 235	LSTEQVVAIASNKGKQALEAVKAQLLELRGAPYA
SEQ ID NO: 236	LSTEQVVAIASNNGGKQALEAVKALLLELRAPYE
SEQ ID NO: 237	LSTEQVVAIASNNGGKQALEAVKAQLLALRAAPYE
SEQ ID NO: 238	LSTEQVVAIASNNGGKQALEAVKAQLLDLRGAPYA
SEQ ID NO: 239	LSTEQVVAIASNNGGKQALEAVKAQLLVLRAPYG
SEQ ID NO: 240	LSTEQVVAIASNNGGKQALEAVKAQLPALRAAPYE
SEQ ID NO: 241	LSTEQVVAIASNNGGKQALEAVKAQLPVLRRAPCG
SEQ ID NO: 242	LSTEQVVAIASNNGGKQALEAVKAQLPVLRRAPYG
SEQ ID NO: 243	LSTEQVVAIASNNGGKQALEAVKARLLDLRGAPYA
SEQ ID NO: 244	LSTEQVVAIASNNGGKQALEAVKTQLLALRTAPYE
SEQ ID NO: 245	LSTEQVVAIASNPGGKQALEAVRALFPDLRAAPYA
SEQ ID NO: 246	LSTEQVVAIASNHGGKQALEAVRALFPDLRAAPYA
SEQ ID NO: 247	LSTEQVVAIASNHGGKQALEAVRALLPVLRAATPYD
SEQ ID NO: 248	LSTEQVVAIVASHNGGKQALEAVRAQLLDLRAAPYE
SEQ ID NO: 249	LSTEQVVAIVASNKGKQALAAVEAQLRLRAAPYE
SEQ ID NO: 250	LSTEQVVAIVASNKGKQALEEVEAQLRLRAAPYE
SEQ ID NO: 251	LSTEQVVAIVASNKGKQVLEAVGAQLLALRAVPYE
SEQ ID NO: 252	LSTEQVVAIVASNNGGKQALKAVKAQLLALRAAPYE
SEQ ID NO: 253	LSTEQVVVIANSIGGKQALEAVKVQLPVLRAAPYE
SEQ ID NO: 254	LSTGQVVAIASNNGGGRQALEAVREQLLALRAVPYE
SEQ ID NO: 255	LSVAQVVTIASHNGGKQALEAVRAQLLALRAAPYG
SEQ ID NO: 256	LTIAQVVAVASHNGGKQALEAIGAQLLALRAAPYA
SEQ ID NO: 257	LTIAQVVAVASHNGGKQALEVIGAQLLALRAAPYA
SEQ ID NO: 258	LTPQQVVAIAANTGGKQALGAITTQLPILRAAPYE
SEQ ID NO: 259	LTPQQVVAIASNTGGKQALEAVTVQLRVLRGARYG
SEQ ID NO: 260	LTPQQVVAIASNTGGKRALEAVCVQLPVLRAAPYR
SEQ ID NO: 261	LTPQQVVAIASNTGGKRALEAVRVQLPVLRAAPYE
SEQ ID NO: 262	LTTAQVVAIASNDGGKQALEAVGAQLLVLRAPYE

SEQ ID NO	Sequence
SEQ ID NO: 263	LTTAQVVAIASNDGGKQTLEVAGAQLLALRAVPYE
SEQ ID NO: 332	LSTAQVVAVASGSGGKPALEAVRAQLLALRAAPYG
SEQ ID NO: 333	LSTAQVVAVASGSGGKPALEAVRAQLLALRAAPYG
SEQ ID NO: 334	LNTAQVVAIASHDGGKPALEAVWAKLPVLRGAPYA
SEQ ID NO: 335	LNTAQVVAIASHDGGKPALEAVRAKLPVLRGVPYA
SEQ ID NO: 336	LNTAQVVAIASHDGGKPALEAVWAKLPVLRGVPYA
SEQ ID NO: 337	LNTAQVVAIASHDGGKPALEAVWAKLPVLRGVPYE
SEQ ID NO: 338	LSTAQVVVAIASHDGGKPALEAVWAKLPVLRGAPYA
SEQ ID NO: 339	LSTAQVVAVASHDGGKPALEAVRKQLPVLRGVPHQ
SEQ ID NO: 340	LSTAQVVAVASHDGGKPALEAVRKQLPVLRGVPHQ
SEQ ID NO: 341	LNTAQVVAIASHDGGKPALEAVWAKLPVLRGVPYA
SEQ ID NO: 342	LSTEQVVAIASHNGGKLALEAVKAHLLDLRGAPYA
SEQ ID NO: 343	LSTEQVVAIASHNGGKPALEAVKAHLLALRAAPYA
SEQ ID NO: 344	LNTAQVVAIASHYGGKPALEAVWAKLPVLRGVPYA
SEQ ID NO: 345	LNTEQVVAIASNNGGKPALEAVKAQLELRAAPYE
SEQ ID NO: 346	LSPEQVVAIASNNGGKPALEAVKALLLALRAAPYE
SEQ ID NO: 347	LSPEQVVAIASNNGGKPALEAVKAQLELRAAPYE
SEQ ID NO: 348	LSTEQVVAIASNNGGKPALEAVKALLLALRAAPYE
SEQ ID NO: 349	LSTEQVVAIASNNGGKPALEAVKALLLELRAAPYE
SEQ ID NO: 350	LSPEQVVAIASNNGGKPALEAVKALLLALRAAPYE
SEQ ID NO: 351	LSPEQVVAIASNNGGKPALEAVKAQLELRAAPYE
SEQ ID NO: 352	LSTEQVVAIASNNGGKPALEAVKALLLELRAAPYE

[0060] In some embodiments, an RNBD of the present disclosure can comprise between 1 to 50 *Ralstonia solanacearum*-derived repeat units. In some embodiments, an RNBD of the present disclosure can comprise between 9 and 36 *Ralstonia solanacearum*-derived repeat units. Preferably, in some embodiments, an RNBD of the present disclosure can comprise between 12 and 30 *Ralstonia solanacearum*-derived repeat units. An RNBD described herein can comprise between 5 to 10 *Ralstonia solanacearum*-derived repeat units, between 10 to 15 *Ralstonia solanacearum*-derived repeat units, between 15 to 20 *Ralstonia solanacearum*-derived repeat units, between 20 to 25 *Ralstonia solanacearum*-derived repeat units, between 25 to 30 *Ralstonia solanacearum*-derived repeat units, or between 30 to 35 *Ralstonia solanacearum*-derived repeat units, between 35 to 40 *Ralstonia solanacearum*-derived repeat units. An RNBD described herein can comprise 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, , or 40 *Ralstonia solanacearum*-derived repeat units.

[0061]

[0062] A *Ralstonia solanacearum*-derived repeat unit can be derived from a wild-type repeat unit, such as any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352. A *Ralstonia solanacearum*- repeat unit can have at least 80% sequence identity with any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352. A *Ralstonia solanacearum*-derived repeat unit can also comprise a modified *Ralstonia solanacearum*-derived repeat unit enhanced for specific recognition of a nucleotide or base pair. An RNBD described herein can comprise one or more wild-type *Ralstonia solanacearum*-derived repeat units, one or more modified *Ralstonia solanacearum*-derived repeat units, or a combination thereof. In some embodiments, a modified *Ralstonia solanacearum*-derived repeat unit can comprise 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, or 29 mutations that can enhance recognition of a specific nucleotide or base pair. In some embodiments, a modified *Ralstonia solanacearum*-derived repeat unit can comprise more than 1 modification, for example 1 to 5 modifications, 5 to 10 modifications, 10 to 15 modifications, 15 to 20 modifications, 20 to 25 modification, or 25-29 modifications. In some embodiments, an RNBD can comprise more than one modified *Ralstonia solanacearum*-derived repeat units, wherein each of the modified *Ralstonia solanacearum*-derived repeat units can have a different number of modifications.

[0063] The *Ralstonia solanacearum*-derived repeat units comprise amino acid residues at positions 12 and 13, what is referred to herein as, a repeat variable diresidue (RVD). The RVD can modulate binding affinity of the repeat unit for a particular nucleic acid base (e.g., adenosine, guanine, cytosine, thymidine, or uracil). In some embodiments, a single amino acid residue can modulate binding to the target nucleic acid base. In some embodiments, two amino acid residues (RVD) can modulate binding to the target nucleic acid base. In some embodiments, any repeat unit disclosed herein can have an RVD selected from HD, HG, HK, HN, ND, NG, NH, NK, NN, NP, NT, QN, RN, RS, SH, SI, or SN. In some embodiments, an RVD of HD can bind to cytosine. In some embodiments, an RVD of NG can bind to thymidine. In some embodiments, an RVD of NK can bind to guanine. In some embodiments, an RVD of SI can bind to adenosine. In some embodiments, an RVD of RS can bind to adenosine. In some embodiments, an RVD of NT can bind to adenosine.

[0064] In some embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 209 can be included in a DNA binding domain of the present disclosure to bind to cytosine. In some embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 197 can be included in a DNA binding domain of the present disclosure to bind to thymidine. In some embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 233 can be included in a DNA binding domain of the present disclosure to bind to guanine. In some

embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 253 can be included in a DNA binding domain of the present disclosure to bind to adenosine. In some embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 203 can be included in a DNA binding domain of the present disclosure to bind to adenosine. In some embodiments, a repeat unit having at least 80% sequence identity with SEQ ID NO: 218 can be included in a DNA binding domain of the present disclosure to bind to guanine. In some embodiments, an RVD of HN can bind to guanine. In some embodiments, the repeat unit of SEQ ID NO: 209 can be included in a DNA binding domain of the present disclosure to bind to cytosine. In some embodiments, the repeat unit of SEQ ID NO: 197 can be included in a DNA binding domain of the present disclosure to bind to thymidine. In some embodiments, the repeat unit of SEQ ID NO: 233 can be included in a DNA binding domain of the present disclosure to bind to guanine. In some embodiments, the repeat unit of SEQ ID NO: 253 can be included in a DNA binding domain of the present disclosure to bind to adenosine. In some embodiments, the repeat unit of SEQ ID NO: 203 can be included in a DNA binding domain of the present disclosure to bind to adenosine. In some embodiments, the repeat unit of SEQ ID NO: 218 can be included in a DNA binding domain of the present disclosure to bind to guanine.

[0065] In some embodiments, the present disclosure provides repeat units as set forth in SEQ ID NO: 267 – SEQ ID NO: 279. Unspecified amino acid residues in SEQ ID NO: 267 – SEQ ID NO: 279 can be any amino acid residues. In particular embodiments, unspecified amino acid residues in SEQ ID NO: 267 – SEQ ID NO: 279 can be those set forth in the Variable Definition column of **TABLE 3**.

[0066] **TABLE 3** shows consensus sequences of *Ralstonia*-derived repeat units.

TABLE 3: Consensus Sequences of *Ralstonia*-derived Repeat Units

RVD	Consensus Sequence	Variable Definition
HN	LX ₁ X ₂ X ₃ QVVX ₄ X ₅ ASHNGX ₆ KQALEX ₇ X ₈ X ₉ X ₁₀ X ₁₁ LX ₁₂ X ₁₃ LX ₁₄ X ₁₅ X ₁₆ PYX ₁₇ (SEQ ID NO: 267)	X ₁ : D N S T, X ₂ : I T V, X ₃ : A E, X ₄ : A T, X ₅ : I V, X ₆ : G S, X ₇ : A V, X ₈ : I V, X ₉ : G K R, X ₁₀ : A T, X ₁₁ : D H Q, X ₁₂ : L P, X ₁₃ : A D E V, X ₁₄ : L R, X ₁₅ : A G R, X ₁₆ : A V, X ₁₇ : A E G V
NN	LX ₁ X ₂ X ₃ QVVAX ₄ AX ₅ NNGGKQALX ₆ AVX ₇ X ₈ X ₉ LX ₁₀ X ₁₁ LRX ₁₂ AX ₁₃ X ₁₄ X ₁₅ (SEQ ID NO: 268)	X ₁ : N S, X ₂ : P T, X ₃ : A E, X ₄ : I V, X ₅ : A S, X ₆ : E K, X ₇ : K R, X ₈ : A T, X ₉ : H L Q R, X ₁₀ : L P, X ₁₁ : A D E V, X ₁₂ : A G R T V, X ₁₃ : P R, X ₁₄ : C Y, X ₁₅ : A E G
NP	LX ₁ TX ₂ QX ₃ VX ₄ IASNPGGKQALEAX ₅ RAX ₆ F ₇ X ₈ X ₉ RAAPYA (SEQ ID NO: 269)	X ₁ : N S, X ₂ : A E, X ₃ : L V, X ₄ : A S, X ₅ : I V, X ₆ : L P, X ₇ : P R, X ₈ : D E, X ₉ : L V

RVD	Consensus Sequence	Variable Definition
SH	LX ₁ TX ₂ QVVAIASSHGGKQALEAVRALX ₃ X ₄ X ₅ LRAX ₆ PYX ₇ (SEQ ID NO: 270)	X ₁ : N S, X ₂ : A E, X ₃ : F L, X ₄ : P R, X ₅ : D E V, X ₆ : A T, X ₇ : A D G
NK	LX ₁ TEQVVAX ₂ ASNKGGKQX ₃ LX ₄ X ₅ VX ₆ AX ₇ LLX ₈ LX ₉ X ₁₀ X ₁₁ PYX ₁₂ (SEQ ID NO: 271)	X ₁ : N S, X ₁₀ : A G, X ₁₁ : A V, X ₁₂ : A E V, X ₂ : I V, X ₃ : A V, X ₄ : A E, X ₅ : A E, X ₆ : E G K, X ₇ : D H Q, X ₈ : A D E R, X ₉ : L R
HD	LSX ₁ X ₂ QVX ₃ AIAX ₄ HDGGX ₅ QX ₆ LEAX ₇ X ₈ X ₉ QLVX ₁₀ LX ₁₁ AAPYA (SEQ ID NO: 272)	X ₁ : A T, X ₂ : A E, X ₃ : A V, X ₄ : G S, X ₅ : K N, X ₆ : A P, X ₇ : A V, X ₈ : G V, X ₉ : A G T V, X ₁₀ : A E V, X ₁₁ : L R
RS	LSX ₁ AQVVAX ₂ AX ₃ RSGGKQALEAVRAQLL X ₄ LRAAPYG(SEQ ID NO: 273)	X ₁ : I T, X ₂ : I V, X ₃ : S T, X ₄ : A D
NH	LSX ₁ EQVVAIASNHGGKQALEAVRALFRX ₂ L RAAPYX ₃ (SEQ ID NO: 274)	X ₁ : P T, X ₂ : E G, X ₃ : A G
SI	LSTX ₁ QVX ₂ X ₃ IAAX ₄ SIGGX ₅ QALEAX ₆ KVQLP VLRAAPYX ₇ (SEQ ID NO: 275)	X ₁ : A E, X ₂ : A V, X ₃ : T V, X ₄ : N S, X ₅ : K R, X ₆ : L V, X ₇ : E G
ND	LX ₁ TAQVVAIASNDGGKQX ₂ LEX ₃ X ₄ X ₅ AQLL X ₆ LRAX ₇ PYE (SEQ ID NO: 276)	X ₁ : S T, X ₂ : A T, X ₃ : A E V, X ₄ : A V, X ₅ : E G, X ₆ : A V, X ₇ : A V
SN	LSTAQVVX ₁ X ₂ ASSNGGKQALEAVWALLPV LRATPYD (SEQ ID NO: 277)	X ₁ : A T, X ₂ : I V
NG	LSTX ₁ QVVAIAX ₂ NGGGX ₃ QALEX ₄ X ₅ X ₆ X ₇ QL X ₈ X ₉ LRX ₁₀ X ₁₁ PX ₁₂ X ₁₃ (SEQ ID NO: 278)	X ₁ : A E G, X ₂ : G S, X ₃ : K R, X ₄ : A G, X ₅ : I V, X ₆ : G R, X ₇ : E K, X ₈ : L Q R, X ₉ : A E K, X ₁₀ : A T, X ₁₁ : A V, X ₁₂ : H Y, X ₁₃ : E G
NT	LTPQQVVAIAX ₁ NTGGKX ₂ ALX ₃ AX ₄ X ₅ X ₆ QL X ₇ X ₈ LRX ₉ AX ₁₀ YX ₁₁ (SEQ ID NO: 279)	X ₁ : A S, X ₁₀ : P R, X ₁₁ : E G R, X ₂ : Q R, X ₃ : E G, X ₄ : I V, X ₅ : C R T, X ₆ : T V, X ₇ : P R, X ₈ : I V, X ₉ : A G

[0067] In some embodiments, the present disclosure provides a modular nucleic acid binding domain (e.g., RNBD, or MAP-NBD), wherein the modular nucleic acid binding domain comprises a repeat unit with a sequence of A₁₋₁₁X₁X₂B₁₄₋₃₅ (SEQ ID NO: 443), wherein A₁₋₁₁ comprises 11 amino acid residues and wherein each amino acid residue of A₁₋₁₁ can be any amino acid. In some embodiments, A₁₋₁₁ can be any amino acids in position 1 through position 11 of any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352. X₁X₂ comprises any repeat

variable diresidue (RVD) disclosed herein and comprises at least one amino acid at position 12 or position 13. As described herein, this RVD contacts and binds to a target nucleic acid base of a target site. Said RVD can be the RVD of any repeat unit disclosed herein, such as position 12 and position 13 of any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352. B₁₄₋₃₅ can comprise 22 amino acid residues and each amino acid residue of B₁₄₋₃₅ can be any amino acid. In some embodiments, B₁₄₋₃₅ can be any amino acid in position 14 through position 35 of any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352. In particular embodiments, a modular nucleic acid binding domain (e.g., RNBD or MAP-NBD) having the above sequence of A₁₋₁₁X₁X₂B₁₄₋₃₅ (SEQ ID NO: 443) can have a first repeat unit with at least one residue in A₁₋₁₁, B₁₄₋₃₅, or a combination thereof that differs from a corresponding residue in a second repeat unit in the modular nucleic acid binding domain (e.g., RNBD or MAP-NBD). In other words, at least two repeat units in a modular nucleic acid binding domain (e.g., RNBD or MAP-NBD) described herein can have different amino acid residues with respect to each other, at the same position outside the RVD region. Thus, in some embodiments, a modular nucleic acid binding domain (e.g., RNBD or MAP-NBD) described herein can have variant backbones with respect to each repeat unit in the plurality of repeat units that make up the modular nucleic acid binding domain. In some embodiments, an RNBD of the present disclosure can have a sequence of GGKQALEAVRAQLLDLRAAPYG (SEQ ID NO: 280) at B₁₄₋₃₅.

[0068] In some embodiments, the present disclosure provides a composition comprising a modular nucleic acid binding domain and a functional domain, wherein: the modular nucleic acid binding domain comprises a plurality of repeat units; at least one repeat unit of the plurality comprises a sequence of A₁₋₁₁X₁X₂B₁₄₋₃₅ (SEQ ID NO: 443); each amino acid residue of A₁₋₁₁ comprises any amino acid residue; X₁X₂ comprises a binding region configured to bind to a target nucleic acid base within a target site; each amino acid residue of B₁₄₋₃₅ comprises any amino acid; and a first repeat unit of the plurality of repeat units comprises at least one residue in A₁₋₁₁, B₁₄₋₃₅, or a combination thereof that differs from a corresponding residue in a second repeat unit of the plurality of repeat units. In some embodiments, the binding region comprises an amino acid residue at position 13 or an amino acid residue at position 12 and the amino acid residue at position 13. In further aspects, the amino acid residue at position 13 binds to the target nucleic acid base. In some aspects, the amino acid residue at position 12 stabilizes the configuration of the binding region.

[0069] In some embodiments, a modular nucleic acid binding sequence (e.g., RNBD) can comprise one or more of the following characteristics: the modular nucleic acid binding sequence (e.g., RNBD) can bind a nucleic acid sequence, wherein the target site comprises a 5' guanine, the

modular nucleic acid binding domain (e.g., RNBD) can comprise 7 repeat units to 25 repeat units, a first modular nucleic acid binding sequence (e.g., RNBD) can bind a target nucleic acid sequence and be separated from a second modular nucleic acid binding domain (e.g., RNBD) from 2 to 50 base pairs, or any combination thereof.

[0070] In some embodiments, an RNBD of the present disclosure can have the full length naturally occurring N-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein. In some embodiments, any truncation of the full length naturally occurring N-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein can be used at the N-terminus of an RNBD of the present disclosure. For example, in some embodiments, amino acid residues at positions 1 (H) to position 137 (F) of the naturally occurring *Ralstonia solanacearum*-derived protein N-terminus can be used. In particular embodiments, said truncated N-terminus from position 1 (H) to position 137 (F) can have a sequence as follows:

FGKLVALGYSREQIRKLKQESLSEIAKYHTTLTGQGFTHADICRISRRRQSLRVVARNYPELAAALPE
LTRAHIVDIARQRSGDLALQALLPVATALTAAPLRLSASQIATVAQYGERPAIQALYRLRRKLTRAPL

H (SEQ ID NO: 264). In some embodiments, the naturally occurring N-terminus of *Ralstonia solanacearum* can be truncated to any length and used at the N-terminus of the engineered DNA binding domain. For example, the naturally occurring N-terminus of *Ralstonia solanacearum* can be truncated to amino acid residues at position 1 (H) to position 120 (K) as follows:

KQESLSEIAKYHTTLTGQGFTHADICRISRRRQSLRVVARNYPELAAALPELTRAHIVDIARQRSGDL
ALQALLPVATALTAAPLRLSASQIATVAQYGERPAIQALYRLRRKLTRAPLH (SEQ ID NO: 303) and

used at the N-terminus of the RNBD. The naturally occurring N-terminus of *Ralstonia solanacearum* can be truncated amino acid residues at positions 1 to 115 and used at the N-terminus of the engineered DNA binding domain as set forth in SEQ ID NO: 20. The naturally occurring N-terminus of *Ralstonia solanacearum* can be truncated to amino acid residues at positions 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the N-terminus of the engineered DNA binding domain. Truncation of the N-termini can be particularly advantageous for obtaining DNA binding domains, which are smaller in size including number of amino acids and overall molecular weight. A reduced number of amino acids can allow for more efficient packaging into a viral vector and a smaller molecular weight can result in more efficient loading of the DNA binding domains in non-viral vectors for delivery.

[0071] In some embodiments, the N-terminus, referred to as the amino terminus or the “NH₂” domain, can recognize a guanine. In some embodiments, the N-terminus can be engineered to bind a cytosine, adenosine, thymidine, guanine, or uracil.

[0072] In some embodiments, an RNBD of the present disclosure can have a DNA binding domain, in which the final full length repeat unit of 33-35 amino acid residues is followed by a half-repeat also derived from *Ralstonia solanacearum*. The half repeat can have 15 to 23 amino acid residues, for example, the half repeat can have 19 amino acid residues. In particular embodiments, the half-repeat can have a sequence as follows: LSTAQVVAIACISGQQALE (SEQ ID NO: 265).

[0073] In some embodiments, an RNBD of the present disclosure can have the full length naturally occurring C-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein. In some embodiments, any truncation of the full length naturally occurring C-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein can be used at the C-terminus of an RNBD of the present disclosure. For example, in some embodiments, the RNBD can comprise amino acid residues at position 1 (A) to position 63 (S) as follows:

AIEAHMPTLRQASHSLSPERVAIAICIGGRSAVEAVRQGLPVKAIRRIKREKAPVAGPPPAS (SEQ ID NO: 266) of the naturally occurring *Ralstonia solanacearum*-derived protein C-terminus. In some embodiments, the naturally occurring C-terminus of *Ralstonia solanacearum* can be truncated to any length and used at the C-terminus of the RNBD. For example, the naturally occurring C-terminus of *Ralstonia solanacearum* can be truncated to amino acid residues at positions 1 to 63 and used at the C-terminus of the RNBD. The naturally occurring C-terminus of *Ralstonia solanacearum* can be truncated amino acid residues at positions 1 to 50 and used at the C-terminus of the RNBD. The naturally occurring C-terminus of *Ralstonia solanacearum* can be truncated to amino acid residues at positions 1 to 63, 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the C-terminus of the RNBD.

[0074] TABLE 4 shows N-termini, C-termini, and half-repeats derived from *Ralstonia*.

TABLE 4: *Ralstonia*-Derived N-terminus, C-terminus, and Half-Repeat

SEQ ID NO	Description	Sequence
SEQ ID NO: 315	Truncated N-terminus; positions 1 (H) to 115 (S) of the naturally occurring <i>Ralstonia solanacearum</i> -derived protein N-terminus	SEIAKYHTTLTGQGFTHADICRISRRRQSLRVARNYPELAAALPELTRAHIVDIARQRSGDLALQALLPVATALTAAPLRLSASQIATVAQYGERPAIQALYRLRRKLTRAPLH
SEQ ID NO: 264	Truncated N-terminus; positions 1 (H) to 137 (F) of the naturally occurring <i>Ralstonia solanacearum</i> -derived protein N-terminus	FGKLVALGYSREQIRKLKQESLSEIAKYHTTLTGQGFTHADICRISRRRQSLRVARNYPELAAALPELTRAHIVDIARQRSGDLALQALLPVATALTAAPLRLSASQIATVAQYGERPAIQALYRLRRKLTRAPLH
SEQ ID NO: 303	Truncated N-terminus; positions 1 (H) to 120 (K) of the naturally occurring <i>Ralstonia solanacearum</i> -derived protein N-terminus	KQESLSEIAKYHTTLTGQGFTHADICRISRRRQSLRVARNYPELAAALPELTRAHIVDIARQRSGDLALQALLPVATALTAAPL

SEQ ID NO	Description	Sequence
		RLSASQIATVAQYGERPAIQALYRLRRK LTRAPLH
SEQ ID NO: 265	Half-repeat	LSTAQVVAIACISGQQALE
SEQ ID NO: 266	Truncated C-terminus; positions 1 (A) to 63 (S) of the naturally occurring <i>Ralstonia solanacearum</i> -derived protein C-terminus	AIEAHMPTLRQASHSLSPERVAAIACIGG RSAVEAVRQGLPVKAIRIRREKAPVAG PPPAS

B. *Xanthomonas* Derived Transcription Activator Like Effector (TALE)

[0075] The present disclosure provides non-naturally occurring fusion proteins of a nuclease (e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81) to a modular *Xanthomonas*-derived transcription activator-like effector (TALE) protein.

[0076] The present disclosure provides a modular nucleic acid binding domain derived from *Xanthomonas spp.*, also referred to herein as a transcription activator-like effector (TALE) protein, can comprise a plurality of repeat units. A repeat unit of the plurality of repeat units recognizes a single target nucleotide, base pair, or both. A repeat unit from *Xanthomonas spp.* can comprise 33-35 amino acid residues. In some embodiments, a repeat unit can be from *Xanthomonas spp.* and have a sequence of

MDPIRSRTPSPARELLPGPQPDGVQPTADRGVSPAGGPLDGLPARRTMSRTRLSPSPAPSPA
FSAGSFSDLLRQFDPSLFNTSLFDSLPPFGAHHTEAATGEWDEVQSGLRAADAPPPTMRVAV
TAARPPRAKPAPRRRAAQPSDASPAQVDLRTLGYSSQQQEKIKPKVIRSTVAQHHEALVG
HGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVA
GELRGPPQLDGTGQLLKIARKGGVTAVEAVHAWRNALTGAPLNLTPEQVVAIASHDGGKQ
ALETVQRLLPVLCQAHGLTPQQVVAIASNGGGKQALETVQRLLPVLCQAHGLTPQQVVAI
SNSGGKQALETVQRLLPVLCQAHGLTPEQVVAIASNGGGKQALETVQRLLPVLCQAHGLTP
EQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASNIGGKQALETVQALLPVLCQ
AHGLTPEQVVAIASNIGGKQALETVQALLPVLCQAHGLTPEQVVAIASHDGGKQALETVQR
LLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPQQVVAIASNGGGK
QALETVQRLLPVLCQAHGLTPEQVVAIASNSGGKQALETVQALLPVLCQAHGLTPEQVVAI
ASNSGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPVLCQAHGL
TPEQVVAIASHDGGKQALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALETVQRLLPV
LCQAHGLTPQQVVAIASNGGGRPALETVQRLLPVLCQAHGLTPEQVVAIASHDGGKQALET
VQRLLPVLCQAHGLTPQQVVAIASNGGGRPALESIVAQLSRPDPALAALTN DHLVALACLG
GRPALDAVKKGLPHAPALIKRTNRRIPERTSHRVADHAQVVRVLGFFQCHSHPAQAFDDA
MTQFGMSRHLLQLFRRVGVTELEARSGTLPPASQRWDRILQASGMKRAKPSPTSTQTPDQ

ASLHAFADSLERDLDA P SPMHEGDQTRASSRKRSRSDRAVTGPSAQQSFEVRVPEQRDALH
LPLSWRVKRPRTSIGGGLPDPGTPTAADLAASSTVMREQDEDPFAGAADDFPAFNEEELAW
LMELLPQ (SEQ ID NO: 299).

[0077] In some embodiments, a TALE of the present disclosure can comprise between 1 to 50 *Xanthomonas spp.*-derived repeat units. In some embodiments, a TALE of the present disclosure can comprise between 9 and 36 *Xanthomonas spp.*-derived repeat units. Preferably, in some embodiments, a TALE of the present disclosure can comprise between 12 and 30 *Xanthomonas spp.*-derived repeat units. A TALE described herein can comprise between 5 to 10 *Xanthomonas spp.*-derived repeat units, between 10 to 15 *Xanthomonas spp.*-derived repeat units, between 15 to 20 *Xanthomonas spp.*-derived repeat units, between 20 to 25 *Xanthomonas spp.*-derived repeat units, between 25 to 30 *Xanthomonas spp.*-derived repeat units, or between 30 to 35 *Xanthomonas spp.*-derived repeat units, between 35 to 40 *Xanthomonas spp.*-derived repeat units. A TALE described herein can comprise 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 *Xanthomonas spp.*-derived repeat units.

[0078] A *Xanthomonas spp.*-derived repeat units can be derived from a wild-type repeat unit, such as any one of SEQ ID NO: 318 – SEQ ID NO: 321. For example, a *Xanthomonas spp.*-derived repeat units can have a sequence of LTPDQVVAIASNHGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 318) comprising an RVD of NH, which recognizes guanine. A *Xanthomonas spp.*-derived repeat units can have a sequence of LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 319) comprising an RVD of NG, which recognizes thymidine. A *Xanthomonas spp.*-derived repeat units can have a sequence of LTPDQVVAIASNIGGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 320) comprising an RVD of NI, which recognizes adenosine. A *Xanthomonas spp.*-derived repeat units can have a sequence of LTPDQVVAIASHDGGKQALETVQRLLPVLCQDHG (SEQ ID NO: 321) comprising an RVD of HD, which recognizes cytosine.

[0079] A *Xanthomonas spp.*-derived repeat unit can also comprise a modified *Xanthomonas spp.*-derived repeat units enhanced for specific recognition of a nucleotide or base pair. A TALE described herein can comprise one or more wild-type *Xanthomonas spp.*-derived repeat units, one or more modified *Xanthomonas spp.*-derived repeat units, or a combination thereof. In some embodiments, a modified *Xanthomonas spp.*-derived repeat units can comprise 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, or 29 mutations that can enhance recognition of a specific nucleotide or base pair. In some embodiments, a modified *Xanthomonas spp.*-derived repeat unit can comprise more than 1 modification, for example 1 to 5 modifications, 5 to 10 modifications, 10 to 15 modifications, 15 to 20 modifications, 20 to 25

modification, or 25-29 modifications. In some embodiments, A TALE can comprise more than one modified *Xanthomonas spp.*-derived repeat units, wherein each of the modified *Xanthomonas spp.*-derived repeat units can have a different number of modifications.

[0080] In some embodiments, a TALE of the present disclosure can have the full length naturally occurring N-terminus of a naturally occurring *Xanthomonas spp.*-derived protein, such as the N-terminus of SEQ ID NO: 299. In some embodiments, a TALE of the present disclosure can comprise the amino acid residues at position 1 (N) through position 137 (M) of the naturally occurring *Xanthomonas spp.*-derived protein as follows:

MVDLRTLGYSSQQQKEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAGRGGVTA
VEAVHAWRNALTGAPLN (SEQ ID NO: 300). In some embodiments, the N-terminus can be truncated to position 1 (N) through position 120 (K) of the naturally occurring *Xanthomonas spp.*-derived protein as follows:

KPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLN (SEQ ID NO: 301). In some embodiments, the N-terminus can be truncated to position 1 (N) through position 115 (S) of the naturally occurring *Xanthomonas spp.*-derived protein as follows:
STVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAGRGGVTAVEAVHAWRNALTGAPLN (SEQ ID NO: 316). In some embodiments, any truncation of the naturally occurring *Xanthomonas spp.*-derived protein can be used at the N-terminus of a TALE disclosed herein. The naturally occurring N-terminus of *Xanthomonas spp.* can be truncated to amino acid residues at positions 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the N-terminus of the TALE.

[0081] In some embodiments, a TALE of the present disclosure can have a DNA binding domain, in which the final full length repeat unit of 33-35 amino acid residues is followed by a half-repeat also derived from *Xanthomonas spp.* The half repeat can have 15 to 23 amino acid residues, for example, the half repeat can have 19 amino acid residues. In particular embodiments, the half-repeat can have a sequence as set forth in LTPQQVVVAIASNGGGRPALE (SEQ ID NO: 297), SEQ ID NO: 327, 328, 329, 330, 331, 332, 333, or 334.

TABLE 5: *Xanthomonas* Repeat Sequences

SEQ ID NO	Amino Acid Sequence	Description
SEQ ID NO: 323	LTPDQVVVAIASNHGGKQALETVQRLLPVLCQDHG	RVD of NH recognizing guanine
SEQ ID NO: 324	LTPDQVVVAIASNNGGKQALETVQRLLPVLCQDHG	RVD of NH recognizing thymidine
SEQ ID NO: 325	LTPDQVVVAIASNIGGKQALETVQRLLPVLCQDHG	RVD of NI recognizing adenosine
SEQ ID NO: 326	LTPDQVVVAIASHDGGKQALETVQRLLPVLCQDHG	RVD of HD recognizing cytosine
SEQ ID NO: 297	LTPQQVVAIASNNGGGRPALE	Half repeat
SEQ ID NO: 327	LTPEQVVAIASNNGGGRPALE	Half repeat
SEQ ID NO: 328	LTPDQVVAIASNNGGGRPALE	Half repeat
SEQ ID NO: 329	LTPEQVVAIASNIGGGRPALE	Half repeat
SEQ ID NO: 330	LTPDQVVAIASNIGGGRPALE	Half repeat
SEQ ID NO: 331	LTPEQVVAIASHDGGRPALE	Half repeat
SEQ ID NO: 332	LTPDQVVAIASHDGGRPALE	Half repeat
SEQ ID NO: 333	LTPEQVVAIASNHGGRPALE	Half repeat
SEQ ID NO: 334	LTPDQVVAIASNHGGRPALE	Half repeat

[0082] In some embodiments, a TALE of the present disclosure can have the full length naturally occurring C-terminus of a naturally occurring *Xanthomonas spp.*-derived protein, such as the C-terminus of SEQ ID NO: 299. In some embodiments, the C-terminus can be positions 1 (S) through position 278 (Q) of the naturally occurring *Xanthomonas spp.*-derived protein as follows:

SIVAQLSRPDPALAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRV
ADHAQVVRVLGFFQCHSHPAQAFDDAMTQFGMSRHGLLQLFRRVGVTELEARSGTLPPAS
QRWDRILQASGMKRAKPSPTSTQTPDQASLHAFADSLERDLDA PSPTHEGDQRRASSRKRS
RSDRAVTGPSAQQSFEVRAPEQRDALHLPLSWRVKRPRTSIGGGLPDPGTPTAADLAASSTV
MREQDEDPFAGAADDFFPAFNEEELAWLMELL PQ (SEQ ID NO: 302). In some embodiments, any truncation of the full length naturally occurring C-terminus of a naturally occurring *Xanthomonas spp.* -derived protein can be used at the C-terminus of a TALE of the present

disclosure. For example, in some embodiments, the naturally occurring N-terminus of *Xanthomonas spp.* can be truncated to amino acid residues at position 1 (S) to position 63 (X) as follows:

SIVAQLSRPDPALAAALTNDHLVALACLGGRPALDAVKKGLPHAPALIKRTNRRIPERTSHRV A (SEQ ID NO: 298). The naturally occurring C-terminus of *Xanthomonas spp.* can be truncated amino acid residues at positions 1 to 50 and used at the C-terminus of the engineered DNA binding domain. The naturally occurring C-terminus of *Xanthomonas spp.* can be truncated to amino acid residues at positions 1 to 63, 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the C-terminus of the engineered DNA binding domain.

C. Animal Pathogen Derived Nucleic Acid Binding Domains (MAP-NBDs)

[0083] The present disclosure provides non-naturally occurring fusion proteins of a nuclease (e.g., any one of SEQ ID NO: 1 – SEQ ID NO: 81) to a modular animal pathogen-derived nucleic acid binding domain (MAP-NBD).

[0084] The present disclosure provides a modular nucleic acid binding domain derived from an animal pathogen protein (MAP-NBD) can comprise a plurality of repeat units, wherein a repeat unit of the plurality of repeat units recognizes a single target nucleotide, base pair, or both.

[0085] In some embodiments, the repeat unit can be derived from an animal pathogen, and can be referred to as a non-naturally occurring modular nucleic acid binding domain derived from an animal pathogen protein (MAP-NBD), or “modular animal pathogen-nucleic acid binding domain” (MAP-NBD). For example, in some cases, the animal pathogen can be from the Gram-negative bacterium genus, *Legionella*. In other cases, the animal pathogen can be from *Burkholderia*. In some cases, the animal pathogen can be from *Paraburkholderia*. In other cases, the animal pathogen can be from *Francisella*.

[0086] In particular embodiments, the repeat unit can be derived from a species of the genus of *Legionella*, such as *Legionella quateirensis*, the genus of *Burkholderia*, the genus of *Paraburkholderia*, or the genus of *Francisella*. In some embodiments, the repeat unit can comprise from 19 amino acid residues to 35 amino acid residues. In particular embodiments, the repeat unit can comprise 33 amino acid residues. In other embodiments, the repeat unit can comprise 35 amino acid residues. In some embodiments, the MAP-NBD is non-naturally occurring, and comprises a plurality of repeat units and wherein a repeat unit of the plurality of repeat units recognizes a single target nucleic acid.

[0087] In some embodiments, a repeat unit can be derived from a *Legionella quateirensis* protein with the following sequence:

MPDLELNFAIPLHLFDDTVFTHDATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSK
 EPANTKSRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSA
 VQQLLCKHEELNLISPDGLGHKELIKIAARNGGGNNLIAVLSCYAKLKEMGFSSQQIIRMV
 SHAGGANLKA V T A N H D D L Q N M G F N V E Q I V R M V S H N G G S K N L K A V T D N H D D L K N M G F N
 A E Q I V R M V S H G G G S K N L K A V T D N H D D L K N M G F N A E Q I V S M V S N N G G S K N L K A V T D N H D D
 L K N M G F N A E Q I V S M V S N G G G S L N L K A V K K Y H D A L K D R G F N T E Q I V R M V S H D G G S L N L K A
 V K K Y H D A L R E R K F N V E Q I V S I V S H G G G S L N L K A V K K Y H D V L K D R E F N A E Q I V R M V S H D G G
 S L N L K A V T D N H D D L K N M G F N A E Q I V R M V S H K G G S K N L A L V K E Y F P V F S S F H F T A D Q I V A L I
 C Q S K Q C F R N L K K N H Q Q W K N K G L S A E Q I V D L I L Q E T P P K P N F N N T S S S T P S P S A P S F F Q G P S T P
 I P T P V L D N S P A P I F S N P V C F F S S R S E N N T E Q Y L Q D S T L D L D S Q L G D P T K N F N V N N F W S L F P F D
 D V G Y H P H S N D V G Y H L H S D E E S P F F D F (SEQ ID NO: 281).

[0088] In some embodiments, a repeat from a *Legionella quateirensis* protein can comprise a repeat with a canonical RVD or a non-canonical RVD. In some embodiments, a canonical RVD can comprise NN, NG, HD, or HD. In some embodiments, a non-canonical RVD can comprise RN, HA, HN, HG, HG, or HK.

[0089] In some embodiments, a repeat of SEQ ID NO: 282 comprises an RVD of HA and primarily recognizes a base of adenine (A). In some embodiments, a repeat of SEQ ID NO: 283 comprises an RVD of HN and recognizes a base comprising guanine (G). In some embodiments, a repeat of S SEQ ID NO: 284 comprises an RVD of HG and recognizes a base comprising thymine (T). In some embodiments, a repeat of SEQ ID NO: 285 comprises an RVD of NN and recognizes a base comprising guanine (G). In some embodiments, a repeat of SEQ ID NO: 286 comprises an RVD of NG and recognizes a base comprising thymine (T). In some embodiments, a repeat of SEQ ID NO: 287 comprises an RVD of HD and recognizes a base comprising cytosine (C). In some embodiments, a repeat of SEQ ID NO: 288 comprises an RVD of HG and recognizes a base comprising thymine (T). In some embodiments, a repeat of SEQ ID NO: 289 comprises an RVD of HD and recognizes a base comprising cytosine (C). In some embodiments, a half-repeat of SEQ ID NO: 290 comprises an RVD of HK and recognizes a base comprising guanine (G). In some embodiments, a repeat of SEQ ID NO: 353 comprises an RVD of RN and recognizes a base comprising guanine (G).

[0090] **TABLE 6** illustrates exemplary repeats from *Legionella quateirensis*, *Burkholderia*, *Paraburkholderia*, or *Francisella* that can make up a MAP-NBD of the present disclosure and the RVD at position 12 and 13 of the particular repeat. A MAP-NBD of the present disclosure can comprise at least one of the repeats disclosed in **TABLE 5** including any one of SEQ ID NO: 353,

SEQ ID NO: 282 – SEQ ID NO: 290, or SEQ ID NO: 354 – SEQ ID NO: 442. A MAP-NBD of the present disclosure can comprise any combination of repeats disclosed in **TABLE 5** including any one of SEQ ID NO: 353, SEQ ID NO: 282 – SEQ ID NO: 290, or SEQ ID NO: 354 – SEQ ID NO: 442.

TABLE 6 – Animal Pathogen Derived Repeat Units

SEQ ID NO	Organism	Repeat Unit Sequence	RVD
353	<i>L. quateirensis</i>	LGHKELIKIAARNGGGNNLIAVLSCYAKLKEMG	RN
282	<i>L. quateirensis</i>	FSSQQIIRMVSHAGGANNLKAVTANHDDLQNMG	HA
283	<i>L. quateirensis</i>	FNVEQIVRMVSHNGGSKNLKAVTDNHDDLKNMG	HN
284	<i>L. quateirensis</i>	FNAEQIVRMVSHGGGSKNLKAVTDNHDDLKNMG	HG
285	<i>L. quateirensis</i>	FNAEQIVSMVSNNGGSKNLKAVTDNHDDLKNMG	NN
286	<i>L. quateirensis</i>	FNAEQIVSMVSNGGGSLNLKAVKKYHDALKDRG	NG
287	<i>L. quateirensis</i>	FNTEQIVRMVSHDGGSLNLKAVKKYHDALRERK	HD
288	<i>L. quateirensis</i>	FNVEQIVSIVSHGGGSLNLKAVKKYHDLKDRE	HG
289	<i>L. quateirensis</i>	FNAEQIVRMVSHDGGSLNLKAVTDNHDDLKNMG	HD
290	<i>L. quateirensis</i>	FNAEQIVRMVSHKGGSKNL	HK
354	<i>L. quateirensis</i>	FSAEQIVRIA AHDGGSRNIEAVQQAQHVLKELG	HD
355	<i>L. quateirensis</i>	FSAEQIVSIV AHDGGSRNIEAVQQAQHILKELG	HD
356	<i>L. quateirensis</i>	FSRQQILRIASHDGGSKNIAAVQKFLPKLMNFGFN	HD
357	<i>L. quateirensis</i>	FSAEQIVRIA AHDGGSLNIDAVQQAQQALKELG	HD
358	<i>L. quateirensis</i>	FSTEQIVCIAGHGGGSLNIKAVLLAQQALKDLG	HG
359	<i>L. quateirensis</i>	FSSEQIVRVA AHGGGSLNIKAVLQAHQALKELD	HG
360	<i>L. quateirensis</i>	FSAEQIVHIA AHGGGSLNIKAILQAHQTLKELN	HG
361	<i>L. quateirensis</i>	FSAEQIVRIA AHIGGSRNIEAIQQAHHALKELG	HI
362	<i>L. quateirensis</i>	FSAEQIVRIA AHIGGSHNLKAVLQAQQALKELD	HI
363	<i>L. quateirensis</i>	FSAKHIVRIA AHIGGSLNIKAVQQAQQALKELG	HI
364	<i>L. quateirensis</i>	FNAEQIVRMVSHKGGSKNLALVKEYFPVFSSFH	HK
365	<i>L. quateirensis</i>	FNAEQIVRMVSHKGGSKNLALVKEYFPVFSSFHFT	HK
366	<i>L. quateirensis</i>	FSADQIVRIA AHKGGSHNIVAVQQAQQALKELD	HK
367	<i>L. quateirensis</i>	FNVEQIVRMVSHNGGSKNLKAVTDNHDDLKNMGFN	HN
368	<i>L. quateirensis</i>	FSADQVVKIAGHSGGSNNIAVMLAVFPRLRDFGFK	HS
369	<i>L. quateirensis</i>	FSAEQIVSIA AHVGGSHNIEAVQKAHQALKELD	HV
370	<i>L. quateirensis</i>	FNAEQIVSMVSNNGGSKNLKAVTDNHDDLKNMGFN	NN
371	<i>L. quateirensis</i>	FSHKELIKIAARNGGGNNLIAVLSCYAKLKEMG	RN
372	<i>L. quateirensis</i>	FSHKELIKIAARNGGGNNLIAVLSCYAKLKEMGFS	RN
373	<i>Burkholderia</i>	FSSGETVGATVGAGGTETVAQGGTASNTTVSSGGY	GA

SEQ ID NO	Organism	Repeat Unit Sequence	RVD
374	<i>Burkholderia</i>	FSGGMATSTTVGSGGTQDVLAGGAAVGGTVGTGGV	GS
375	<i>Burkholderia</i>	FSAADIVKIAGKIGGAQALQAFITHRAALIQAGFS	KI
376	<i>Burkholderia</i>	FNPTDIVKIAGNDGGAQALQAVLELEPALRERGF	ND
377	<i>Burkholderia</i>	FNPTDIVRMAGNDGGAQALQAVFELEPAFRERSFS	ND
378	<i>Burkholderia</i>	FNPTDIVRMAGNDGGAQALQAVLELEPAFRERGF	ND
379	<i>Burkholderia</i>	FSQVDIVKIASNDGGAQALYSVLDVEPTFRERGF	ND
380	<i>Burkholderia</i>	FSRADIVKIAGNDGGAQALYSVLDVEPPLRERGF	ND
381	<i>Burkholderia</i>	FSRGDIVKIAGNDGGAQALYSVLDVEPPLRERGF	ND
382	<i>Burkholderia</i>	FNRADIVRIAGNGGGAQALYSVRDAGPTLGKRGF	NG
383	<i>Burkholderia</i>	FRQADIVKIASNGGSAQALNAVIKLGPTLRQRGF	NG
384	<i>Burkholderia</i>	FRQADIVKMASNGGSAQALNAVIKLGPTLRQRGF	NG
385	<i>Burkholderia</i>	FSRADIVKIAGNGGGAQALQAVLELEPTFRERGF	NG
386	<i>Burkholderia</i>	FSRADIVRIAGNGGGAQALYSVLDVGP TLGKRGF	NG
387	<i>Burkholderia</i>	FSRGDIVRIAGNGGGAQALQAVLELEPTLGERGF	NG
388	<i>Burkholderia</i>	FSRADIVKIAGNGGGAQALQAVITHRAALTQAGF	NG
389	<i>Burkholderia</i>	FSRGDTVKIAGNIGGAQALQAVLELEPTLRERGF	NI
390	<i>Burkholderia</i>	FNPTDIVKIAGNIGGAQALQAVLELEPAFRERGF	NI
391	<i>Burkholderia</i>	FSAADIVKIAGNIGGAQALQAFITHRAALIQAGF	NI
392	<i>Burkholderia</i>	FSAADIVKIAGNIGGAQALQAVITHRATLTQAGF	NI
393	<i>Burkholderia</i>	FSATDIVKIASNIGGAQALQAVISRRAALIQAGF	NI
394	<i>Burkholderia</i>	FSQPDIVKIAGNIGGAQALQAVLELEPAFRERGF	NI
395	<i>Burkholderia</i>	FSRADIVKIAGNIGGAQALQAVLELESTFRERSFN	NI
396	<i>Burkholderia</i>	FSRADIVKIAGNIGGAQALQAVLELESTLRERSFN	NI
397	<i>Burkholderia</i>	FSRGDIVKMAGNIGGAQALQAGLELEPAFRERGF	NI
398	<i>Burkholderia</i>	FSRGDIVKMAGNIGGAQALQAVLELEPAFHRSFC	NI
399	<i>Burkholderia</i>	FTLTDIVKMAGNIGGAQALKAVLEHGPTLRQRDLS	NI
400	<i>Burkholderia</i>	FTLTDIVKMAGNIGGAQALKVVLEHGPTLRQRDLS	NI
401	<i>Burkholderia</i>	FNPTDIVKIAGNNGGAQALQAVLELEPALRERGF	NN
402	<i>Burkholderia</i>	FNPTDIVKIAGNNGGAQALQAVLELEPALRERSFS	NN
403	<i>Burkholderia</i>	FNPTDMVKIAGNNGGAQALQAVLELEPALRERGF	NN
404	<i>Burkholderia</i>	FSAADIVKIASNNGGAQALQALIDHWSTLSGKTKA	NN
405	<i>Burkholderia</i>	FSAADIVKIASNNGGAQALQAVISRRAALIQAGF	NN
406	<i>Burkholderia</i>	FSAADIVKIASNNGGAQALQAVITHRAALAQAGF	NN
407	<i>Burkholderia</i>	FSAADIVKIASNNGGARALQALIDHWSTLSGKTKA	NN
408	<i>Burkholderia</i>	FTLTDIVEMAGNNGGAQALKAVLEHGSTLDERGFT	NN
409	<i>Burkholderia</i>	FTLTDIVKMAGNNGGAQALKAVLEHGPTLDERGFT	NN
410	<i>Burkholderia</i>	FTLTDIVKMAGNNGGAQALKVVLEHGPTLRQRGF	NN
411	<i>Burkholderia</i>	FTLTDIVKMASNNGGAQALKAVLEHGPTLDERGFT	NN
412	<i>Burkholderia</i>	FSAADIVKIAGNSGGAQALQAVISHRAALTQAGF	NS
413	<i>Burkholderia</i>	FSGGDAVSTVVRSGGAQSVASGGTASGTTVSAGAT	RS
414	<i>Burkholderia</i>	FRQTDIVKMAGSGGSAQALNAVIKHGPTLRQRGF	SG
415	<i>Burkholderia</i>	FSLIDIVEIASNNGGAQALKAVLKYGPVLTQAGRS	SN
416	<i>Burkholderia</i>	FSGGDAAGTVVSSGGAQNV TGGLASGTTVASGGAA	SS
417	<i>Paraburkholderia</i>	FNLTDIVEMAANSGGAQALKAVLEHGPTLRQRGLS	NS
418	<i>Paraburkholderia</i>	FNRASIVKIAGNSGGAQALQAVLKHGPTLDERGFN	NS

SEQ ID NO	Organism	Repeat Unit Sequence	RVD
419	<i>Paraburkholderia</i>	FSQANIVKMAGNSGGAQALQAVLDLELVFRERGF	NS
420	<i>Paraburkholderia</i>	FSQPDIVKMAGNSGGAQALQAVLDLELAFRERGF	NS
421	<i>Paraburkholderia</i>	FSLIDIVEIASNGGAQALKAVLKYGPVLMQAGRS	SN
422	<i>Francisella</i>	YKSEDIIRLASHDGGSVNLEAVLRLHSQRLG	HD
423	<i>Francisella</i>	YKPEDIIRLASHGGGSVNLEAVLRLNPQLIGL	HG
424	<i>Francisella</i>	YKSEDIIRLASHGGGSVNLEAVLRLHSQRLG	HG
425	<i>Francisella</i>	YKSEDIIRLASHGGGSVNLEAVLRLNPQLIGL	HG
426	<i>Paraburkholderia</i>	FNLTDIVEMAGKGGGAQALKAVLEHGPTLRQGF	KG
427	<i>Paraburkholderia</i>	FRQADIIRKIAGNDGGAQALQAVIEHGPTLRQHGF	ND
428	<i>Paraburkholderia</i>	FSQADIVKIAGNDGGTQALHAVLDLERMLGERGF	ND
429	<i>Paraburkholderia</i>	FSRADIVKIAGNGGGAQALKAVLEHEATLDERGF	NG
430	<i>Paraburkholderia</i>	FSRADIVRIAGNGGGAQALYSVLDVEPTLGKRGF	NG
431	<i>Paraburkholderia</i>	FSQPDIVKMASNIGGAQALQAVLELEPALRERGF	NI
432	<i>Paraburkholderia</i>	FSQPDIVKMAGNIGGAQALQAVLSLGPALRERGF	NI
433	<i>Paraburkholderia</i>	FSQPEIVKIAGNIGGAQALHTVLELEPTLHKRGF	NI
434	<i>Paraburkholderia</i>	FSQSDIVKIAGNIGGAQALQAVLDLESMLGKRGF	NI
435	<i>Paraburkholderia</i>	FSQSDIVKIAGNIGGAQALQAVLELEPTLRESDF	NI
436	<i>Paraburkholderia</i>	FNPTDIVKIAGNKGGAQALQAVLELEPALRERGF	NK
437	<i>Paraburkholderia</i>	FSPTDIIRKIAGNNGGAQALQAVLDLELMLRERGF	NN
438	<i>Paraburkholderia</i>	FSQADIVKIAGNNGGAQALYSVLDVEPTLGKRGF	NN
439	<i>Paraburkholderia</i>	FSRGDIVTIAGNNGGAQALQAVLELEPTLRERGF	NN
440	<i>Paraburkholderia</i>	FSRIDIVKIAANNNGGAQALHAVLDLGPTLRECGF	NN
441	<i>Paraburkholderia</i>	FSQADIVKIVGNNGGAQALQAVFELEPTLRERGF	NN
442	<i>Paraburkholderia</i>	FSQPDIVRITGNRGGGAQALQAVLALELTLRERGF	NR

[0091] In any one of the animal pathogen-derived repeat domains of SEQ ID NO: 353, SEQ ID NO: 282 – SEQ ID NO: 290, or SEQ ID NO: 354 – SEQ ID NO: 442, there can be considerable sequence divergence between repeats of a MAP-NBD outside of the RVD.

[0092] In some embodiments, a MAP-NBD of the present disclosure can comprise between 1 to 50 animal pathogen-derived repeat units. In some embodiments, a MAP-NBD of the present disclosure can comprise between 9 and 36 animal pathogen-derived repeat units. In some embodiments, a MAP-NBD of the present disclosure can comprise between 12 and 30 animal pathogen-derived repeat units. A MAP-NBD described herein can comprise between 5 to 10, between 10 to 15, between 15 to 20, between 20 to 25, between 25 to 30, between 30 to 35, or between 35 to 40 animal pathogen-derived repeat units. A MAP-NBD described herein can comprise 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 animal pathogen-derived repeat units.

[0093] An animal pathogen-derived repeat unit can be derived from a wild-type repeat unit, such as any one of SEQ ID NO: 353, SEQ ID NO: 282 – SEQ ID NO: 290, or SEQ ID NO: 354 – SEQ ID

NO: 442. A animal pathogen-derived repeat unit can also comprise a modified animal pathogen-derived repeat units enhanced for specific recognition of a nucleotide or base pair. A MAP-NBD described herein can comprise one or more wild-type animal pathogen-derived repeat units, one or more modified animal pathogen-derived repeat units, or a combination thereof. In some embodiments, a modified animal pathogen-derived repeat units can comprise 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, or 29 mutations that can enhance recognition of a specific nucleotide or base pair. In some embodiments, a modified animal pathogen-derived repeat unit can comprise more than 1 modification, for example 1 to 5 modifications, 5 to 10 modifications, 10 to 15 modifications, 15 to 20 modifications, 20 to 25 modification, or 25-29 modifications. In some embodiments, A MAP-NBD can comprise more than one modified animal pathogen-derived repeat units, wherein each of the modified animal pathogen-derived repeat units can have a different number of modifications.

[0094] In some embodiments, a MAP-NBD of the present disclosure can have the full length naturally occurring N-terminus of a naturally occurring *Legionella quateirensis*-derived protein, such as the N-terminus of SEQ ID NO: 281. A N-terminus can be the full length N-terminus sequence and can have a sequence of

MPDLELNFAIPLHLFDDETVFTHDATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSK
EPANTKSRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSA
VQQLLCKHEELLNLISPDG (SEQ ID NO: 291). In some embodiments, any truncation of SEQ ID NO: 291 can be used as the N-terminus in a MAP-NBD of the present disclosure. For example, in some embodiments, a MAP-NBD comprises a truncated N-terminus including amino acid residues at position 1 (G) to position 137 (S) of the naturally occurring *Legionella quateirensis* N-terminus as follows:

NFAIPLHLFDDETVFTHDATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSKEPANTK
SRRANSQNNKLSLADRLTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSAVQQLLCK
KHEELLNLISPDG (SEQ ID NO: 331). For example, in some embodiments, a MAP-NBD comprises a truncated N-terminus including amino acid residues at position 1 (G) to position 120 (S) of the naturally occurring *Legionella quateirensis* N-terminus as follows:

DATNDNSQASSSYSSKSSPASANARKRTSRKEMSGPPSKEPANTKSRRANSQNNKLSLADR
LTKYNIDEEFYQTRSDSLLSLNYTKKQIERLILYKGRTSAVQQLLCKHEELLNLISPDG (SEQ
ID NO: 304). In some embodiments, a MAP-NBD comprises a truncated N-terminus including amino acid residues at position 1 (G) to position 115 (K) of the naturally occurring *Legionella quateirensis* N-terminus as follows:

NSQASSSYSSKSSPASANARKRTSRKEMSGPPSKEPANTKSRRANSQNNKLSLADRLTKYNI DEEFYQTRSDSLSLNYTKKQIERLILYKGRTSAVQQLCKHEELLNLISPDG (SEQ ID NO: 317). In some embodiments, any truncation of the naturally occurring *Legionella quateirensis*-derived protein can be used at the N-terminus of a DNA binding domain disclosed herein. The naturally occurring N-terminus of *Legionella quateirensis* can be truncated to amino acid residues at positions 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the N-terminus of the MAP-NBD.

[0095] In some embodiments, a MAP-NBD of the present disclosure can have the full length naturally occurring C-terminus of a naturally occurring *Legionella quateirensis*-derived protein. In some embodiments, A MAP-NBD of the present disclosure can have at its C-terminus amino acid residues at position 1 (A) to position 176 (F) of the naturally occurring *Legionella quateirensis*-derived protein as follows:

ALVKEYFPVFSSFHFTADQIVALICQSKQCFRNLKKNHQQWKNKGLSAEQIVDLILQETPPK PNFNNTSSSTSPSPAPSFFQGPSTPIPTPVLDNSPAPIFSNPVCFSSRSENTEQYLQDSTLDDL SQLGDPTKNFNVNNFWSLFPFDDVGYHPHSNDVGYHLHSDEESPFDF (SEQ ID NO: 305).

In some embodiments, a MAP-NBD of the present disclosure can have at its C-terminus amino acid residues at position 1 (A) to position 63 (P) of the naturally occurring *Legionella quateirensis*-derived protein as follows:

ALVKEYFPVFSSFHFTADQIVALICQSKQCFRNLKKNHQQWKNKGLSAEQIVDLILQETPPK P (SEQ ID NO: 306).

[0096] In some embodiments, the present disclosure provides methods for identifying an animal pathogen-derived repeat unit. For example, a consensus sequence can be defined comprising a first repeat motif, a spacer, and a second repeat motif. The consensus sequence can be

1xxx211x1xxx33x2x1xxxxxxxx1xxx1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 292),

1xxx211x1xxx33x2x1xxxxxxxx1xxxx1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 293),

1xxx211x1xxx33x2x1xxxxxxxx1xxxx1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 294),

1xxx211x1xxx33x2x1xxxxxxxx1xxxx1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 295),

1xxx211x1xxx33x2x1xxxxxxxx1xxxx1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 296).

For any one of SEQ ID NO: 292 – SEQ ID NO: 296, x can be any amino acid residue, 1, 2, and 3 are flexible residues that are defined as follows: 1 can be selected from any one of A, F, I, L, M, T, or V, 2 can be selected from any one of D, E, K, N, M, S, R, or Q, and 3 can be selected from any one of A, G, N, or S. Thus, in some embodiments, a MAP-NBD can be derived from an animal pathogen comprising the consensus sequence of SEQ ID NO: 292, SEQ ID NO: 293, SEQ ID NO: 294, SEQ

ID NO: 295, or SEQ ID NO: 296. Any one of consensus sequences of SEQ ID NO: 292 – SEQ ID NO: 296 can be compared against all sequences downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>), MGRast (<https://www.mg-rast.org/>), JGI (<https://jgi.doe.gov/our-science/science-programs/metagenomics/>), and EBI (<https://www.ebi.ac.uk/metagenomics/>) databases to identify matches corresponding to animal pathogen proteins containing repeat units of a DNA-binding repeat unit.

[0097] In some embodiments, a MAP-NBD repeat unit can itself have a consensus sequence of 1xxx211x1xxx33x2x1xxxxxxxx1 (SEQ ID NO: 293), wherein x can be any amino acid residue, 1, 2, and 3 are flexible residues that are defined as follows: 1 can be selected from any one of A, F, I, L, M, T, or V, 2 can be selected from any one of D, E, K, N, M, S, R, or Q, and 3 can be selected from any one of A, G, N, or S.

D. Mixed DNA Binding Domains

[0098] In some embodiments, the present disclosure provides DNA binding domains in which the repeat units, the N-terminus, and the C-terminus can be derived from any one of *Ralstonia solanacearum*, *Xanthomonas spp.*, *Legionella quateirensis*, *Burkholderia*, *Paraburkholderia*, or *Francisella*. For example, the present disclosure provides a DNA binding domain wherein the plurality of repeat units are selected from any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352 and can further comprise an N-terminus and/or C-terminus from *Xanthomonas spp.*, (N-termini: SEQ ID NO: 300, SEQ ID NO: 301, and SEQ ID NO: 316; C-termini: SEQ ID NO: 302 and SEQ ID NO: 298) or *Legionella quateirensis* (N-termini: SEQ ID NO: 304 and SEQ ID NO: 317; C-termini: SEQ ID NO: 305 and SEQ ID NO: 306). In some embodiments, the present disclosure provides modular DNA binding domains in which the repeat units can be from *Ralstonia solanacearum* (e.g., any one of SEQ ID NO: 168 – SEQ ID NO: 263 or SEQ ID NO: 332 – SEQ ID NO: 352), *Xanthomonas spp.* (e.g., any one of SEQ ID NO: 318 – SEQ ID NO: 329), an animal pathogen such as *Legionella quateirensis*, *Burkholderia*, *Paraburkholderia*, or *Francisella* (e.g., any one of SEQ ID NO: 353, SEQ ID NO: 282 – SEQ ID NO: 290, or SEQ ID NO: 354 – SEQ ID NO: 442), or any combination thereof.

Zinc Finger Proteins

[0099] The present disclosure provides for novel endonucleases. In some aspects, the novel endonucleases can be fused to a DNA binding domain of a zinc finger protein. The present disclosure provides for non-naturally occurring fusion protein, e.g., anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) that are

fused to particular zinc finger proteins. A non-naturally occurring fusion protein, *e.g.*, any one of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) fused to a ZFP can include multiple components including the zinc finger protein (ZFP), an optional linker, and the linkage to the endonuclease. The ZFPs described herein can be used for genome editing.

[0100] The ZFP can also be referred to as a zinc finger DNA binding domain. The zinc finger DNA binding domain can comprise a set of zinc finger motifs. Each zinc finger motif can be about 30 amino acids in length and can fold into a $\beta\beta\alpha$ structure in which the α -helix can be inserted into the major groove of the DNA double helix and can engage in sequence-specific interaction with the DNA site. In some cases, the sequence-specific recognition can span over 3 base pairs. In some cases, a single zinc finger motif can interact specifically with 1, 2 or 3 nucleotides.

[0101] A zinc finger DNA binding domain of a ZFN can comprise from 1 to 10 zinc finger motifs. A zinc finger DNA binding domain can comprise from 1 to 9, from 2 to 8, from 2 to 6 or from 2 to 4 zinc finger motifs. In some cases, a zinc finger DNA binding domain can comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more zinc finger motifs. A zinc finger DNA binding domain can comprise at least 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 zinc finger motifs. A zinc finger DNA binding domain can comprise 1 zinc finger motif. A zinc finger DNA binding domain can comprise 2 zinc finger motif. A zinc finger DNA binding domain can comprise 3 zinc finger motif. A zinc finger DNA binding domain can comprise 4 zinc finger motif. A zinc finger DNA binding domain can comprise 5 zinc finger motif. A zinc finger DNA binding domain can comprise 6 zinc finger motif. A zinc finger DNA binding domain can comprise 7 zinc finger motif. A zinc finger DNA binding domain can comprise 8 zinc finger motif. A zinc finger DNA binding domain can comprise 9 zinc finger motif. A zinc finger DNA binding domain can comprise 10 zinc finger motif.

[0102] A zinc finger motif can be a wild-type zinc finger motif or a modified zinc finger motif enhanced for specific recognition of a set of nucleotides. A ZFN described herein can comprise one or more wild-type zinc finger motif. A ZFN described herein can comprise one or more modified zinc finger motif enhanced for specific recognition of a set of nucleotides. A modified zinc finger motif can comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, or more mutations that can enhance the motif for specific recognition of a set of nucleotides. In some cases, one or more amino acid residues within the α -helix of a zinc finger motif are modified. In some cases, one or more amino acid residues at positions -1, +1, +2, +3, +4, +5, and/or +6 relative to the N-terminus of the α -helix of a zinc finger motif can be modified.

Clustered Regularly Interspaced Palindromic Repeats-Associated-Deactivated Cas Protein (CRISPR-dCas9)

[0103] The present disclosure provides for novel endonucleases. In some instances the novel endonucleases can be fused to a clustered regularly interspaced palindromic repeats-associated-deactivated Cas protein, such as Cas9 (CRISPR-dCas9), or another suitable Cas protein. A CRISPR-dCas9 can comprise multiple components in a ribonucleoprotein complex, which can include the dCas9 protein that can interact with a single-guide RNA (sgRNA), an optional linker, and a any endonuclease disclosed herein (*e.g.*, SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)). The sgRNA can be made of a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). The dCas9 can be generated from a wild-type Cas9 protein by mutating 2 residues.

[0104] The CRISPR-dCas9s linked to an endonuclease described herein can be used to edit a target gene to which the sgRNA binds. For example, the CRISPR-dCas9s of the present disclosure can be used to knock out a target gene or it can be used to introduce a functional gene.

[0105] The sgRNA can comprise at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, or at least 25 base pairs that are complementary to a target sequences of interest. Thus, this portion of the sgRNA is analogous to the DNA binding domain described above with respect to ZFPs and TALEs. The portion of the sgRNA (*e.g.*, the about 20 base pairs within the sgRNA that bind to a target) bind adjacent to a protospacer adjacent motif (PAM), which can comprise 2-6 base pairs in the target sequence that is bound by dCas9.

[0106] A linker for linking an endonuclease domain of the present disclosure (*e.g.*, SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162)) to a dCas9 can be 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, 30, 35, 40, 45, or 50 amino acid residues in length. A linker can be 10-30 or 10-20 amino acid residues in length.

Linkers

[0107] A nuclease, *e.g.*, anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) fused to a DNA binding domain (*e.g.*, RNBD, TALE, MAP-NBD, ZFP, sgRNA), can further include a linker connecting SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) to the DNA binding domain. A linker used herein can be a short flexible linker comprising 0 base pairs, 3 to 6 base pairs, 6 to 12 base pairs, 12 to 15 base pairs, 15 to 21 base pairs, 21 to 24 base pairs, 24 to 30

base pairs, 30 to 36 base pairs, 36 to 42 base pairs, 42 to 48 base pairs, or 1-48 base pairs. The nucleic acid sequence of the linker can encode for an amino acid sequence comprising 0 residues, 1-3 residues, 4-7 residues, 8-10 residues, 10-12 residues, 12-15 residues, or 1-15 residues. Linkers can include, but are not limited to, residues such as glycine, methionine, aspartic acid, alanine, lysine, serine, leucine, threonine, tryptophan, or any combination thereof.

[0108] A nuclease, *e.g.*, anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) can be connected to a DNA binding domain via a linker, a linker can be between 1 to 70 amino acid residues in length. A linker can be from 5 to 45, from 5 to 40, from 5 to 35, from 5 to 30, from 5 to 25, from 5 to 20, from 5 to 15, from 10 to 40, from 10 to 35, from 10 to 30, from 10 to 25, from 10 to 20, from 12 to 40, from 12 to 35, from 12 to 30, from 12 to 25, from 12 to 20, from 14 to 40, from 14 to 35, from 14 to 30, from 14 to 25, from 14 to 20, from 14 to 16, from 15 to 40, from 15 to 35, from 15 to 30, from 15 to 25, from 15 to 20, from 15 to 18, from 18 to 40, from 18 to 35, from 18 to 30, from 18 to 25, from 18 to 24, from 20 to 40, from 20 to 35, from 20 to 30, from 25 to 30, from 25 to 70, from 30 to 70, from 5 to 70, from 35 to 70, from 40 to 70, from 45 to 70, from 50 to 70, from 55 to 70, from 60 to 70, or from 65 to 70 amino acid residues in length.

[0109] A linker for linking a nuclease, *e.g.*, anyone of SEQ ID NO: 1 – SEQ ID NO: 81 (or any one of nucleic acid sequences of SEQ ID NO: 82 – SEQ ID NO: 162) to a DNA binding domain can be 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, 30, 35, 40, 45, 50, 55, 60, 65, or 70 amino acid residues in length. A linker can be 10 – 20 or 10-30 amino acid residues in length.

[0110] In some embodiments, the linker can be the N-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or *Legionella quateirensis*-derived protein, wherein any functional domain disclosed herein is fused to the N-terminus of the engineered DNA binding domain. In some embodiments, the linker comprising the N-terminus can comprise the full length naturally occurring N-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or *Legionella quateirensis*-derived protein, or a truncation of the naturally occurring N-terminus, such as amino acid residues at positions 1 to 137 of the naturally occurring *Ralstonia solanacearum*-derived protein N-terminus (*e.g.*, SEQ ID NO: 264), positions 1 (H) to 115 (S) of the naturally occurring *Ralstonia solanacearum*-derived protein N-terminus (SEQ ID NO: 315), positions 1 (N) to 115 (S) of the naturally occurring *Xanthomonas spp.*-derived protein N-terminus (SEQ ID NO: 316), or positions 1 (G) to 115 (K) of the naturally occurring *Legionella quateirensis*-derived protein N-terminus (SEQ

ID NO: 317). In some embodiments, the linker can comprise amino acid residues at positions 1 to 120 of the naturally occurring *Ralstonia solanacearum*-derived protein (SEQ ID NO: 303), *Xanthomonas spp.*-derived protein (SEQ ID NO: 301), or *Legionella quateirensis*-derived protein (SEQ ID N): 304). In some embodiments, the linker can comprise the naturally occurring N-terminus of *Ralstonia solanacearum* truncated to any length. For example, the naturally occurring N-terminus of *Ralstonia solanacearum* can be truncated to amino acid residues at positions 1 to 120, 1 to 115, 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the N-terminus of the engineered DNA binding domain as a linker to a nuclease.

[0111] In other embodiments, the linker can be the C-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or animal pathogen-derived protein, wherein any functional domain disclosed herein is fused to the C-terminus of the engineered DNA binding domain. In some embodiments, the linker comprising the C-terminus can comprise the full length naturally occurring C-terminus of a naturally occurring *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or animal pathogen-derived protein, or a truncation of the naturally occurring C-terminus, such as positions 1 to 63 of the naturally occurring *Ralstonia solanacearum*-derived protein (SEQ ID NO: 266), *Xanthomonas spp.*-derived protein (SEQ ID NO: 298), or *Legionella quateirensis*-derived protein (SEQ ID NO: 306). In some embodiments, the naturally occurring C-terminus of *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or *Legionella quateirensis*-derived protein can be truncated to any length and used at the C-terminus of the engineered DNA binding domain and used as a linker to a nuclease. For example, the naturally occurring C-terminus of *Ralstonia solanacearum*-derived protein, *Xanthomonas spp.*-derived protein, or *Legionella quateirensis*-derived protein can be truncated to amino acid residues at positions 1 to 63, 1 to 50, 1 to 70, 1 to 100, 1 to 120, 1 to 130, 10 to 40, 60 to 100, or 100 to 120 and used at the C-terminus of the engineered DNA binding domain.

Linkers Comprising Recognition Sites

[0112] In some embodiments, the present disclosure provides DNA binding domains (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) with gapped repeat units for use as gene editing complexes. A DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) with gapped repeat can comprise of a plurality of repeat units in which each repeat unit of the plurality of repeat units is separated from a neighboring repeat unit by a linker. This linker can comprise a recognition site for additional functionality and activity. For example, the linker can comprise a recognition site for a small molecule. As another example, the linker can serve as a recognition site for a protease. In yet

another example, the linker can serve as a recognition site for a kinase. In other embodiments, the recognition site can serve as a localization signal.

[0113] Each repeat unit of a DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) comprises a secondary structure in which the RVD interfaces with and binds to a target nucleic acid base on double stranded DNA, while the remainder of the repeat unit protrudes from the surface of the DNA. Thus, the linkers comprising a recognition site between each repeat unit are removed from the surface of the DNA and are solvent accessible. In some embodiments, these solvent accessible linkers comprising recognition sites can have extra activity while mediating gene editing.

[0114] Examples of a left and a right DNA binding domain comprising repeat units derived from *Xanthomonas spp.* are shown below in **TABLE 7** for AAVS1 and GA7. “X,” shown in bold and underlining, represents a linker comprising a recognition site and can comprise 1-40 amino acid residues. An amino acid residue of the linker can comprise a glycine, an alanine, a threonine, or a histidine.

TABLE 7: Exemplary Left or Right Gapped DNA Binding Domains

SEQ ID NO	Construct	Sequence
SEQ ID NO: 307	AAVS1_Left	LTPDQVVAIASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVV AIASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDG GKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGGKQALET VQRLPVLCQDHG <u>X</u> LTPDQVVAIASNNGGKQALETQVRLPV LCQDHG <u>X</u> LTPDQVVAIASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVV AIASNIGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGG KQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGGKQALETQ VRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGGKQALETQVRLPV LCQDHG <u>X</u> LTPDQVVAIASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASNIGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAI ASHDGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASNIGGK QALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASNHGGKQALETQ VRLPVLCQDHG <u>X</u> LTPDQVVAIASNNGG
SEQ ID NO: 308	AAVS1_Right	LTPDQVVAIASNNGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVV AIASNNGGKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASNNG GKQALETQVRLPVLCQDHG <u>X</u> LTPDQVVAIASHDGGKQALET

[illegible]

SEQ ID NO	Construct	Sequence
		AIASNGGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVAIASHDG GKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVAIASNIGGKQALET VQRLPVLCQDHGX ~~L~~ TPDQVVAIASNGGGKQALETQVRLPV LCQDHGX ~~L~~ TPDQVVAIASNGGGKQALETQVRLPVLCQDHGX ~~L~~ LTPDQVVAIASHDGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVV AIASNGGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVAIASHDG GKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVAIASNGGGKQALET VQRLPVLCQDHGX ~~L~~ TPDQVVAIASNIGGKQALETQVRLPV LCQDHGX ~~L~~ TPDQVVAIASHDGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVAIASHDGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVA IASNIGGKQALETQVRLPVLCQDHGX ~~L~~ TPDQVVASASNGGG KQALESIVAQLSRPDPALA

Tunable Repeat Units

[0115] In some embodiments, the present disclosure provides DNA binding domains (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) with expanded repeat units. For example, a DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) comprises a plurality of repeat units in which each repeat unit is usually 33-35 amino acid residues in length. The present disclosure provides repeat units, which are greater than 35 amino acid residues in length. In some embodiments, the present disclosure provides repeat units, which are greater than 39 amino acid residues in length. In some embodiments, the present disclosure provides repeat units which are 35 to 40 amino acid residues long, 39 to 40 amino acid residues long, 35 to 45 amino acid residues long, 39 to 45 amino acid residues long, 35 to 50 amino acid residues long, 39 to 50 amino acid residues long, 35 to 60 amino acid residues long, 39 to 60 amino acid residues long, 35 to 70 amino acid residues long, 39 to 70 amino acid residues long, 35 to 79 amino acid residues long, or 39 to 79 amino acid residues long.

[0116] In other embodiments, the present disclosure provides DNA binding domains (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) with contracted repeat units. For example, the present disclosure provides repeat units, which are less than 32 amino acid residues in length. In some embodiments, the present disclosure provides repeat units, which are 15 to 32 amino acid residues in length, 16 to 32 amino acid residues in length, 17 to 32 amino acid residues in length, 18 to 32 amino acid residues in length, 19 to 32 amino acid residues in length, 20 to 32 amino acid residues in length, 21 to 32 amino acid residues in length, 22 to 32 amino acid residues in length, 23 to 32

amino acid residues in length, 24 to 32 amino acid residues in length, 25 to 32 amino acid residues in length, 26 to 32 amino acid residues in length, 27 to 32 amino acid residues in length, 28 to 32 amino acid residues in length, 29 to 32 amino acid residues in length, 30 to 32 amino acid residues in length, or 31 to 32 amino acid residues in length.

[0117] In some embodiments, said expanded repeat units can be tuned to modulate binding of each repeat unit to its target nucleic acid, resulting in the ability to overall modulate binding of the DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) to a target gene of interest. For example, expanding repeat units can improve binding affinity of the repeat unit to its target nucleic acid base and thereby increase binding affinity of the DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) to a target gene. In other embodiments, contracting repeat units can improve binding affinity of the repeat unit to its target nucleic acid base and thereby increase binding affinity of the DNA binding domain (*e.g.*, RNBDs, MAP-NBDs, TALEs, ZFPs, sgRNAs) for a target gene.

Genes and Indications of Interest

[0118] In some embodiments, genome editing can be performed by fusing a nuclease of the present disclosure with a DNA binding domain for a particular genomic locus of interest. Genetic modification can involve introducing a functional gene for therapeutic purposes, knocking out a gene for therapeutic gene, or engineering a cell *ex vivo* (*e.g.*, HSCs or CAR T cells) to be administered back into a subject in need thereof. For example, the genome editing complex can have a target site within PDCD1, CTLA4, LAG3, TET2, BTLA, HAVCR2, CCR5, CXCR4, TRA, TRB, B2M, albumin, HBB, HBA1, TTR, NR3C1, CD52, erythroid specific enhancer of the BCL11A gene, CBLB, TGFBR1, SERPINA1, HBV genomic DNA in infected cells, CEP290, DMD, CFTR, IL2RG, CS-1, or any combination thereof. In some embodiments, a genome editing complex can cleave double stranded DNA at a target site in order to insert a chimeric antigen receptor (CAR), alpha-L iduronidase (IDUA), iduronate-2-sulfatase (IDS), or Factor 9 (F9). Cells, such as hematopoietic stem cells (HSCs) and T cells, can be engineered *ex vivo* with the genome editing complex. Alternatively, genome editing complexes can be directly administered to a subject in need thereof.

[0119] The subject receiving treatment can be suffering from a disease such as transthyretin amyloidosis (ATTR), HIV, glioblastoma multiforme, cancer, acute lymphoblastic leukemia, acute myeloid leukemia, beta-thalassemia, sickle cell disease, MPSI, MPSII, Hemophilia B, multiple myeloma, melanoma, sarcoma, Leber congenital amaurosis (LCA10), CD19 malignancies, BCMA-

related malignancies, duchenne muscular dystrophy (DMD), cystic fibrosis, alpha-1 antitrypsin deficiency, X-linked severe combined immunodeficiency (X-SCID), or Hepatitis B.

Samples for Analysis

[0120] In some aspects, described herein include methods of modifying the genetic material of a target cell using any DNA binding described herein linked to any nuclease described herein. A sample described herein may be a fresh sample. The sample may be a live sample.

[0121] The sample may be a cell sample. The cell sample may be obtained from the cells or tissue of an animal. The animal cell may comprise a cell from an invertebrate, fish, amphibian, reptile, or mammal. The mammalian cell may be obtained from a primate, ape, equine, bovine, porcine, canine, feline, or rodent. The mammal may be a primate, ape, dog, cat, rabbit, ferret, or the like. The rodent may be a mouse, rat, hamster, gerbil, hamster, chinchilla, or guinea pig. The bird cell may be from a canary, parakeet, or parrot. The reptile cell may be from a turtle, lizard, or snake. The fish cell may be from a tropical fish. For example, the fish cell may be from a zebrafish (such as *Danio rerio*). The amphibian cell may be from a frog. An invertebrate cell may be from an insect, arthropod, marine invertebrate, or worm. The worm cell may be from a nematode (such as *Caenorhabditis elegans*). The arthropod cell may be from a tarantula or hermit crab.

[0122] The cell sample may be obtained from a mammalian cell. For example, the mammalian cell may be an epithelial cell, connective tissue cell, hormone secreting cell, a nerve cell, a skeletal muscle cell, a blood cell, an immune system cell, or a stem cell. A cell may be a fresh cell, live cell, fixed cell, intact cell, or cell lysate. Cell samples can be any primary cell, such as a hematopoietic stem cell (HSCs) or naïve or stimulated T cells (*e.g.*, CD4⁺ T cells).

[0123] Cell samples may be cells derived from a cell line, such as an immortalized cell line. Exemplary cell lines include, but are not limited to, 293A cell line, 293FT cell line, 293F cell line, 293 H cell line, HEK 293 cell line, CHO DG44 cell line, CHO-S cell line, CHO-K1 cell line, Expi293F™ cell line, Flp-In™ T-REx™ 293 cell line, Flp-In™-293 cell line, Flp-In™-3T3 cell line, Flp-In™-BHK cell line, Flp-In™-CHO cell line, Flp-In™-CV-1 cell line, Flp-In™-Jurkat cell line, FreeStyle™ 293-F cell line, FreeStyle™ CHO-S cell line, GripTite™ 293 MSR cell line, GS-CHO cell line, HepaRG™ cell line, T-REx™ Jurkat cell line, Per.C6 cell line, T-REx™-293 cell line, T-REx™-CHO cell line, T-REx™-HeLa cell line, NC-HIMT cell line, PC12 cell line, A549 cells, and K562 cells.

[0124] The cell sample may be obtained from cells of a primate. The primate may be a human, or a non-human primate. The cell sample may be obtained from a human. For example, the cell sample

may comprise cells obtained from blood, urine, stool, saliva, lymph fluid, cerebrospinal fluid, synovial fluid, cystic fluid, ascites, pleural effusion, amniotic fluid, chorionic villus sample, vaginal fluid, interstitial fluid, buccal swab sample, sputum, bronchial lavage, Pap smear sample, or ocular fluid. The cell sample may comprise cells obtained from a blood sample, an aspirate sample, or a smear sample.

[0125] The cell sample may be a circulating tumor cell sample. A circulating tumor cell sample may comprise lymphoma cells, fetal cells, apoptotic cells, epithelia cells, endothelial cells, stem cells, progenitor cells, mesenchymal cells, osteoblast cells, osteocytes, hematopoietic stem cells (HSC) (*e.g.*, a CD34⁺ HSC), foam cells, adipose cells, transcervical cells, circulating cardiocytes, circulating fibrocytes, circulating cancer stem cells, circulating myocytes, circulating cells from a kidney, circulating cells from a gastrointestinal tract, circulating cells from a lung, circulating cells from reproductive organs, circulating cells from a central nervous system, circulating hepatic cells, circulating cells from a spleen, circulating cells from a thymus, circulating cells from a thyroid, circulating cells from an endocrine gland, circulating cells from a parathyroid, circulating cells from a pituitary, circulating cells from an adrenal gland, circulating cells from islets of Langerhans, circulating cells from a pancreas, circulating cells from a hypothalamus, circulating cells from prostate tissues, circulating cells from breast tissues, circulating cells from circulating retinal cells, circulating ophthalmic cells, circulating auditory cells, circulating epidermal cells, circulating cells from the urinary tract, or combinations thereof.

[0126] The cell can be a T cell. For example, in some embodiments, the T cell can be an engineered T cell transduced to express a chimeric antigen receptor (CAR). The CAR T cell can be engineered to bind to BCMA, CD19, CD22, WT1, L1CAM, MUC16, ROR1, or LeY.

[0127] A cell sample may be a peripheral blood mononuclear cell sample.

[0128] A cell sample may comprise cancerous cells. The cancerous cells may form a cancer which may be a solid tumor or a hematologic malignancy. The cancerous cell sample may comprise cells obtained from a solid tumor. The solid tumor may include a sarcoma or a carcinoma. Exemplary sarcoma cell sample may include, but are not limited to, cell sample obtained from alveolar rhabdomyosarcoma, alveolar soft part sarcoma, ameloblastoma, angiosarcoma, chondrosarcoma, chordoma, clear cell sarcoma of soft tissue, dedifferentiated liposarcoma, desmoid, desmoplastic small round cell tumor, embryonal rhabdomyosarcoma, epithelioid fibrosarcoma, epithelioid hemangioendothelioma, epithelioid sarcoma, esthesioneuroblastoma, Ewing sarcoma, extrarenal rhabdoid tumor, extraskeletal myxoid chondrosarcoma, extraskeletal osteosarcoma, fibrosarcoma, giant cell tumor, hemangiopericytoma, infantile fibrosarcoma, inflammatory myofibroblastic tumor,

Kaposi sarcoma, leiomyosarcoma of bone, liposarcoma, liposarcoma of bone, malignant fibrous histiocytoma (MFH), malignant fibrous histiocytoma (MFH) of bone, malignant mesenchymoma, malignant peripheral nerve sheath tumor, mesenchymal chondrosarcoma, myxofibrosarcoma, myxoid liposarcoma, myxoinflammatory fibroblastic sarcoma, neoplasms with perivascular epithelioid cell differentiation, osteosarcoma, parosteal osteosarcoma, neoplasm with perivascular epithelioid cell differentiation, periosteal osteosarcoma, pleomorphic liposarcoma, pleomorphic rhabdomyosarcoma, PNET/extraskelatal Ewing tumor, rhabdomyosarcoma, round cell liposarcoma, small cell osteosarcoma, solitary fibrous tumor, synovial sarcoma, or telangiectatic osteosarcoma.

[0129] Exemplary carcinoma cell samples may include, but are not limited to, cell samples obtained from an anal cancer, appendix cancer, bile duct cancer (i.e., cholangiocarcinoma), bladder cancer, brain tumor, breast cancer, cervical cancer, colon cancer, cancer of Unknown Primary (CUP), esophageal cancer, eye cancer, fallopian tube cancer, gastroenterological cancer, kidney cancer, liver cancer, lung cancer, medulloblastoma, melanoma, oral cancer, ovarian cancer, pancreatic cancer, parathyroid disease, penile cancer, pituitary tumor, prostate cancer, rectal cancer, skin cancer, stomach cancer, testicular cancer, throat cancer, thyroid cancer, uterine cancer, vaginal cancer, or vulvar cancer.

[0130] The cancerous cell sample may comprise cells obtained from a hematologic malignancy. Hematologic malignancy may comprise a leukemia, a lymphoma, a myeloma, a non-Hodgkin's lymphoma, or a Hodgkin's lymphoma. The hematologic malignancy may be a T-cell based hematologic malignancy. The hematologic malignancy may be a B-cell based hematologic malignancy. Exemplary B-cell based hematologic malignancy may include, but are not limited to, chronic lymphocytic leukemia (CLL), small lymphocytic lymphoma (SLL), high risk CLL, a non-CLL/SLL lymphoma, prolymphocytic leukemia (PLL), follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL), mantle cell lymphoma (MCL), Waldenström's macroglobulinemia, multiple myeloma, extranodal marginal zone B cell lymphoma, nodal marginal zone B cell lymphoma, Burkitt's lymphoma, non-Burkitt high grade B cell lymphoma, primary mediastinal B-cell lymphoma (PMBL), immunoblastic large cell lymphoma, precursor B-lymphoblastic lymphoma, B cell prolymphocytic leukemia, lymphoplasmacytic lymphoma, splenic marginal zone lymphoma, plasma cell myeloma, plasmacytoma, mediastinal (thymic) large B cell lymphoma, intravascular large B cell lymphoma, primary effusion lymphoma, or lymphomatoid granulomatosis. Exemplary T-cell based hematologic malignancy may include, but are not limited to, peripheral T-cell lymphoma not otherwise specified (PTCL-NOS), anaplastic large cell lymphoma, angioimmunoblastic lymphoma, cutaneous T-cell lymphoma, adult T-cell leukemia/lymphoma

(ATLL), blastic NK-cell lymphoma, enteropathy-type T-cell lymphoma, hematosplenic gamma-delta T-cell lymphoma, lymphoblastic lymphoma, nasal NK/T-cell lymphomas, or treatment-related T-cell lymphomas.

[0131] A cell sample described herein may comprise a tumor cell line sample. Exemplary tumor cell line sample may include, but are not limited to, cell samples from tumor cell lines such as 600MPE, AU565, BT-20, BT-474, BT-483, BT-549, Evsa-T, Hs578T, MCF-7, MDA-MB-231, SkBr3, T-47D, HeLa, DU145, PC3, LNCaP, A549, H1299, NCI-H460, A2780, SKOV-3/Luc, Neuro2a, RKO, RKO-AS45-1, HT-29, SW1417, SW948, DLD-1, SW480, Capan-1, MC/9, B72.3, B25.2, B6.2, B38.1, DMS 153, SU.86.86, SNU-182, SNU-423, SNU-449, SNU-475, SNU-387, Hs 817.T, LMH, LMH/2A, SNU-398, PLHC-1, HepG2/SF, OCI-Ly1, OCI-Ly2, OCI-Ly3, OCI-Ly4, OCI-Ly6, OCI-Ly7, OCI-Ly10, OCI-Ly18, OCI-Ly19, U2932, DB, HBL-1, RIVA, SUDHL2, TMD8, MEC1, MEC2, 8E5, CCRF-CEM, MOLT-3, TALL-104, AML-193, THP-1, BDCM, HL-60, Jurkat, RPMI 8226, MOLT-4, RS4, K-562, KASUMI-1, Daudi, GA-10, Raji, JeKo-1, NK-92, and Mino.

[0132] A cell sample may comprise cells obtained from a biopsy sample, necropsy sample, or autopsy sample.

[0133] The cell samples (such as a biopsy sample) may be obtained from an individual by any suitable means of obtaining the sample using well-known and routine clinical methods. Procedures for obtaining tissue samples from an individual are well known. For example, procedures for drawing and processing tissue sample such as from a needle aspiration biopsy are well-known and may be employed to obtain a sample for use in the methods provided. Typically, for collection of such a tissue sample, a thin hollow needle is inserted into a mass such as a tumor mass for sampling of cells that, after being stained, will be examined under a microscope.

[0134] A cell may be a live cell. A cell may be a eukaryotic cell. A cell may be a yeast cell. A cell may be a plant cell. A cell may be obtained from an agricultural plant.

EXAMPLES

[0135] These examples are provided for illustrative purposes only and not to limit the scope of the claims provided herein.

EXAMPLE 1

Activity of Endonucleases

[0136] This example describes the activity of endonucleases including fusion proteins comprising DNA binding domains fused to each of SEQ ID NO: 1 (nucleic acid sequence of SEQ ID NO: 82),

SEQ ID NO: 2 (nucleic acid sequence of SEQ ID NO: 83), SEQ ID NO: 3 (nucleic acid sequence of SEQ ID NO: 84), SEQ ID NO: 4 (nucleic acid sequence of SEQ ID NO: 85), SEQ ID NO: 5 (nucleic acid sequence of SEQ ID NO: 86), SEQ ID NO: 6 (nucleic acid sequence of SEQ ID NO: 87), SEQ ID NO: 8 (nucleic acid sequence of SEQ ID NO: 89), described herein in A549 cells as compared with FokI, which has an amino acid sequence of

QLVKSELEEKKSELRHKLKYVPHEYIELIEIARNSTQDRILEMKVMEFFMKVYGYRGKHLG
GSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLPIGQADEMQRYVEENQTRNKHINPNEWW
KVYPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELLIGGEMIKAGTLTLEE
VRRKFNNGEINF (SEQ ID NO: 163) (nucleic acid sequence of

CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGAGCGAGCTGAGGCACAAGCTGAAG
TACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACA
GGATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGCAA
GCACCTGGGCGGCAGCAGGAAGCCCGACGGCGCCATCTACACCGTGGGCAGCCCCATC
GACTACGGCGTGATCGTGGACACCAAGGCCTACAGCGGCGGCTACAACCTGCCCATCG
GCCAGGCCGACGAGATGCAGAGGTACGTGGAGGAGAACCAGACCAGGAACAAGCACA
TCAACCCCAACGAGTGGTGGGAAGGTGTACCCAGCAGCGTGACCGAGTTCAAGTTCCTG
TTCGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCACA
TCACCAACTGCAACGGCGCCGTGCTGAGCGTGGAGGAGCTGCTGATCGGCGGCGAGAT
GATCAAGGCCGGCACCCCTGACCCTGGAGGAGGTGAGGAGGAAGTTCAACAACGGCGAG
ATCAACTTC; SEQ ID NO: 164), which comprises a DNA cleavage domain from FokI. The

resulting genome editing complexes were designed to cut a mutant allele of SMARCA4 present in the genome. The experiment measured the percentage of indels (insertions/deletions) at the SMARCA4 target site, as a result of non-homologous end joining (NHEJ). TALE domains fused to the above nucleases comprised SMARCA4 TALEs including TL4409 and TL4412. Sequences of said TALE domains are shown below in **TABLE 8**.

TABLE 8: SMARCA TALE Domains

TL #:	TALEN Name	Top Strand Sequence	Full Sequence on Top Strand of DNA (With Spacer indicated in lowercase):
TL4409	SMARCA4_5 Left TALEN	GGCGTGTCCCAGGCCCTTGC (SEQ ID NO: 444)	T GGCGTGTCCCAGGCCCTTGC acgtggcctatgctgt CACTGAGAGAGTGGACAAGC A (SEQ ID NO: 446)

TL #:	TALEN Name	Top Strand Sequence	Full Sequence on Top Strand of DNA (With Spacer indicated in lowercase):
TL4412	SMARCA4_5C Right TALEN	ACTGAGAGAGTGGACAAGC (SEQ ID NO: 445)	

[0137] NHEJ is a pathway that repairs double-strand breaks in DNA. NHEJ is referred to as “non-homologous” because the break ends are directly ligated without the need for a homologous template. SMARCA4 (or BRG1) is a chromatin remodeling ATPase frequently mutated in cancer.

[0138] Briefly, double-stranded DNA molecules encoding: SEQ ID NO: 1 (nucleic acid sequence of SEQ ID NO: 82), SEQ ID NO: 2 (nucleic acid sequence of SEQ ID NO: 83), SEQ ID NO: 3 (nucleic acid sequence of SEQ ID NO: 84), SEQ ID NO: 4 (nucleic acid sequence of SEQ ID NO: 85), SEQ ID NO: 5 (nucleic acid sequence of SEQ ID NO: 86), SEQ ID NO: 6 (nucleic acid sequence of SEQ ID NO: 87), SEQ ID NO: 7 (nucleic acid sequence of SEQ ID NO: 88), SEQ ID NO: 8 (nucleic acid sequence of SEQ ID NO: 89), and SEQ ID NO: 163 (nucleic acid sequence of SEQ ID NO: 164) comprising 15 bp overhangs were purchased. The 15 bp overhangs were designed to be compatible with the overhangs of a vector comprising TALE sequences that target the SMARCA4 gene, namely TL 4409 and TL 4412. The identity of each sequence and expression cassette was confirmed by sequencing. mRNA was generated from each expression cassette, which providing individual non-naturally occurring fusion proteins.

[0139] Adenocarcinomic human alveolar basal epithelial cells, A549s cells, were transfected with mRNA generated from each expression cassette. These cells were transfected in parallel with AAVS1 and a negative control (2 ug per TALEN). PCR and Minisequencing experiments were conducted to evaluate non-homologous end joining (NHEJ) efficiency. **TABLE 9** shows the percentages of indels on the target gene, SMARCA4.

TABLE 9 – Percentage of NHEJ on SMARCA4

SEQ ID NO	% Indels
SEQ ID NO: 1 (nucleic acid sequence of SEQ ID NO: 82)	42%
SEQ ID NO: 2 (nucleic acid sequence of SEQ ID NO: 83)	8%
SEQ ID NO: 3 (nucleic acid sequence of SEQ ID NO: 84)	0%
SEQ ID NO: 4 (nucleic acid sequence of SEQ ID NO: 85)	87%
SEQ ID NO: 5 (nucleic acid sequence of SEQ ID NO: 86)	0%
SEQ ID NO: 6 (nucleic acid sequence of SEQ ID NO: 87)	1%

SEQ ID NO	% Indels
SEQ ID NO: 7 (nucleic acid sequence of SEQ ID NO: 88)	0%
SEQ ID NO: 8 (nucleic acid sequence of SEQ ID NO: 89)	38%
SEQ ID NO: 163 (nucleic acid sequence of SEQ ID NO: 164)	74%

[0140] FIG. 2, FIG. 3, FIG. 4, FIG. 5, and FIG. 6 illustrate the base pair cleavage rate at a human SMARCA4 target site with FokI endonuclease, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 5, and SEQ ID NO: 8 respectively. FIG. 7 illustrates the base pair cleavage rate at a control AAVS1 target site. This positive control targets an intronic region of the AAVS1 locus.

[0141] Table 10 ranks the indicated nucleases by percent cutting:

SEQ ID NO	% Cleavage
4	94.5
6	75.2
51	74.7
53	73.7
8	70.4
2	67.6
52	60.6
1	59.7
16	59.2
54	51.9
47	51.7
3	43.3
49	43
56	38.5
58	36
74	33.2
59	32.6
22	31.8
17	28.1
65	27.7
48	26

EXAMPLE 2

Genome Editing with a Nuclease of SEQ ID NO: 7

[0142] This example illustrates genome editing with a nuclease of SEQ ID NO: 7. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 7 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 7, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0143] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 7 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0144] **Factor IX.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0145] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 7 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0146] **CAR.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and

the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0147] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 7 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0148] **TTR.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0149] **SERPINA1.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

Ex Vivo Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region

[0150] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 7 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0151] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0152] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 7 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0153] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 3

Genome Editing with a Nuclease of SEQ ID NO: 68

[0154] This example illustrates genome editing with a nuclease of SEQ ID NO: 68. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 68 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 68, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0155] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 68 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0156] **Factor IX.** The genome editing complex with a cleavage domain of SEQ ID NO: 68 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0157] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 68 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0158] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 68 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0159] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 68 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0160] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 68 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0161] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 68 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

Ex Vivo Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region

[0162] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 68 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0163] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 68 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0164] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 68 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified

CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0165] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 4

Genome Editing with a Nuclease of SEQ ID NO: 9

[0166] This example illustrates genome editing with a nuclease of SEQ ID NO: 9. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 9 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 9, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0167] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 9 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0168] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 9 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0169] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 9 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting

engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0170] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 9 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0171] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 9 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0172] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 9 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0173] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 9 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0174] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 9 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0175] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 9 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0176] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 9 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are

introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0177] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 5

Genome Editing with a Nuclease of SEQ ID NO: 79

[0178] This example illustrates genome editing with a nuclease of SEQ ID NO: 79. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 79 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 79, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0179] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 79 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0180] **Factor IX.** The genome editing complex with a cleavage domain of SEQ ID NO: 79 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0181] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 79 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene

(also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0182] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 79 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0183] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 79 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0184] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 79 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0185] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 79 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (a1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0186] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 79 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0187] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 79 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0188] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 79 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for

introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0189] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 6

Genome Editing with a Nuclease of SEQ ID NO: 74

[0190] This example illustrates genome editing with a nuclease of SEQ ID NO: 74. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 74 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 74, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0191] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 74 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0192] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 74 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0193] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 74

induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0194] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 74 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0195] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 74 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0196] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 74 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0197] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 74 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

Ex Vivo Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region

[0198] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 74 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0199] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 74 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0200] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 74 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0201] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 7

Genome Editing with a Nuclease of SEQ ID NO: 35

[0202] This example illustrates genome editing with a nuclease of SEQ ID NO: 35. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 35 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 35, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0203] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 35 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0204] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 35 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0205] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 35 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0206] **CAR.** The genome editing complex with a cleavage domain of SEQ ID NO: 35 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0207] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 35 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0208] **TTR.** The genome editing complex with a cleavage domain of SEQ ID NO: 35 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0209] **SERPINA1.** The genome editing complex with a cleavage domain of SEQ ID NO: 35 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

Ex Vivo Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region

[0210] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 35 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0211] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 35 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0212] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 35 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0213] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 8

Genome Editing with a Nuclease of SEQ ID NO: 2

[0214] This example illustrates genome editing with a nuclease of SEQ ID NO: 2. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 2 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 2, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0215] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 2 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0216] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 2 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the

factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0217] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 2 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0218] **CAR.** The genome editing complex with a cleavage domain of SEQ ID NO: 2 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0219] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 2 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0220] **TTR.** The genome editing complex with a cleavage domain of SEQ ID NO: 2 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0221] **SERPINA1.** The genome editing complex with a cleavage domain of SEQ ID NO: 2 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (dA1AT def).

Ex Vivo Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region

[0222] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 2 induces a double strand break in the DNA of the target

cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0223] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 2 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0224] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 2 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0225] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 9

Genome Editing with a Nuclease of SEQ ID NO: 80

[0226] This example illustrates genome editing with a nuclease of SEQ ID NO: 80. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 80 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 80, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0227] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 80 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0228] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 80 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double

strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0229] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 80 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0230] **CAR.** The genome editing complex with a cleavage domain of SEQ ID NO: 80 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0231] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 80 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0232] **TTR.** The genome editing complex with a cleavage domain of SEQ ID NO: 80 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0233] **SERPINA1.** The genome editing complex with a cleavage domain of SEQ ID NO: 80 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0234] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell

and the cleavage domain of SEQ ID NO: 80 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0235] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 80 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0236] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 80 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0237] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 10

Genome Editing with a Nuclease of SEQ ID NO: 49

[0238] This example illustrates genome editing with a nuclease of SEQ ID NO: 49. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 49 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 49, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0239] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 49 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0240] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 49 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with

delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0241] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 49 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0242] **CAR.** The genome editing complex with a cleavage domain of SEQ ID NO: 49 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0243] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 49 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0244] **TTR.** The genome editing complex with a cleavage domain of SEQ ID NO: 49 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0245] **SERPINA1.** The genome editing complex with a cleavage domain of SEQ ID NO: 49 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0246] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 49 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0247] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 49 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0248] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 49 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0249] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 11

Genome Editing with a Nuclease of SEQ ID NO: 47

[0250] This example illustrates genome editing with a nuclease of SEQ ID NO: 47. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 47 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 47, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0251] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 47 induces a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0252] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 47 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0253] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 47 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0254] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 47 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0255] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 47 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0256] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 47 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0257] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 47 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (a1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0258] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 47 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0259] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 47 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0260] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 47 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0261] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 12

Genome Editing with a Nuclease of SEQ ID NO: 63

[0262] This example illustrates genome editing with a nuclease of SEQ ID NO: 63. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 63 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 63, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0263] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 63 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0264] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 63 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0265] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 63 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0266] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 63 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0267] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 63 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0268] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 63 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0269] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 63 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0270] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 63 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0271] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 63 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0272] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 63 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0273] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 13

Genome Editing with a Nuclease of SEQ ID NO: 25

[0274] This example illustrates genome editing with a nuclease of SEQ ID NO: 25. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 25 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 25, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0275] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 25 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0276] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 25 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0277] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 25 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0278] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 25 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0279] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 25 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0280] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 25 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0281] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 25 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0282] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 25 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0283] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 25 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0284] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 25 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0285] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 14

Genome Editing with a Nuclease of SEQ ID NO: 12

[0286] This example illustrates genome editing with a nuclease of SEQ ID NO: 12. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 12 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 12, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0287] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 12 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0288] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 12 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0289] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 12 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0290] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 12 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0291] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 12 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0292] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 12 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0293] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 12 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0294] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 12 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0295] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 12 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0296] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 12 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0297] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 15

Genome Editing with a Nuclease of SEQ ID NO: 1

[0298] This example illustrates genome editing with a nuclease of SEQ ID NO: 1. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 1 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 1, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0299] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 1 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0300] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 1 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0301] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 1 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0302] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 1 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0303] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 1 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0304] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 1 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0305] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 1 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0306] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 1 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0307] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 1 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0308] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 1 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0309] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 16

Genome Editing with a Nuclease of SEQ ID NO: 54

[0310] This example illustrates genome editing with a nuclease of SEQ ID NO: 54. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 54 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 54, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0311] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 54 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0312] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 54 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0313] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 54 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0314] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 54 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0315] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 54 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0316] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 54 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0317] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 54 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0318] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 54 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0319] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 54 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0320] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 54 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0321] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 17

Genome Editing with a Nuclease of SEQ ID NO: 15

[0322] This example illustrates genome editing with a nuclease of SEQ ID NO: 15. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 15 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 15, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0323] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 15 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0324] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 15 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0325] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 15 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0326] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 15 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0327] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 15 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0328] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 15 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0329] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 15 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0330] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 15 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0331] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 15 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0332] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 15 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0333] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 18

Genome Editing with a Nuclease of SEQ ID NO: 10

[0334] This example illustrates genome editing with a nuclease of SEQ ID NO: 10. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 10 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 10, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0335] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 10 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0336] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 10 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0337] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 10 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0338] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 10 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0339] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 10 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0340] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 10 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0341] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 10 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0342] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 10 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0343] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 10 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0344] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 10 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0345] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 19

Genome Editing with a Nuclease of SEQ ID NO: 6

[0346] This example illustrates genome editing with a nuclease of SEQ ID NO: 6. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 6 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 6, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0347] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 6 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0348] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 6 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0349] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 6 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0350] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 6 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0351] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 6 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0352] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 6 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0353] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 6 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0354] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 6 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0355] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 6 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0356] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 6 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0357] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 20

Genome Editing with a Nuclease of SEQ ID NO: 23

[0358] This example illustrates genome editing with a nuclease of SEQ ID NO: 23. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 23 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 23, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0359] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 23 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0360] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 23 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0361] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 23 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0362] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 23 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0363] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 23 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0364] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 23 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0365] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 23 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0366] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 23 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0367] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 23 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0368] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 23 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0369] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 21

Genome Editing with a Nuclease of SEQ ID NO: 4

[0370] This example illustrates genome editing with a nuclease of SEQ ID NO: 4. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 4 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 4, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0371] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 4 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0372] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 4 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0373] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 4 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0374] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 4 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0375] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 4 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0376] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 4 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0377] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 4 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0378] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 4 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0379] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 4 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0380] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 4 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0381] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 22

Genome Editing with a Nuclease of SEQ ID NO: 5

[0382] This example illustrates genome editing with a nuclease of SEQ ID NO: 5. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 5 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 5, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0383] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 5 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0384] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 5 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0385] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 5 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0386] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 5 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0387] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 5 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0388] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 5 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0389] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 5 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0390] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 5 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0391] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 5 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0392] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 5 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0393] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 23

Genome Editing with a Nuclease of SEQ ID NO: 8

[0394] This example illustrates genome editing with a nuclease of SEQ ID NO: 8. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 8 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 8, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0395] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 8 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0396] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 8 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0397] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 8 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0398] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 8 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0399] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 8 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0400] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 8 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0401] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 8 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0402] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 8 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0403] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 8 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0404] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 8 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0405] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 24

Genome Editing with a Nuclease of SEQ ID NO: 17

[0406] This example illustrates genome editing with a nuclease of SEQ ID NO: 24. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 24 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 24, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0407] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 24 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0408] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 24 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

***Ex Vivo* Engineering of a Cell to Introduce a Gene**

[0409] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 24 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0410] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 24 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0411] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 24 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0412] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 24 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0413] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 24 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0414] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 24 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0415] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 24 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0416] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 24 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0417] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 25

Genome Editing with a Nuclease of SEQ ID NO: 31

[0418] This example illustrates genome editing with a nuclease of SEQ ID NO: 31. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 31 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 31, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0419] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 31 induces

a double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0420] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 31 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0421] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 31 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0422] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 31 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0423] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 31 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0424] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 31 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0425] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 31 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0426] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 31 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0427] **BCL11A Enhancer.** The genome editing complex with a cleavage domain of SEQ ID NO: 31 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0428] **CCR5.** The genome editing complex with a cleavage domain of SEQ ID NO: 31 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0429] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

EXAMPLE 26

Genome Editing with a Nuclease of SEQ ID NO: 3

[0430] This example illustrates genome editing with a nuclease of SEQ ID NO: 3. A DNA binding domain described herein, such as a TALE protein, an RNBD, a MAP-NBD, a ZFP, or a CRISPR-dCas9 system, is fused to a cleavage domain, such as an endonuclease, of SEQ ID NO: 3 to form a genome editing complex. The DNA binding domain is fused to SEQ ID NO: 3, optionally, via a naturally occurring linker, a variant or truncation of a naturally occurring linker, or a synthetic linker.

Direct Administration to Introduce a Gene

[0431] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 3 induces a

double strand break in the DNA of the target cell to introduce a gene. The introduced gene is a mutated gene or a functional gene.

[0432] Factor IX. The genome editing complex with a cleavage domain of SEQ ID NO: 3 introduces a double strand break into the albumin gene locus (e.g., into intron 1) concomitant with delivery to the cell of an ectopic nucleic acid bearing a cDNA of the factor IX gene. The double strand break leads to the integration of the ectopic nucleic acid into intron 1 of the albumin gene; the factor IX protein is secreted by the cell into the circulation. The target cell is a hepatocyte and the subject in need thereof has Hemophilia B.

Ex Vivo Engineering of a Cell to Introduce a Gene

[0433] The genome editing complex is transfected into cells *ex vivo* along with an ectopic nucleic acid bearing a gene. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 3 induces a double strand break in the DNA of the target cell to introduce an ectopically provided gene (also provided to the cell) into the region cleaved by the genome editing complex. The resulting engineered cells with modified DNA are administered to a subject in need thereof. The subject has a disease.

[0434] CAR. The genome editing complex with a cleavage domain of SEQ ID NO: 3 introduces a chimeric antigen receptor (CAR) by editing the DNA of a target cell. The target cell is a T cell and the subject has cancer, such as a blood cancer. Upon administration of the engineered cells to a subject, the engineered CAR T cells effectively eliminate cancer in the subject.

Direct Administration to Partially or Completely Knock Out a Gene

[0435] The genome editing complex is administered directly to a subject in need thereof and is taken up by a cell. The subject has a disease. The DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 3 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene.

[0436] TTR. The genome editing complex with a cleavage domain of SEQ ID NO: 3 partially or completely knocks out the transthyretin (TTR) gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has transthyretin amyloidosis (ATTR).

[0437] SERPINA1. The genome editing complex with a cleavage domain of SEQ ID NO: 3 partially or completely knocks out the SERPINA1 gene by editing the DNA of a target cell. The target cell is a liver cell and the subject in need thereof has alpha-1 antitrypsin deficiency (A1AT def).

***Ex Vivo* Engineering of a Cell to Partially or Completely Knock Out a Gene or a Gene Regulatory Region**

[0438] The genome editing complex is transfected in cells *ex vivo*. Upon transfection of cells *ex vivo*, the DNA binding domain of the genome editing complex binds a region of DNA in a target cell and the cleavage domain of SEQ ID NO: 3 induces a double strand break in the DNA of the target cell to partially or completely knock out a gene or a gene regulatory region. The subject has a disease.

[0439] BCL11A Enhancer. The genome editing complex with a cleavage domain of SEQ ID NO: 3 partially or completely knocks out the BCL11A erythroid enhancer by editing the DNA of a target cell. The target cell is an HPSC and the subject in need thereof has b-thalassemia or sickle cell disease.

[0440] CCR5. The genome editing complex with a cleavage domain of SEQ ID NO: 3 partially or completely knocks the CCR5 gene by editing the DNA of a target cell, thereby allowing for introduction of a mutated version of CCR5. Target cells, in which mutated versions of CCR5 are introduced via the action of the genome editing complex, are not infected by HIV via the modified CCR5 receptor. The target cell is a T cell or a hematopoietic stem cell (HPSC) and the subject has HIV.

[0441] Upon administration of the genome editing complex directly to a subject or upon administration of an engineered cell with DNA that has been modified with the genome editing complex, the disease symptoms are eliminated or reduced.

[0442] While preferred embodiments of the present invention have been shown and described herein, it will be apparent to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A non-naturally occurring fusion protein comprising a nucleic acid binding domain and a cleavage domain, wherein the cleavage domain comprises at least 33.3% divergence from SEQ ID NO: 163 and is immunologically orthogonal to SEQ ID NO: 163.
2. The non-naturally occurring fusion protein of claim 1 comprising one or more of the following characteristics:
 - (a) induces greater than 1% indels at a target site;
 - (b) the cleavage domain comprises a molecular weight of less than 23 kDa;
 - (c) the cleavage domain comprises less than 196 amino acids; and
 - (d) capable of cleaving across a spacer region greater than 24 base pairs.
3. The non-naturally occurring fusion protein of claim 2, wherein the non-naturally occurring fusion protein induces greater than 5%, greater than 10%, greater than 20%, greater than 30%, greater than 40%, greater than 50%, greater than 60%, greater than 70%, greater than 80%, or greater than 90% indels at the target site.
4. The non-naturally occurring fusion protein of any one of claims 1-3, wherein the cleavage domain comprises at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 75% divergence from SEQ ID NO: 163.
5. The non-naturally occurring fusion protein of any one of claims 1-4, wherein the cleavage domain comprises a sequence selected from SEQ ID NO: 311 – SEQ ID NO: 314.
6. The non-naturally occurring fusion protein of any one of claims 1-5, wherein the cleavage domain comprises an amino acid sequence having at least 80% sequence identity to the amino acid sequence set forth in one of SEQ ID NO: 1 – SEQ ID NO: 81.
7. The non-naturally occurring fusion protein of any one of claims 1-6, wherein the cleavage domain comprises an amino acid sequence having at least 90% sequence identity to any one of the amino acid sequence set forth in one of SEQ ID NO: 1 – SEQ ID NO: 81.

8. The non-naturally occurring fusion protein of any one of claims 6-7, wherein the cleavage domain is encoded by a nucleic acid sequence comprising at least 80% sequence identity with any one of SEQ ID NO: 82 – SEQ ID NO: 162.
9. The non-naturally occurring fusion protein of any one of claims 6 or 7, wherein the cleavage domain is encoded by a nucleic acid sequence selected from SEQ ID NO: 82 – SEQ ID NO: 162.
10. The non-naturally occurring fusion protein of any one of claims 1-9, wherein the nucleic acid binding domain binds a first region of double stranded genomic DNA.
11. The non-naturally occurring fusion protein of any one of claims 1-10, wherein the cleavage domain binds a second region of double stranded genomic DNA.
12. The non-naturally occurring fusion protein of claim 11, wherein the second region of double stranded genomic DNA is within at most 50 bp of the first region of double stranded genomic DNA.
13. The non-naturally occurring fusion protein of any one of claims 11-12, wherein the second region of double stranded genomic DNA is within at most 15 bp of the first region of double stranded genomic DNA.
14. The non-naturally occurring fusion protein of any one of claims 1-13, wherein the nucleic binding domain comprises a modular nucleic acid binding domain comprising a potency for a target site greater than 65% and a specificity ratio for the target site of 50:1; and a functional domain; wherein the modular nucleic acid binding domain comprises a plurality of repeat units, wherein at least one repeat unit of the plurality comprises a binding region configured to bind to a target nucleic acid base in the target site, wherein the potency comprises indel percentage at the target site, and wherein the specificity ratio comprises indel percentage at the target site over indel percentage at a top-ranked off-target site of the non-naturally occurring fusion protein.
15. The non-naturally occurring fusion protein of any one of claims 1-13, wherein the nucleic binding domain comprises a sequence from a zinc finger protein (ZFP).
16. The non-naturally occurring fusion protein of any one of claims 1-13, wherein the cleavage domain is fused to a catalytically inactive Cas9 (dCas9).

17. The non-naturally occurring fusion protein of claim 16, wherein the nucleic acid binding domain comprises a guide RNA or a truncated guide RNA.

18. The non-naturally occurring fusion protein of claim 14, wherein the at least one repeat unit comprises a sequence of $A_{1-11}X_1X_2B_{14-35}$ (SEQ ID NO: 443),

wherein each amino acid residue of A_{1-11} comprises any amino acid residue;

wherein X_1X_2 comprises the binding region;

wherein each amino acid residue of B_{14-35} comprises any amino acid; and

wherein a first repeat unit of the plurality of repeat units comprises at least one residue in A_{1-11} , B_{14-35} , or a combination thereof that differs from a corresponding residue in a second repeat unit of the plurality of repeat units.

19. The non-naturally occurring fusion protein of any one of claims 14 or 18, wherein the binding region comprises an amino acid residue at position 13 or an amino acid residue at position 12 and the amino acid residue at position 13.

20. The non-naturally occurring fusion protein of claim 19, wherein the amino acid residue at position 13 binds to the target nucleic acid base.

21. The non-naturally occurring fusion protein of any one of claims 19-20, wherein the amino acid residue at position 12 stabilizes the configuration of the binding region.

22. The non-naturally occurring fusion protein of any one of claims 14 or 19-21, wherein the indel percentage is measured by deep sequencing.

23. The non-naturally occurring fusion protein of any one of claims 14 or 18-22, wherein the modular nucleic acid binding domain further comprises one or more properties selected from the following:

(a) binds the target site, wherein the target site comprises a 5' guanine;

(b) comprises from 7 repeat units to 25 repeat units; and

(c) upon binding to the target site, the modular nucleic acid binding domain is separated from a second modular nucleic acid binding domain bound to a second target site by from 2 to 50 base pairs.

24. The non-naturally occurring fusion protein of any one of claims 14 or 18-23, wherein the plurality of repeat units comprises a *Ralstonia* repeat unit, a *Xanthomonas* repeat unit, a *Legionella* repeat unit, or any combination thereof.
25. The non-naturally occurring fusion protein of claim 24, wherein the *Ralstonia* repeat unit is a *Ralstonia solanacearum* repeat unit, the *Xanthomonas* repeat unit is a *Xanthomonas spp.* repeat unit, and the *Legionella* repeat unit is a *Legionella quateirensis* repeat unit.
26. The non-naturally occurring fusion protein of any one of claims 18-25, wherein the B₁₄₋₃₅ of at least one repeat unit of the plurality of repeat units has at least 92% sequence identity to GGKQALEAVRAQLDLRAAPYG (SEQ ID NO: 280).
27. The non-naturally occurring fusion protein of any one of claims 14 or 18-26, wherein the binding region comprises HD binding to cytosine, NG binding to thymidine, NK binding to guanine, SI binding to adenosine, RS binding to adenosine, HN binding to guanine, or NT binds to adenosine.
28. The non-naturally occurring fusion protein of any one of claims 14 or 18-27, wherein the at least one repeat unit comprises any one of SEQ ID NO: 267 – SEQ ID NO: 279.
29. The non-naturally occurring fusion protein of any one of claims 14 or 18-28, wherein the at least one repeat unit comprises at least 80% sequence identity with any one of SEQ ID NO: 168 – SEQ ID NO: 263.
30. The non-naturally occurring fusion protein of any one of claims 14 or 18-29, wherein the at least one repeat unit comprises at least 80% sequence identity with SEQ ID NO: 209, SEQ ID NO: 197, SEQ ID NO: 233, SEQ ID NO: 253, SEQ ID NO: 203, or SEQ ID NO: 218.
31. The non-naturally occurring fusion protein of any one of claims 14 or 18-30, wherein the at least one repeat unit comprises any one of SEQ ID NO: 168 – SEQ ID NO: 263.
32. The non-naturally occurring fusion protein of any one of claims 14 or 18-31, wherein the at least one repeat unit comprises SEQ ID NO: 209, SEQ ID NO: 197, SEQ ID NO: 233, SEQ ID NO: 253, SEQ ID NO: 203, or SEQ ID NO: 218.
33. The non-naturally occurring fusion protein of any one of claims 14 or 18-32, wherein the target nucleic acid base is cytosine, guanine, thymidine, adenosine, uracil, or a combination thereof.

34. The non-naturally occurring fusion protein of any one of claims 14 or 18-33, wherein the modular nucleic acid binding domain comprises an N-terminus amino acid sequence, a C-terminus amino acid sequence, or a combination thereof.
35. The non-naturally occurring fusion protein of claim 34, wherein the N-terminus amino acid sequence is from *Xanthomonas spp.*, *Legionella quateirensis*, or *Ralstonia solanacearum*.
36. The non-naturally occurring fusion protein of any one of claims 34-35, wherein the N-terminus amino acid sequence comprises at least 80% sequence identity to SEQ ID NO: 264, SEQ ID NO: 300, SEQ ID NO: 331, SEQ ID NO: 303, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 315, SEQ ID NO: 316, or SEQ ID NO: 317.
37. The non-naturally occurring fusion protein of any one of claims 34-36, wherein the N-terminus amino acid sequence comprises SEQ ID NO: 264, SEQ ID NO: 300, SEQ ID NO: 331, SEQ ID NO: 303, SEQ ID NO: 301, SEQ ID NO: 304, SEQ ID NO: 315, SEQ ID NO: 316, or SEQ ID NO: 317.
38. The non-naturally occurring fusion protein of any one of claims 34-37, wherein the C-terminus amino acid sequence is from *Xanthomonas spp.*, *Legionella quateirensis*, or *Ralstonia solanacearum*.
39. The non-naturally occurring fusion protein of any one of claims 34-38, wherein the C-terminus amino acid sequence comprises at least 80% sequence identity to SEQ ID NO: 266, SEQ ID NO: 298, or SEQ ID NO: 306.
40. The non-naturally occurring fusion protein of any one of claims 34-39, wherein the C-terminus amino acid sequence comprises SEQ ID NO: 266, SEQ ID NO: 298, or SEQ ID NO: 306.
41. The non-naturally occurring fusion protein of any one of claims 34-40, wherein the C-terminus amino acid sequence serves as a linker between the modular nucleic acid binding domain and the cleavage domain.
42. The non-naturally occurring fusion protein of any one of claims 14 or 18-41, wherein the modular nucleic acid binding domain comprises a half repeat.

43. The non-naturally occurring fusion protein of claim 42, wherein the half repeat comprises at least 80% sequence identity to SEQ ID NO: 265, SEQ ID NO: 322 – SEQ ID NO: 329, or SEQ ID NO: 290.
44. The non-naturally occurring fusion protein of any one of claims 14 or 18-43, wherein the at least one repeat unit comprises 1-20 additional amino acid residues at the C-terminus.
45. The non-naturally occurring fusion protein of any one of claims 14 or 18-44, wherein the at least one repeat unit of the plurality of repeat units is separated from a neighboring repeat unit by a linker.
46. The non-naturally occurring fusion protein of claim 45, wherein the linker comprises a recognition site.
47. The non-naturally occurring fusion protein of claim 46, wherein the recognition site is for a small molecule, a protease, or a kinase.
48. The non-naturally occurring fusion protein of any one of claims 46-47, wherein the recognition site serves as a localization signal.
49. The non-naturally occurring fusion protein of any one of claims 14 or 18-48, wherein the plurality of repeat units comprises 3 to 60 repeat units.
50. The non-naturally occurring fusion protein of any one of claims 2-49, wherein the target site is a nucleic acid sequence within a PDCD1 gene, a CTLA4 gene, a LAG3 gene, a TET2 gene, a BTLA gene, a HAVCR2 gene, a CCR5 gene, a CXCR4 gene, a TRA gene, a TRB gene, a B2M gene, an albumin gene, a HBB gene, a HBA1 gene, a TTR gene, a NR3C1 gene, a CD52 gene, an erythroid specific enhancer of the BCL11A gene, a CBLB gene, a TGFBR1 gene, a SERPINA1 gene, a HBV genomic DNA in infected cells, a CEP290 gene, a DMD gene, a CFTR gene, or an IL2RG gene.
51. The non-naturally occurring fusion protein of any one of claims 1-50, wherein the nucleic acid binding domain and the cleavage domain are linked by a synthetic linker comprising 0-15 residues of glycine, methionine, aspartic acid, alanine, lysine, serine, leucine, threonine, tryptophan, or any combination thereof.

52. A nucleic acid comprising a nucleotide sequence encoding the non-naturally occurring fusion protein of any one of claims 1-51.
53. An expression cassette comprising a nucleotide sequence encoding for a sequence having an amino acid sequence identity of at least 80% to any one of SEQ ID NO: 1 – SEQ ID NO: 81.
54. The expression cassette of claim 53, wherein the nucleotide sequence is selected from SEQ ID NO: 82 – SEQ ID NO: 162.
55. An expression cassette comprising a nucleotide sequence encoding a cleavage domain, wherein the cleavage domain comprises at least 33.3% divergence from SEQ ID NO: 163 and is immunologically orthogonal to SEQ ID NO: 163.
56. The expression cassette of claim 55, wherein the expression cassette encodes for a non-naturally occurring fusion protein according to any one of claims 1-51.
57. The nucleic acid sequence of claim 52 or the expression cassette of any one of claims 53-56, wherein the expression cassette is a part of a viral vector.
58. A method of genome editing, the method comprising:
administering the non-naturally occurring fusion protein of any one of claims 1-51, the nucleic acid sequence of claim 52 or 57, or the expression cassette of any one of claims 53-57; and inducing a double stranded break.
59. A method of genome editing, the method comprising:
administering a first non-naturally occurring fusion protein of any one of claims 1-51 or a nucleic acid encoding the first non-naturally occurring fusion protein and a second non-naturally occurring fusion protein of any one of claims 1-51 or a nucleic acid encoding the second non-naturally occurring fusion protein;
dimerizing a cleavage domain of the first non-naturally occurring fusion protein and a cleavage domain of the second non-naturally occurring fusion protein; and
cleaving a double stranded genomic DNA at a target site,
wherein the first non-naturally occurring fusion protein comprises a first nucleic acid binding domain and a first cleavage domain and the second non-naturally occurring fusion protein comprises a second nucleic binding domain and a second cleavage domain and wherein the first cleavage

domain and the second cleavage domain comprise at least 33.3% divergence from SEQ ID NO: 163 and are immunologically orthogonal to SEQ ID NO: 163.

60. The method of claim 59, wherein the first nucleic acid binding domain recognizes a first region of a top strand of the double stranded genomic DNA at the target site and the first cleavage domain binds a second region of the top strand of the double stranded genomic DNA at the target site.

61. The method of any one of claims 59-60, wherein the second nucleic acid binding domain recognizes a first region of a bottom strand of the double stranded genomic DNA at the target site and the second cleavage domain binds a second region of the bottom strand of the double stranded genomic DNA at the target site.

62. The method of any one of claims 59-61, wherein the cleaving the double stranded genomic DNA at the target site comprises a double strand break.

63. The method of any one of claims 59-62, wherein the method further comprises introducing an ectopic nucleic acid encoding a gene.

64. The method of claim 63, wherein the ectopic nucleic acid encoding the gene comprises a chimeric antigen receptor (CAR), alpha-L iduronidase (IDUA), iduronate-2-sulfatase (IDS), or Factor 9 (F9).

65. The method of any one of claims 59-64, wherein the cleaving the double stranded genomic DNA at the target site partially or completely knocked out a target gene.

66. The method of claim 65, wherein the target gene is a nucleic acid sequence within a PDCD1 gene, a CTLA4 gene, a LAG3 gene, a TET2 gene, a BTLA gene, a HAVCR2 gene, a CCR5 gene, a CXCR4 gene, a TRA gene, a TRB gene, a B2M gene, an albumin gene, a HBB gene, a HBA1 gene, a TTR gene, a NR3C1 gene, a CD52 gene, an erythroid specific enhancer of the BCL11A gene, a CBLB gene, a TGFBR1 gene, a SERPINA1 gene, a HBV genomic DNA in infected cells, a CEP290 gene, a DMD gene, a CFTR gene, or an IL2RG gene.

67. The method of any one of claims 59-66, wherein the administering comprises direct administration to a subject in need thereof.

68. The method of any one of claims 59-66, wherein the administering comprises transfecting a cell ex vivo with the first non-naturally occurring fusion protein and the second non-naturally occurring fusion protein, thereby obtaining a cell comprising modified DNA.
69. The method of claim 68, wherein the cell comprising modified DNA is administered to a subject in need thereof.
70. The method of any one of claims 59-66, wherein the administering comprises intravenous, subcutaneous, intramuscular, or mucosal administration to a subject in need thereof.
71. The method of any one of claims 59-70, wherein the target site is in a cell.
72. The method of claim 71, wherein the cell comprises a T-cell, a hematopoietic stem cell (HPSC), or a liver cell.
73. The method of any one of claims 67 or 69-72, wherein the subject is a human.
74. The method of claim 73, wherein the human has cancer, blood cancer, a CD19 malignancy, a BCMA malignancy, transthyretin amyloidosis, HIV, glioblastoma multiforme, acute lymphoblastic leukemia, acute myeloid leukemia, b-thalassemia, sickle cell disease, MPSI, MPSII, hemophilia B, multiple myeloma, melanoma, sarcoma, Leber congenital amaurosis type 10 (LCA10), Duchenne muscular dystrophy, cystic fibrosis, alpha-1 antitrypsin deficiency (dA1AT def), X-linked severe combined immunodeficiency (X-SCID), or Hepatitis B.
75. The composition of any one of the preceding claims, wherein a repeat unit of the plurality of repeat units recognizes a target nucleic acid base and wherein the plurality of repeat units has one or more of the following characteristics:
- (a) at least one repeat unit comprising greater than 39 amino acid residues;
 - (b) at least one repeat unit comprising greater than 35 amino acid residues derived from the genus of *Ralstonia*;
 - (c) at least one repeat unit comprising less than 32 amino acid residues; and
 - (d) each repeat unit of the plurality of repeat units is separated from a neighboring repeat unit by a linker comprising a recognition site.

76. The composition of claim 75, wherein the at least one repeat unit comprises an amino acid selected from glycine, alanine, threonine or histidine at a position after an amino acid residue at position 35.

77. The composition any one of claims 75-76, wherein the at least one repeat unit comprises an amino acid selected from glycine, alanine, threonine or histidine at a position after an amino acid residue at position 39.

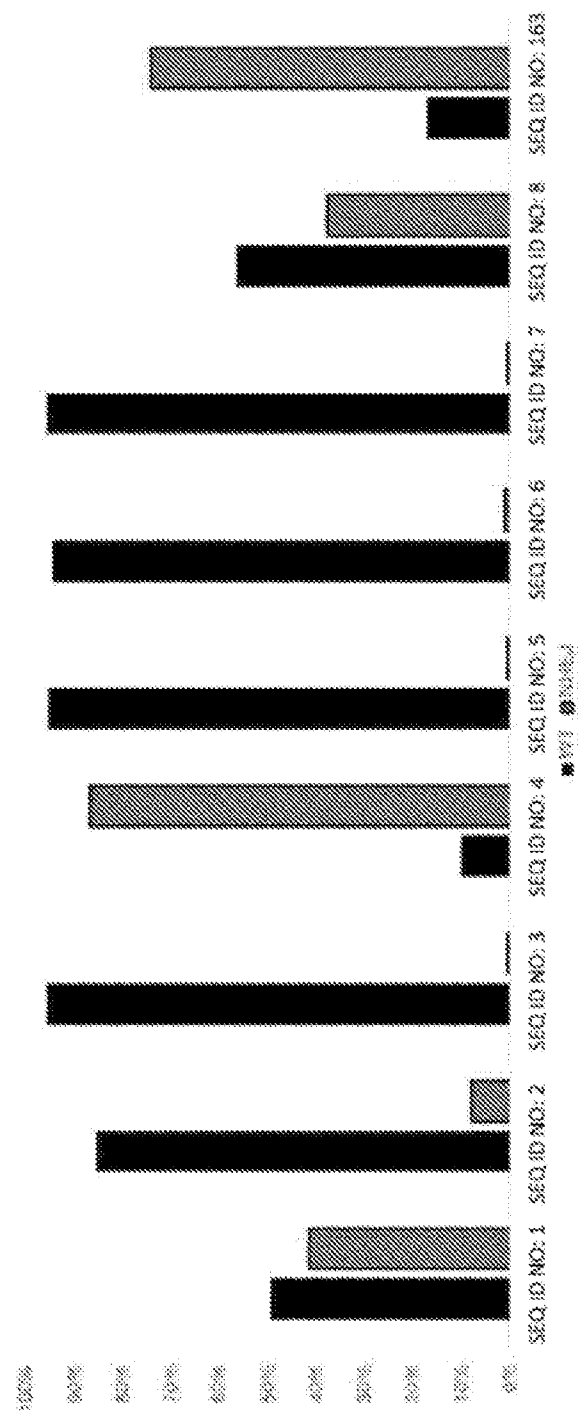


FIG. 1

2/11

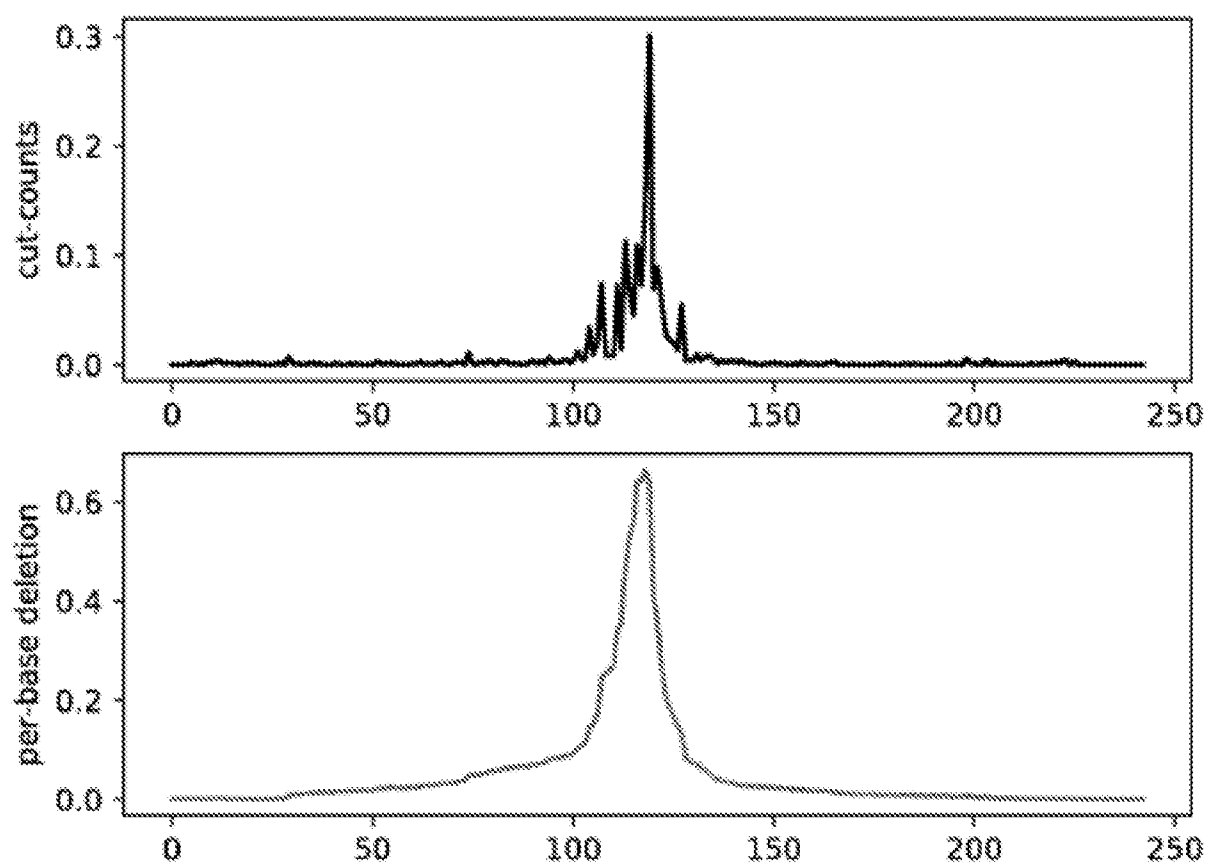


FIG. 2

3/11

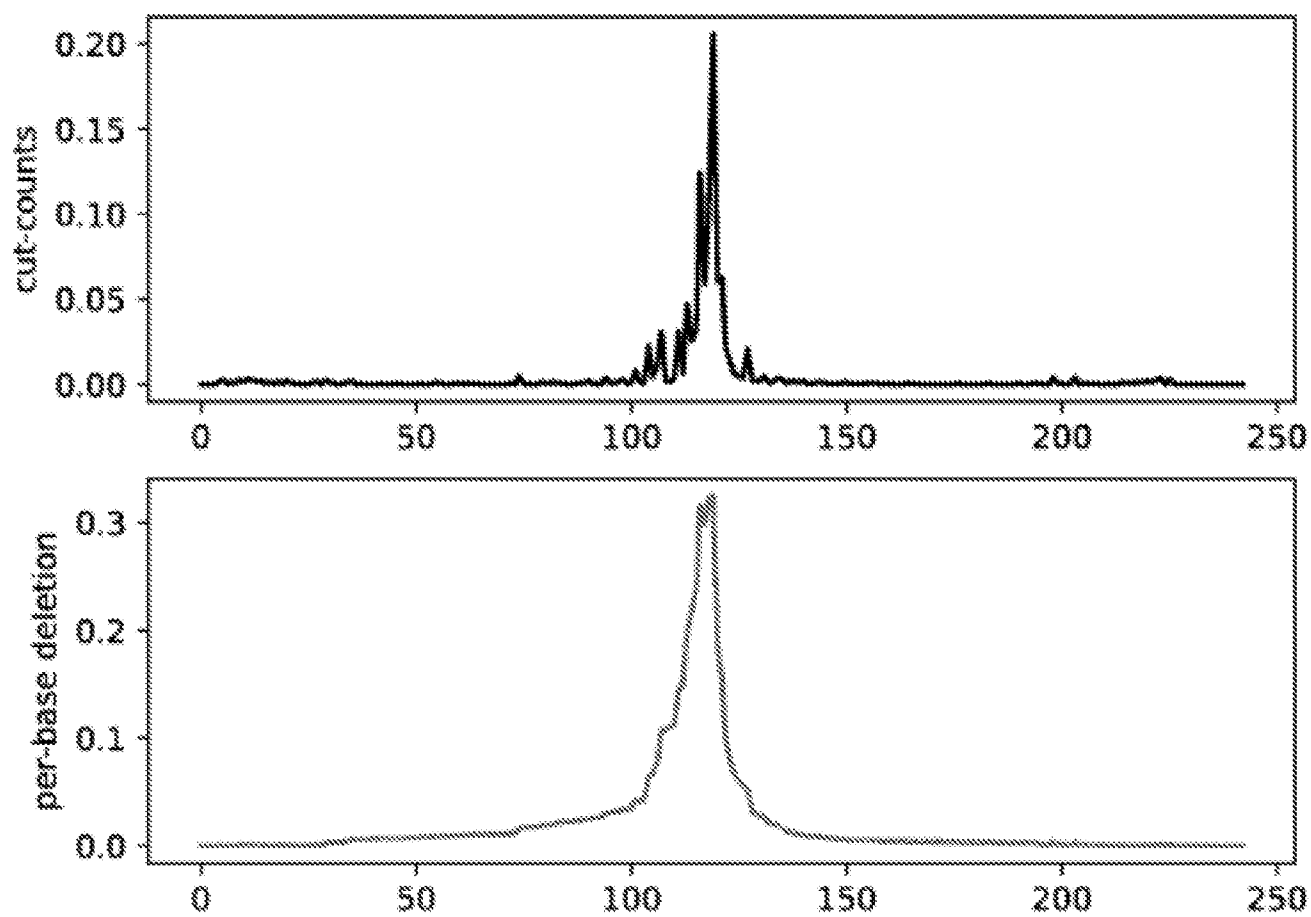


FIG. 3

4/11

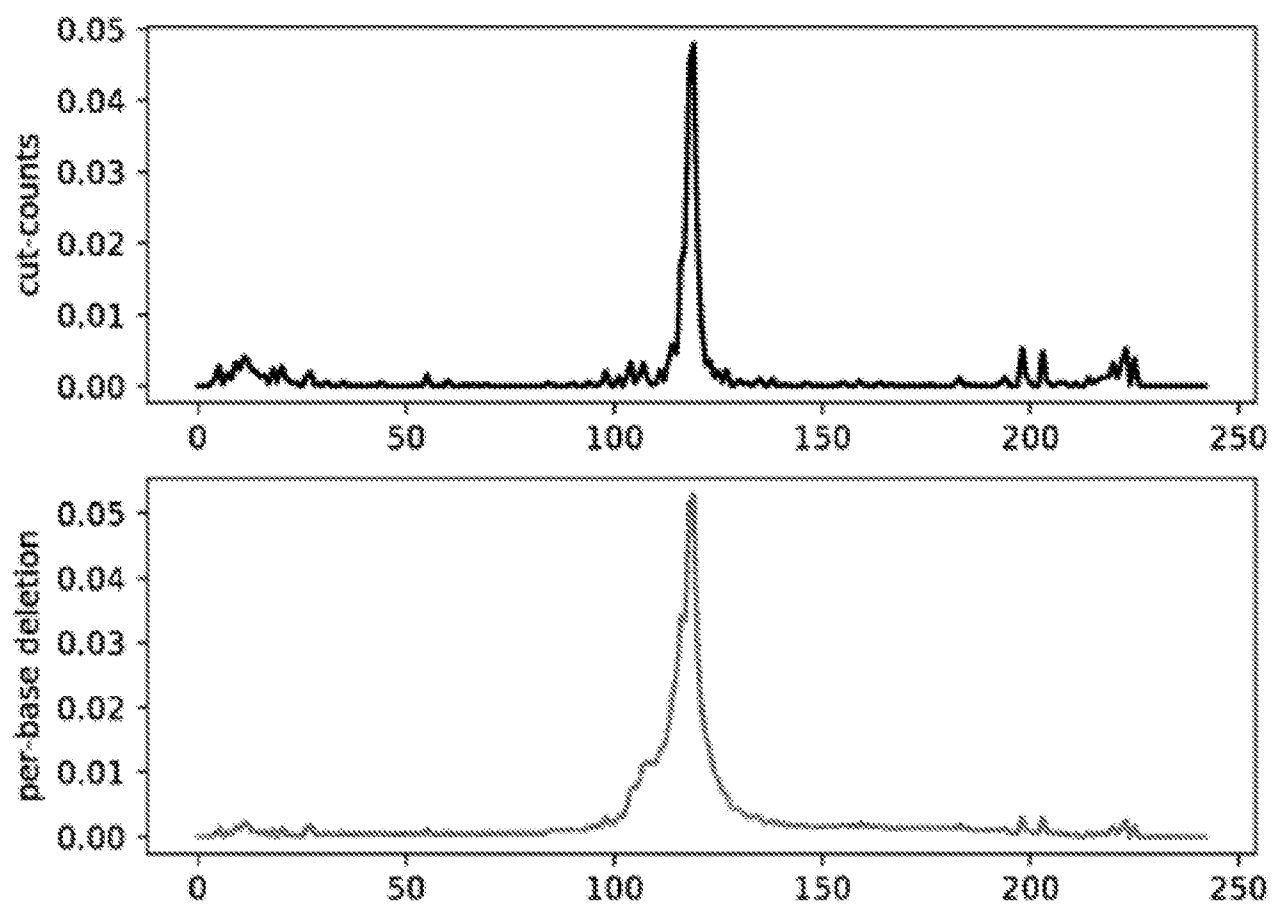


FIG. 4

5/11

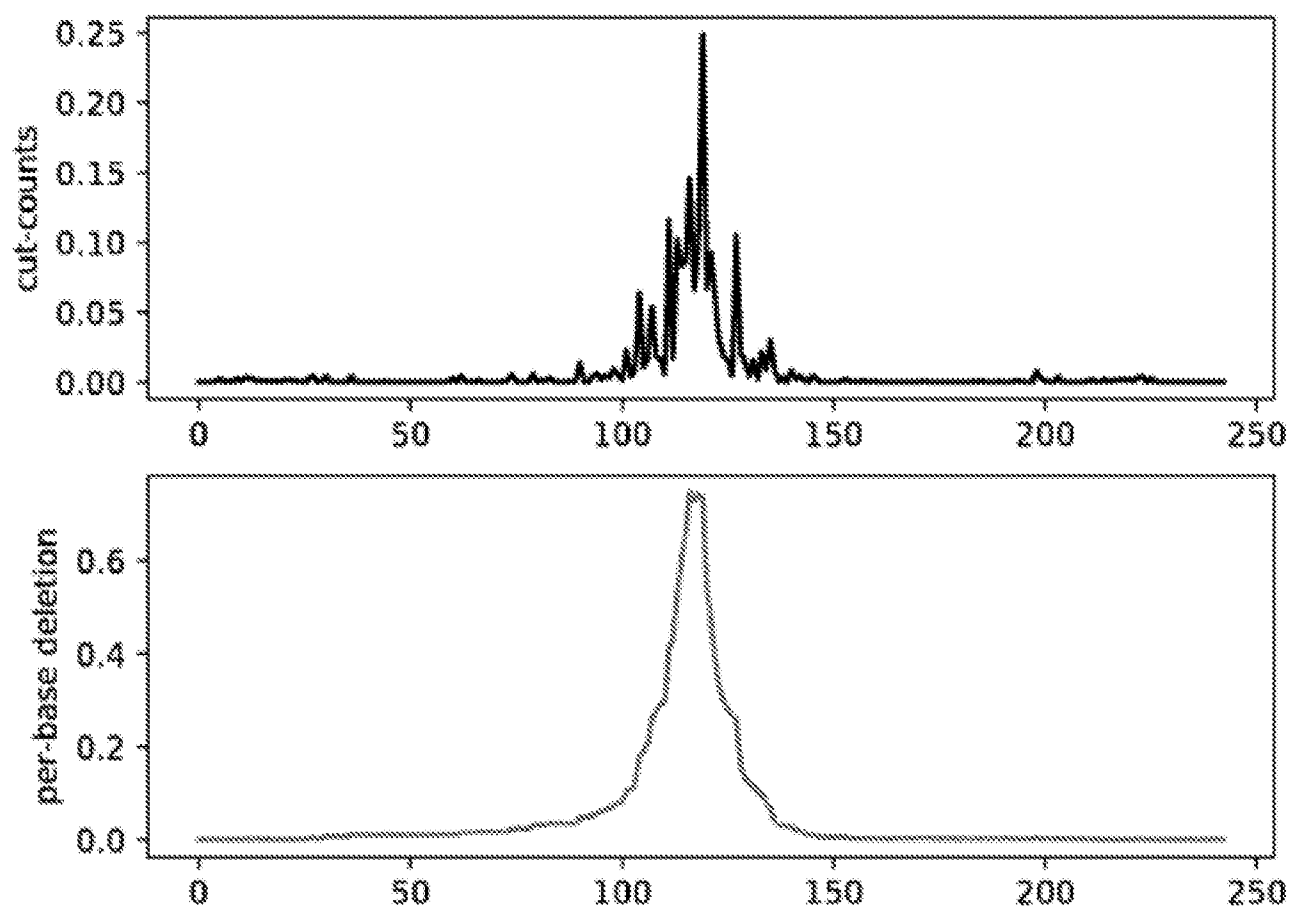


FIG. 5

6/11

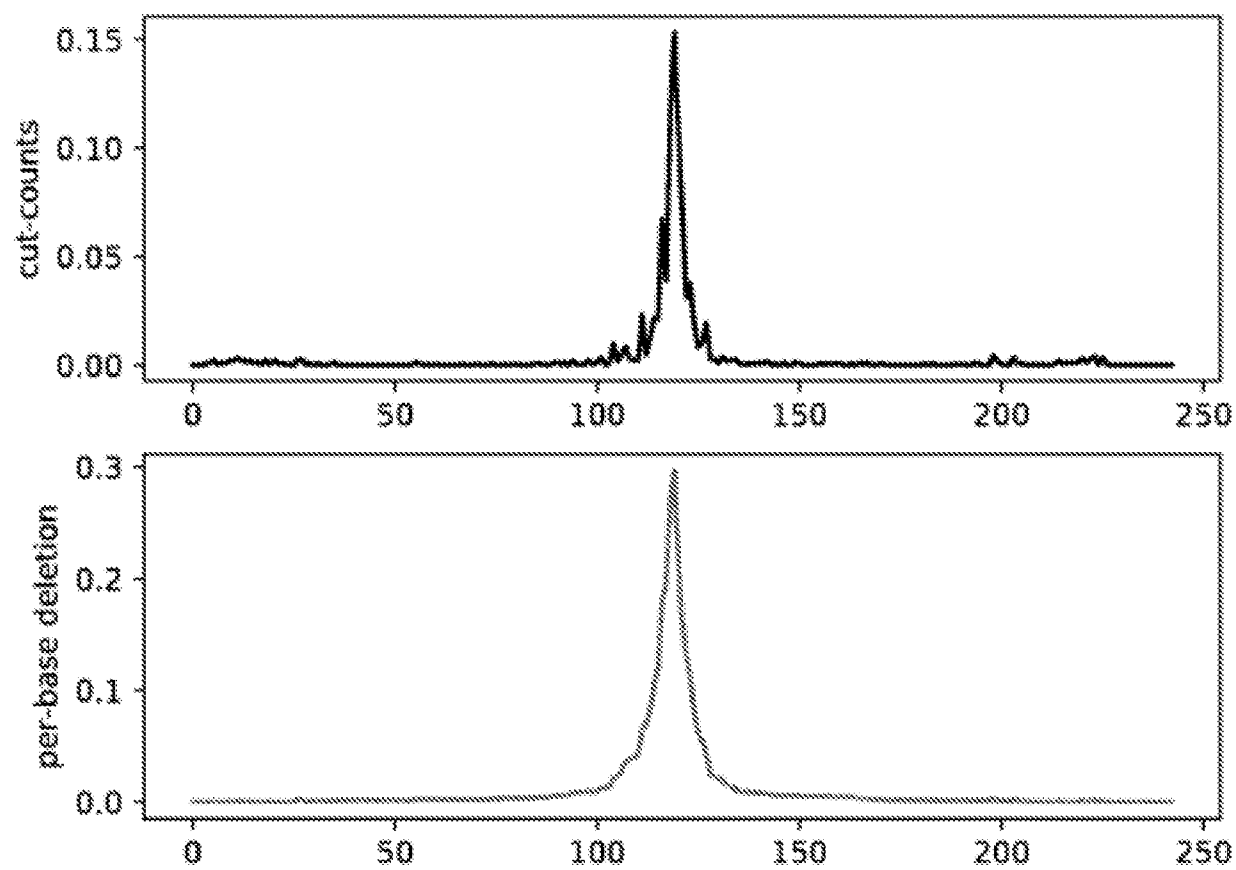


FIG. 6

7/11

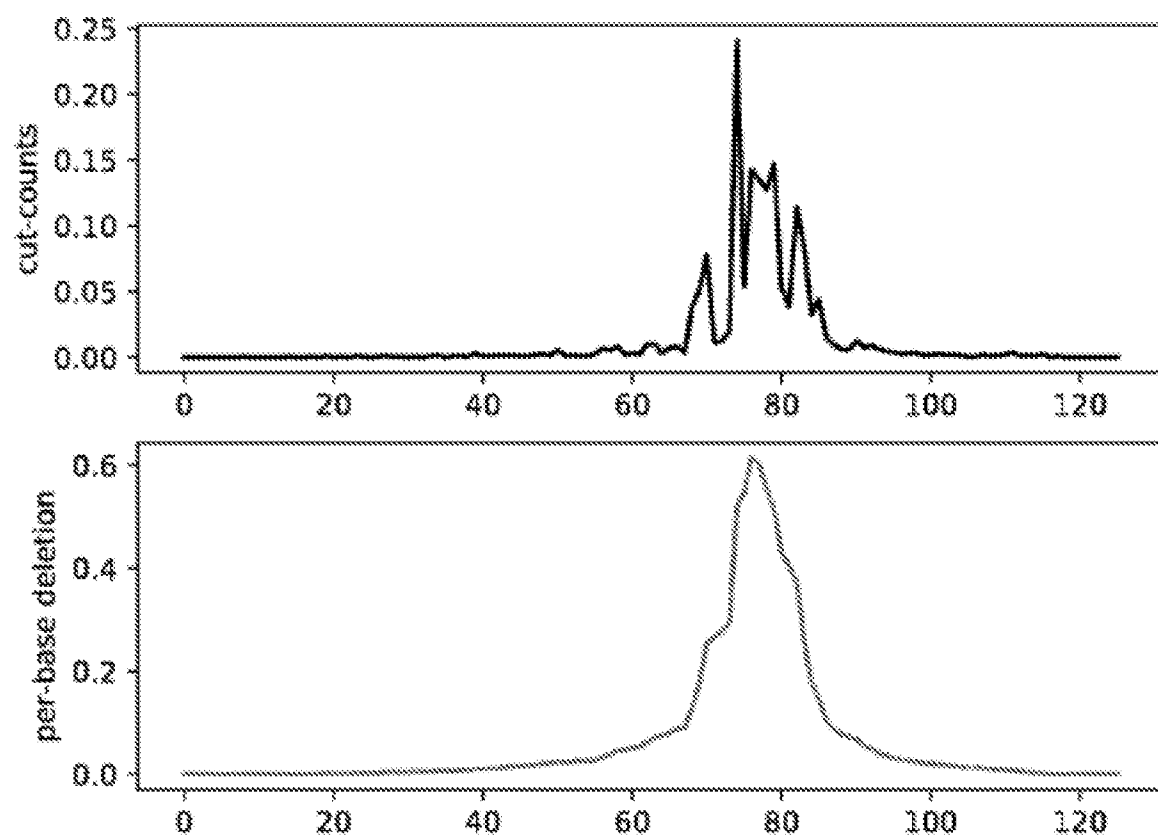


FIG. 7

65 AEADLNSEKIKNHY-RKITNLPKEYIELLDIAFDHRR-----HQDFEIVTAGLFKDC 51
74 QIKDKYLEDLKLEL-YKKTNLPNKYYEMVDIAYDGKR-----NREFEITYTSDLMQEI 51
1 FLVKGAMEIKKSELRHKL RHPHEYIELIEIAQDSKQ-----NRLLEFKVVEFFKKI 52
2 QIVKSSIEMSKANMRDNLQMLPHDYIELIEISQDPYQ-----NRIFEMKVMDFINE 52
3 QTIKSSIEELKSELRTQLNVISHDYQLVDISQDSQ-----NRLFEMKVMDFINE 52
16 KAEKSEFLIEKDKLREKLDLTPHDYLSMVDLAYDSKQ-----NRLFEMKTIELLINE 52
6 --AKSSQSETKEKLRKLRNLPHEYLSLVDLAYDSKQ-----NRLFEMKVIELLTEE 50
4 --GKSEVETIKEQMRGELTHLSHEYLGLLDLAYDSKQ-----NRLFELKTMQLLTEE 50
8 NLTCSDLTEIKEEVRNALTHLSHEYALIDLAYDSTQ-----NRLFEMKTLQLLVEE 52
56 VKEKTDAAALVKERVRLQLHNINHKYALIDYAFSGKN-----NSRDFEVYTTIDLLVNE 53
58 VKEKTDAAALVKERVRLQLHNINHKYALIDYAFSGKN-----NSRDFEVYTTIDLLVNE 53
59 VKETTDVVIKDRVRLKLHVNHKYLTIDYAFSGKN-----NCMDFEVYTTIDLLVNE 53
48 ECVKDNVVDIKDRVRNKLHLHDKYALIDLAYSDAASRAKKNADAREFEIQTADLFTKE 60
17 -ISKTNVLELKDKVRDKLYVDHRYLALIDLAYDGT-----NRDFEIQTIDLLINE 51
22 -ISKTNVLELKDKVRDKLYVDHRYLSLIDLAYDGN-----NRDFEIQTIDLLINE 51
49 -IAKSDFSIIKDNIRRKLYVNHKYLILLDLGFDSDS-----NRDYEIQTAEELLTE 51
51 VPVKSEVSLCKDYLRSHLNHVDHKYLIILLDLGFDGTS-----DRDYEIQTACLITGE 52
53 VPVKSEVSLCKDYLRSHLNHVDHKYLIILLDLGFDGTS-----DRDYEIQTACLITGE 52
47 VPVKSEVSLCKDYLRSHLVHVDHKYLVLLDLGFDGTS-----DRDYEIQTACLITGE 52
52 VPVKSEVSLCKDYLRTHLLHVDHRYLIILLDLGFDGTS-----DRDYEIQTACLITGE 52
54 VPVKSEMSLLCKDYLRTHLLHVDHRYLIILLDLGFDGAS-----DRDYEIQTACLITGE 52

* : . * : : : . :
* . . :

FIG. 8A

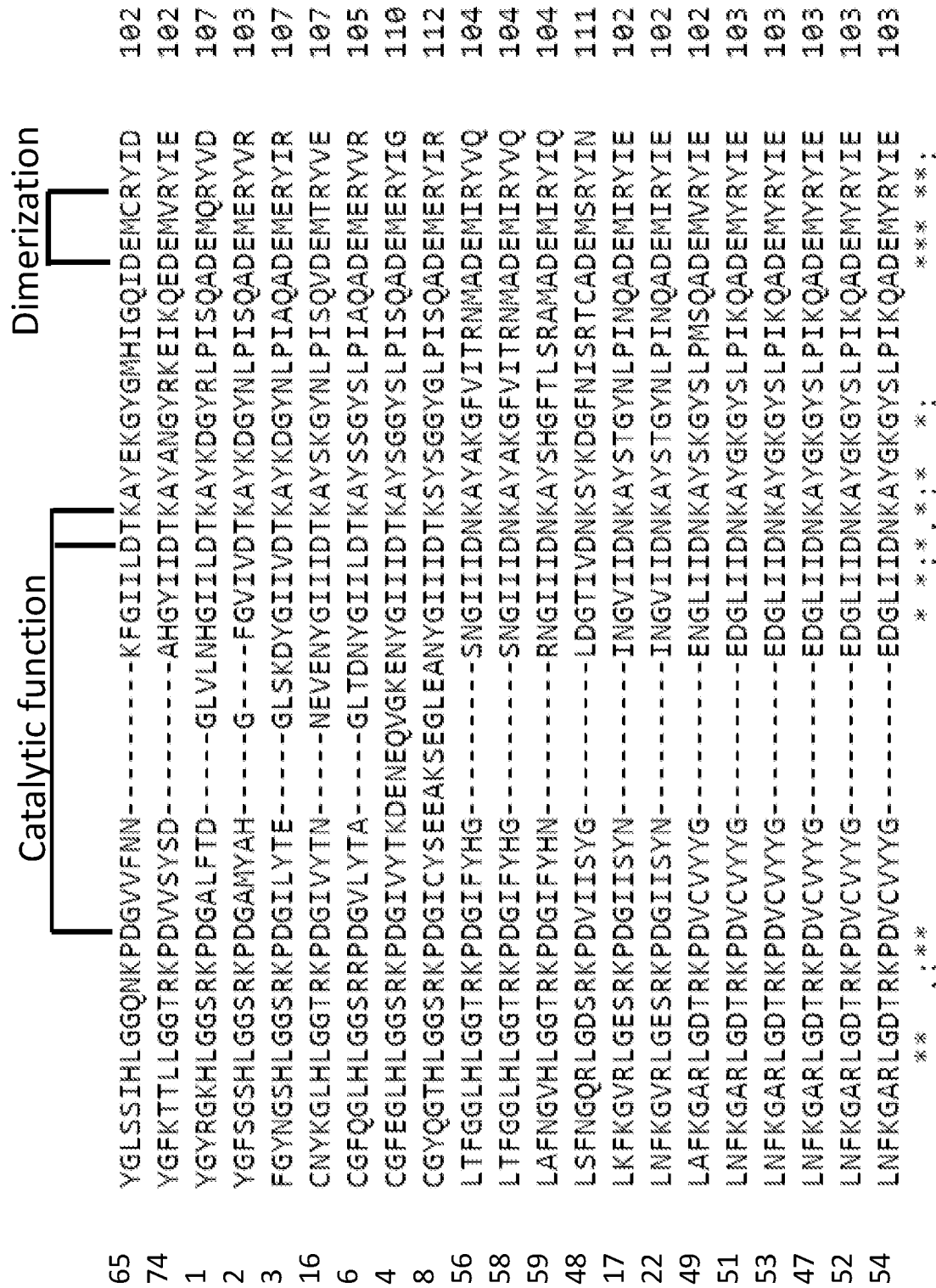


FIG. 8B

65	VSQLLL	GANAAM	KGKLDV	NTPKH	MNQV	IKL-	193
74	IVQLLL	GAHLVY	SGEISK	DQFAA	YMMN	TEINF-	193
1	VEQLLL	GGKEIK	EGSLT	LEEVG	KKFK	NDEIVF-	196
2	VEHLLL	GAEYIK	RGILT	LYDFK	NSFL	NKEIQF-	192
3	VEHLLL	GAEYFK	RGILS	LEDVR	DKFC	NTEIEF-	196
16	VTTLLY	IANEIK	ANRMKK	SDIMEY	FNDKV	----	193
6	ISQLLL	ADAI	RGGRMD	RERLR	HLMF	QNEEFL-	194
4	IQQLLR	IVNDYK	AGKLT	HEDMK	LKIF	HY----	195
8	IKTLLL	TANRCK	TGELDH	AGIESC	FFNN	CRL--	200
56	IENLLY	YADAIK	GGTLSK	ATFING	FNANK	EMVF	194
58	IENLLY	YADAIK	GGTLSK	ATFING	FNANK	EMVF	194
59	AENLLY	FAEAMK	SGVMPK	TEFISY	FGAG	KEIQF	195
48	VESLLY	LSEGIK	AGRISH	ADFYSN	FNKN	KEMIY-	201
17	VENLLY	FAEELK	AGRLSY	VDST	MYDN	DEIYV-	191
22	VENLLY	FAEELK	AGRLSY	ADSFT	MYDN	DEIYV-	191
49	TINLLL	AAEELK	SGRMSY	PKCFSY	FDTN	DEVHI	192
51	SMNLLL	MAEELK	SGRLGY	KECFKL	FDCN	DEIVL	193
53	SMNLLL	MAEELK	SGRLGY	KECFKL	FDCN	DEIVL	193
47	SMNLLL	AAEELK	SGRLGY	KECFKL	FDCN	DEIVL	193
52	SMNLLL	MAEELK	SGRLGY	KECFKL	FDCN	DEIVL	193
54	SMNLLL	MAEELK	SGRLGY	KECFKL	FDCN	DEIVL	193
	**	,	.	:			

FIG. 8D