**(12) STANDARD PATENT** (11) Application No. **AU 2019201344 C1**
**(19) AUSTRALIAN PATENT OFFICE**

(54) Title
**CRISPR-BASED GENOME MODIFICATION AND REGULATION**

(51) International Patent Classification(s)
***C12N 15/10*** (2006.01) ***C12N 9/22*** (2006.01)
***A61K 45/00*** (2006.01) ***C12N 15/63*** (2006.01)

(21) Application No: **2019201344** (22) Date of Filing: **2019.02.26**

(43) Publication Date: **2019.03.21**
(43) Publication Journal Date: **2019.03.21**
(44) Accepted Journal Date: **2020.09.03**
(44) Amended Journal Date: **2020.12.24**

(62) Divisional of:
**2018229489**

(71) Applicant(s)
**Sigma-Aldrich Co. LLC**

(72) Inventor(s)
**Chen, Fuqiang;Davis, Gregory D.;Kang, Qiaohua;Knight, Scott W.**

(74) Agent / Attorney
**Pizzeys Patent and Trade Mark Attorneys Pty Ltd, Level 15 241 Adelaide Street, BRISBANE, QLD, 4000, AU**

(56) Related Art
**WO 2014/065596 A1**
**WO 2013/176772 A1**
**US 2010/0076057 A1**

**ABSTRACT**

The present invention provides RNA-guided endonucleases, which are engineered for expression in eukaryotic cells or embryos, and methods of using the RNA-guided endonuclease for targeted genome modification in in eukaryotic cells or embryos. Also provided are fusion proteins, wherein each fusion protein comprises a CRISPR/Cas-like protein or fragment thereof and an effector domain. The effector domain can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. Also provided are methods for using the fusion proteins to modify a chromosomal sequence or regulate expression of a chromosomal sequence.

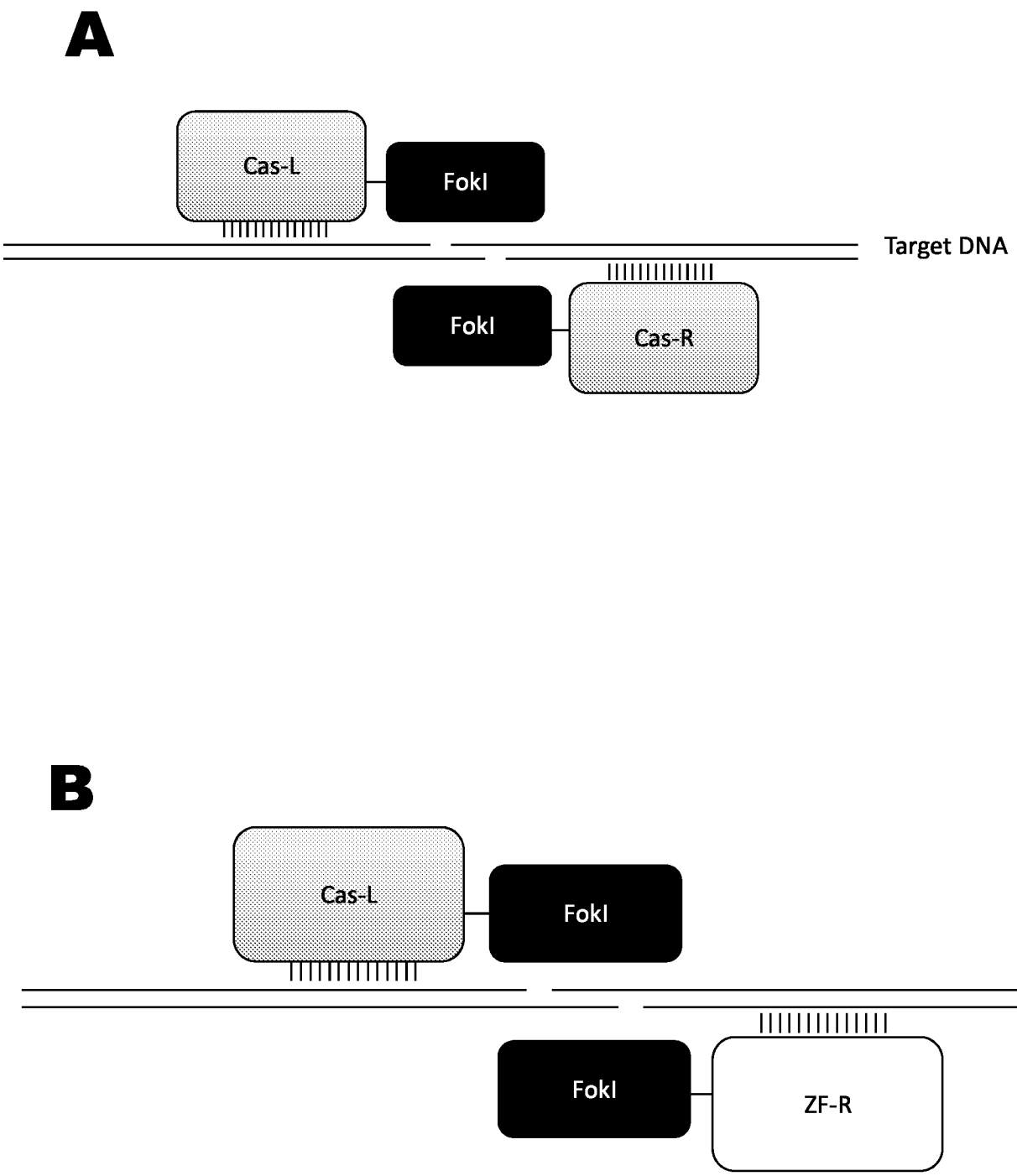


FIG. 1

# CRISPR-BASED GENOME MODIFICATION AND REGULATION

## FIELD OF THE INVENTION

[0001] The present disclosure relates targeted genome modification. In particular, the disclosure relates to RNA-guided endonucleases or fusion proteins comprising CRISPR/Cas-like protein and methods of using said proteins to modify or regulate targeted chromosomal sequences.

## BACKGROUND OF THE INVENTION

[0002] Targeted genome modification is a powerful tool for genetic manipulation of eukaryotic cells, embryos, and animals. For example, exogenous sequences can be integrated at targeted genomic locations and/or specific endogenous chromosomal sequences can be deleted, inactivated, or modified. Current methods rely on the use of engineered nuclease enzymes, such as, for example, zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs). These chimeric nucleases contain programmable, sequence-specific DNA-binding modules linked to a nonspecific DNA cleavage domain. Each new genomic target, however, requires the design of a new ZFN or TALEN comprising a novel sequence-specific DNA-binding module. Thus, these custom designed nucleases tend to be costly and time-consuming to prepare. Moreover, the specificities of ZFNs and TALENS are such that they can mediate off-target cleavages.

[0003] Thus, there is a need for a targeted genome modification technology that does not require the design of a new nuclease for each new targeted genomic location. Additionally, there is a need for a technology with increased specificity with few or no off-target effects.

## SUMMARY OF THE INVENTION

[0004] Among the various aspects of the present disclosure is the provision of an isolated RNA-guided endonuclease, wherein the endonuclease comprises at least one nuclear localization signal, at least one nuclease domain, and at

least one domain that interacts with a guide RNA to target the endonuclease to a specific nucleotide sequence for cleavage. In one embodiment, the endonuclease can be derived from a Cas9 protein. In another embodiment, the endonuclease can be modified to lack at least one functional nuclease domain. In other embodiments, the endonuclease can further comprise a cell-penetrating domain, a marker domain, or both. In a further embodiment, the endonuclease can be part of a protein-RNA complex comprising the guide RNA. In some instances, the guide RNA can be a single molecule comprising a 5' region that is complementary to a target site. Also provided is an isolated nucleic acid encoding any of the RNA-guided endonucleases disclosed herein. In some embodiments, the nucleic acid can be codon optimized for translation in mammalian cells, such as, for example, human cells. In other embodiments, the nucleic acid sequence encoding the RNA-guided endonuclease can be operably linked to a promoter control sequence, and optionally, can be part of a vector. In other embodiments, a vector comprising sequence encoding the RNA-guided endonuclease, which can be operably linked to a promoter control sequence, can also comprise sequence encoding a guide RNA, which can be operably linked to a promoter control sequence.

[0005] Another aspect of the present invention encompasses a method for modifying a chromosomal sequence in a eukaryotic cell or embryo. The method comprises introducing into a eukaryotic cell or embryo (i) at least one RNA-guided endonuclease comprising at least one nuclear localization signal or nucleic acid encoding at least one RNA-guided endonuclease as defined herein, (ii) at least one guide RNA or DNA encoding at least one guide RNA, and, optionally, (iii) at least one donor polynucleotide comprising a donor sequence. The method further comprises culturing the cell or embryo such that each guide RNA directs a RNA-guided endonuclease to a targeted site in the chromosomal sequence where the RNA-guided endonuclease introduces a double-stranded break in the targeted site, and the double-stranded break is repaired by a DNA repair process such that the chromosomal sequence is modified. In one embodiment, the RNA-guided endonuclease can be derived from a Cas9 protein. In another embodiment, the nucleic acid encoding the RNA-guided endonuclease introduced into the cell or embryo can be mRNA. In a

further embodiment, wherein the nucleic acid encoding the RNA-guided endonuclease introduced into the cell or embryo can be DNA. In a further embodiment, the DNA encoding the RNA-guided endonuclease can be part of a vector that further comprises a sequence encoding the guide RNA. In certain embodiments, the eukaryotic cell can be a human cell, a non-human mammalian cell, a stem cell, a non-mammalian vertebrate cell, an invertebrate cell, a plant cell, or a single cell eukaryotic organism. In certain other embodiments, the embryo is a non-human one cell animal embryo.

[0006] A further aspect of the disclosure provides a fusion protein comprising a CRISPR/Cas-like protein or fragment thereof and an effector domain. In general, the fusion protein comprises at least one nuclear localization signal. The effector domain of the fusion protein can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. In one embodiment, the CRISPR/Cas-like protein of the fusion protein can be derived from a Cas9 protein. In one iteration, the Cas9 protein can be modified to lack at least one functional nuclease domain. In an alternate iteration, the Cas9 protein can be modified to lack all nuclease activity. In one embodiment, the effector domain can be a cleavage domain, such as, for example, a FokI endonuclease domain or a modified FokI endonuclease domain. In another embodiment, one fusion protein can form a dimer with another fusion protein. The dimer can be a homodimer or a heterodimer. In another embodiment, the fusion protein can form a heterodimer with a zinc finger nuclease, wherein the cleavage domain of both the fusion protein and the zinc finger nucleases is a FokI endonuclease domain or a modified FokI endonuclease domain. In still another embodiment, the fusion protein comprises a CRISPR/Cas-like protein derived from a Cas9 protein modified to lack all nuclease activity, and the effector domain is a FokI endonuclease domain or a modified FokI endonuclease domain. In still another embodiment, the fusion protein comprises a CRISPR/Cas-like protein derived from a Cas9 protein modified to lack all nuclease activity, and the effector domain can be an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. In additional embodiments, any of the fusion proteins disclosed herein can comprise at least one additional domain chosen from a nuclear localization signal, a cell-penetrating domain, and a marker domain.

Also provided are isolated nucleic acids encoding any of the fusion proteins provided herein.

[0007] Still another aspect of the disclosure encompasses a method for modifying a chromosomal sequence or regulating expression of a chromosomal sequence in a cell or embryo. The method comprises introducing into the cell or embryo (a) at least one fusion protein or nucleic acid encoding at least one fusion protein, wherein the fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof and an effector domain, and (b) at least one guide RNA or DNA encoding at least one guide RNA, wherein the guide RNA guides the CRISPR/Cas-like protein of the fusion protein to a targeted site in the chromosomal sequence and the effector domain of the fusion protein modifies the chromosomal sequence or regulates expression of the chromosomal sequence. In one embodiment, the CRISPR/Cas-like protein of the fusion protein can be derived from a Cas9 protein. In another embodiment, the CRISPR/Cas-like protein of the fusion protein can be modified to lack at least one functional nuclease domain. In still another embodiment, the CRISPR/Cas-like protein of the fusion protein can be modified to lack all nuclease activity. In one embodiment in which the fusion protein comprises a Cas9 protein modified to lack all nuclease activity and a FokI cleavage domain or a modified FokI cleavage domain, the method can comprise introducing into the cell or embryo one fusion protein or nucleic acid encoding one fusion protein and two guide RNAs or DNA encoding two guide RNAs, and wherein one double-stranded break is introduced in the chromosomal sequence. In another embodiment in which the fusion protein comprises a Cas9 protein modified to lack all nuclease activity and a FokI cleavage domain or a modified FokI cleavage domain, the method can comprise introducing into the cell or embryo two fusion proteins or nucleic acid encoding two fusion proteins and two guide RNAs or DNA encoding two guide RNAs, and wherein two double-stranded breaks are introduced in the chromosomal sequence. In still another one embodiment in which the fusion protein comprises a Cas9 protein modified to lack all nuclease activity and a FokI cleavage domain or a modified FokI cleavage domain, the method can comprise introducing into the cell or embryo one fusion protein or nucleic acid encoding one fusion protein, one guide RNA or nucleic acid encoding one guide RNA, and one zinc finger nuclease or nucleic acid

encoding one zinc finger nuclease, wherein the zinc finger nuclease comprises a FokI cleavage domain or a modified a FokI cleavage domain, and wherein one double-stranded break is introduced into the chromosomal sequence. In certain embodiments in which the fusion protein comprises a cleavage domain, the method can further comprise introducing into the cell or embryo at least one donor polynucleotide. In embodiments in which the fusion protein comprises an effector domain chosen from an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain, the fusion protein can comprise a Cas9 protein modified to lack all nuclease activity, and the method can comprise introducing into the cell or embryo one fusion protein or nucleic acid encoding one fusion protein, and one guide RNA or nucleic acid encoding one guide RNA, and wherein the structure or expression of the targeted chromosomal sequence is modified. In certain embodiments, the eukaryotic cell can be a human cell, a non-human mammalian cell, a stem cell, a non-mammalian vertebrate cell, an invertebrate cell, a plant cell, or a single cell eukaryotic organism. In certain other embodiments, the embryo is a non-human one cell animal embryo.

[0008] Other aspects and iterations of the disclosure are detailed below.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0009] **FIG. 1** diagrams genome modification using protein dimers. (**A**) depicts a double stranded break created by a dimer composed of two fusion proteins, each of which comprises a Cas-like protein for DNA binding and a FokI cleavage domain. (**B**) depicts a double stranded break created by a dimer composed of a fusion protein comprising a Cas-like protein and a FokI cleavage domain and a zinc finger nuclease comprising a zinc finger (ZF) DNA-binding domain and a FokI cleavage domain.

[0010] **FIG. 2** illustrates regulation of gene expression using RNA-guided fusion proteins comprising gene regulatory domains. (**A**) depicts a fusion protein comprising a Cas-like protein used for DNA binding and an "A/R" domain that activates or represses gene expression. (**B**) diagrams a fusion protein comprising a Cas-like protein for DNA binding and a epigenetic modification domain ("Epi-mod') that affects epigenetic states by covalent modification of proximal DNA or proteins.

[0011] **FIG. 3** diagrams genome modification using two RNA-guided endonuclease. (**A**) depicts a double stranded break created by two RNA-guided endonuclease that have been converted into nickases. (**B**) depicts two double stranded breaks created by two RNA-guided endonuclease having endonuclease activity.

[0012] **FIG. 4** presents fluorescence-activated cell sorting (FACS) of human K562 cells transfected with Cas9 nucleic acid, Cas9 guiding RNA, and AAVS1-GFP DNA donor. The Y axis represents the auto fluorescence intensity at a red channel, and the X axis represents the green fluorescence intensity. (**A**) K562 cells transfected with 10 μg of Cas9 mRNA transcribed with an Anti-Reverse Cap Analog, 0.3 nmol of pre-annealed crRNA-tracrRNA duplex, and 10 μg of AAVS1-GFP plasmid DNA; (**B**) K562 cells transfected 10 μg of Cas9 mRNA transcribed with an Anti-Reverse Cap Analog, 0.3 nmol of chimeric RNA, and 10 μg of AAVS1-GFP plasmid DNA; (**C**) K562 cells transfected 10 μg of Cas9 mRNA that was capped by post-transcription capping reaction, 0.3 nmol of chimeric RNA, and 10 μg of AAVS1-GFP plasmid DNA; (**D**) K562 cells transfected with 10 μg of Cas9 plasmid DNA, 5 μg of U6-chimeric RNA plasmid DNA, and 10 μg of AAVS1-GFP plasmid DNA; (**E**) K562 cells transfected with 10 μg of AAVS1-GFP plasmid DNA; (**F**) K562 cells transfected with transfection reagents only.

[0013] **FIG. 5** presents a junction PCR analysis documenting the targeted integration of GFP into the AAVS1 locus in human cells. Lane M: 1 kb DNA molecular markers; Lane A: K562 cells transfected with 10 μg of Cas9 mRNA transcribed with an Anti-Reverse Cap Analog, 0.3 nmol of pre-annealed crRNA-tracrRNA duplex, and 10 μg of AAVS1-GFP plasmid DNA; Lane B: K562 cells transfected 10 μg of Cas9 mRNA transcribed with an Anti-Reverse Cap Analog, 0.3 nmol of chimeric RNA, and 10 μg of AAVS1-GFP plasmid DNA; Lane C: K562 cells transfected 10 μg of Cas9 mRNA that was capped by post-transcription capping reaction, 0.3 nmol of chimeric RNA, and 10 μg of AAVS1-GFP plasmid DNA; Lane D: K562 cells transfected with 10 μg of Cas9 plasmid DNA, 5 μg of U6-chimeric RNA plasmid DNA, and 10 μg of AAVS1-GFP plasmid DNA; Lane E: K562 cells transfected with 10 μg of AAVS1-GFP plasmid DNA; Lane F: K562 cells transfected with transfection reagents only.

## DETAILED DESCRIPTION OF THE INVENTION

[0014] Provided herein are RNA-guided endonucleases, which comprise at least one nuclear localization signal, at least one nuclease domain, and at least one domain that interacts with a guide RNA to target the endonuclease to a specific nucleotide sequence for cleavage. Also provided are nucleic acids encoding the RNA-guided endonucleases, as well as methods of using the RNA-guided endonucleases to modify chromosomal sequences of eukaryotic cells or embryos. The RNA-guided endonuclease interacts with specific guide RNAs, each of which directs the endonuclease to a specific targeted site, at which site the RNA-guided endonuclease introduces a double-stranded break that can be repaired by a DNA repair process such that the chromosomal sequence is modified. Since the specificity is provided by the guide RNA, the RNA-based endonuclease is universal and can be used with different guide RNAs to target different genomic sequences. The methods disclosed herein can be used to target and modify specific chromosomal sequences and/or introduce exogenous sequences at targeted locations in the genome of cells or embryos. Furthermore, the targeting is specific with limited off target effects.

[0015] The present disclosure provides fusion proteins, wherein a fusion protein comprises a CRISPR/Cas-like protein or fragment thereof and an effector domain. Suitable effector domains include, without limit, cleavage domains, epigenetic modification domains, transcriptional activation domains, and transcriptional repressor domains. Each fusion protein is guided to a specific chromosomal sequence by a specific guide RNA, wherein the effector domain mediates targeted genome modification or gene regulation. In one aspect, the fusion proteins can function as dimers thereby increasing the length of the target site and increasing the likelihood of its uniqueness in the genome (thus, reducing off target effects). For example, endogenous CRISPR systems modify genomic locations based on DNA binding word lengths of approximately 13-20 bp (Cong et al., Science, 339:819-823). At this word size, only 5-7% of the target sites are unique within the genome (Iseli et al, PLos One 2(6):e579). In contrast, DNA binding word sizes for zinc finger nucleases typically range from 30-36 bp, resulting in target sites that are approximately 85-87% unique within the human genome. The smaller sized DNA binding sites utilized by CRISPR-based systems limits

and complicates design of targeted CRISP-based nucleases near desired locations, such as disease SNPs, small exons, start codons, and stop codons, as well as other locations within complex genomes. The present disclosure not only provides means for expanding the CRISPR DNA binding word length (i.e., so as to limit off-target activity), but further provides CRISPR fusion proteins having modified functionality. According, the disclosed CRISPR fusion proteins have increased target specificity and unique functionality(ies). Also provided herein are methods of using the fusion proteins to modify or regulate expression of targeted chromosomal sequences.

### *(I) RNA-Guided Endonucleases*

[0016] One aspect of the present disclosure provides RNA-guided endonucleases comprising at least one nuclear localization signal, which permits entry of the endonuclease into the nuclei of eukaryotic cells and embryos such as, for example, non-human one cell embryos. RNA-guided endonucleases also comprise at least one nuclease domain and at least one domain that interacts with a guide RNA. An RNA-guided endonuclease is directed to a specific nucleic acid sequence (or target site) by a guide RNA. The guide RNA interacts with the RNA-guided endonuclease as well as the target site such that, once directed to the target site, the RNA-guided endonuclease is able to introduce a double-stranded break into the target site nucleic acid sequence. Since the guide RNA provides the specificity for the targeted cleavage, the endonuclease of the RNA-guided endonuclease is universal and can be used with different guide RNAs to cleave different target nucleic acid sequences. Provided herein are isolated RNA-guided endonucleases, isolated nucleic acids (i.e., RNA or DNA) encoding the RNA-guided endonucleases, vectors comprising nucleic acids encoding the RNA-guided endonucleases, and protein-RNA complexes comprising the RNA-guided endonuclease plus a guide RNA.

[0017] The RNA-guided endonuclease can be derived from a clustered regularly interspersed short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system. The CRISPR/Cas system can be a type I, a type II, or a type III system. Non-limiting examples of suitable CRISPR/Cas proteins include Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10,

Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (or CasA), Cse2 (or CasB), Cse3 (or CasE), Cse4 (or CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3,Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cu1966.

[0018] In one embodiment, the RNA-guided endonuclease is derived from a type II CRISPR/Cas system. In specific embodiments, the RNA-guided endonuclease is derived from a Cas9 protein. The Cas9 protein can be from *Streptococcus pyogenes, Streptococcus thermophilus, Streptococcus sp., Nocardiopsis dassonvillei, Streptomyces pristinaespiralis, Streptomyces viridochromogenes, Streptomyces viridochromogenes, Streptosporangium roseum, Streptosporangium roseum, Alicyclobacillus acidocaldarius, Bacillus pseudomycoides, Bacillus selenitireducens, Exiguobacterium sibiricum, Lactobacillus delbrueckii, Lactobacillus salivarius, Microscilla marina, Burkholderiales bacterium, Polaromonas naphthalenivorans, Polaromonas sp., Crocosphaera watsonii, Cyanothece sp., Microcystis aeruginosa, Synechococcus sp., Acetohalobium arabaticum, Ammonifex degensii, Caldicelulosiruptor becscii, Candidatus Desulforudis, Clostridium botulinum, Clostridium difficile, Finegoldia magna, Natranaerobius thermophilus, Pelotomaculum thermopropionicum, Acidithiobacillus caldus, Acidithiobacillus ferrooxidans, Allochromatium vinosum, Marinobacter sp., Nitrosococcus halophilus, Nitrosococcus watsoni, Pseudoalteromonas haloplanktis, Ktedonobacter racemifer, Methanohalobium evestigatum, Anabaena variabilis, Nodularia spumigena, Nostoc sp., Arthrospira maxima, Arthrospira platensis, Arthrospira sp., Lyngbya sp., Microcoleus chthonoplastes, Oscillatoria sp., Petrotoga mobilis, Thermosipho africanus*, or *Acaryochloris marina*.

[0019] In general, CRISPR/Cas proteins comprise at least one RNA recognition and/or RNA binding domain. RNA recognition and/or RNA binding domains interact with guide RNAs. CRISPR/Cas proteins can also comprise nuclease domains (i.e., DNase or RNase domains), DNA binding domains, helicase domains, RNAse domains, protein-protein interaction domains, dimerization domains, as well as other domains.

[0020] The CRISPR/Cas-like protein can be a wild type CRISPR/Cas protein, a modified CRISPR/Cas protein, or a fragment of a wild type or modified CRISPR/Cas protein. The CRISPR/Cas-like protein can be modified to increase nucleic acid binding affinity and/or specificity, alter an enzymatic activity, and/or change another property of the protein. For example, nuclease (i.e., DNase, RNase) domains of the CRISPR/Cas-like protein can be modified, deleted, or inactivated. Alternatively, the CRISPR/Cas-like protein can be truncated to remove domains that are not essential for the function of the fusion protein. The CRISPR/Cas-like protein can also be truncated or modified to optimize the activity of the effector domain of the fusion protein.

[0021] In some embodiments, the CRISPR/Cas-like protein can be derived from a wild type Cas9 protein or fragment thereof. In other embodiments, the CRISPR/Cas-like protein can be derived from modified Cas9 protein. For example, the amino acid sequence of the Cas9 protein can be modified to alter one or more properties (e.g., nuclease activity, affinity, stability, etc.) of the protein. Alternatively, domains of the Cas9 protein not involved in RNA-guided cleavage can be eliminated from the protein such that the modified Cas9 protein is smaller than the wild type Cas9 protein.

[0022] In general, a Cas9 protein comprises at least two nuclease (i.e., DNase) domains. For example, a Cas9 protein can comprise a RuvC-like nuclease domain and a HNH-like nuclease domain. The RuvC and HNH domains work together to cut single strands to make a double-stranded break in DNA. (Jinek et al., Science, 337: 816-821). In some embodiments, the Cas9-derived protein can be modified to contain only one functional nuclease domain (either a RuvC-like or a HNH-like nuclease domain). For example, the Cas9-derived protein can be modified such that one of the nuclease domains is deleted or mutated such that it is no longer functional (i.e., the nuclease activity is absent). In some embodiments in which one of the nuclease domains is inactive, the Cas9-derived protein is able to introduce a nick into a double-stranded nucleic acid (such protein is termed a “nickase”), but not cleave the double-stranded DNA. For example, an aspartate to alanine (D10A) conversion in a RuvC-like domain converts the Cas9-derived protein into a nickase. Likewise, a histidine to alanine (H840A or H839A) conversion in a HNH domain converts the Cas9-derived

protein into a nickase. Each nuclease domain can be modified using well-known methods, such as site-directed mutagenesis, PCR-mediated mutagenesis, and total gene synthesis, as well as other methods known in the art.

[0023] The RNA-guided endonuclease disclosed herein comprises at least one nuclear localization signal. In general, an NLS comprises a stretch of basic amino acids. Nuclear localization signals are known in the art (see, e.g., Lange et al., J. Biol. Chem., 2007, 282:5101-5105). For example, in one embodiment, the NLS can be a monopartite sequence, such as PKKKRKV (SEQ ID NO:1) or PKKKRRV (SEQ ID NO:2). In another embodiment, the NLS can be a bipartite sequence. In still another embodiment, the NLS can be KRPAATKKAGQAKKKK (SEQ ID NO:3). The NLS can be located at the N-terminus, the C-terminal, or in an internal location of the RNA-guided endonuclease.

[0024] In some embodiments, the RNA-guided endonuclease can further comprise at least one cell-penetrating domain. In one embodiment, the cell-penetrating domain can be a cell-penetrating peptide sequence derived from the HIV-1 TAT protein. As an example, the TAT cell-penetrating sequence can be GRKKRRQRRRPPQPKKKRKV (SEQ ID NO:4). In another embodiment, the cell-penetrating domain can be TLM (PLSSIFSRIGDPPKKKRKV; SEQ ID NO:5), a cell-penetrating peptide sequence derived from the human hepatitis B virus. In still another embodiment, the cell-penetrating domain can be MPG (GALFLGWLGAAGSTMGAPKKKRKV; SEQ ID NO:6 or GALFLGFLGAAGSTMGAWSQPKKKRKV; SEQ ID NO:7). In an additional embodiment, the cell-penetrating domain can be Pep-1 (KETWWETWWTEWSQPKKKRKV; SEQ ID NO:8), VP22, a cell penetrating peptide from Herpes simplex virus, or a polyarginine peptide sequence. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or in an internal location of the protein.

[0025] In still other embodiments, the RNA-guided endonuclease can also comprise at least one marker domain. Non-limiting examples of marker domains include fluorescent proteins, purification tags, and epitope tags. In some embodiments, the marker domain can be a fluorescent protein. Non limiting examples of suitable

fluorescent proteins include green fluorescent proteins (e.g., GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (e.g. YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1,), blue fluorescent proteins (e.g. EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire,), cyan fluorescent proteins (e.g. ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRasberry, mStrawberry, Jred), and orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) or any other suitable fluorescent protein. In other embodiments, the marker domain can be a purification tag and/or an epitope tag. Exemplary tags include, but are not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, biotin carboxyl carrier protein (BCCP), and calmodulin.

[0026] In certain embodiments, the RNA-guided endonuclease may be part of a protein-RNA complex comprising a guide RNA. The guide RNA interacts with the RNA-guided endonuclease to direct the endonuclease to a specific target site, wherein the 5' end of the guide RNA base pairs with a specific protospacer sequence.

### *(II) Fusion Proteins*

[0027] Another aspect of the present disclosure provides a fusion protein comprising a CRISPR/Cas-like protein or fragment thereof and an effector domain. The CRISPR/Cas-like protein is directed to a target site by a guide RNA, at which site the effector domain can modify or effect the targeted nucleic acid sequence. The effector domain can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. The fusion protein can further comprise at least one additional domain chosen from a nuclear localization signal, a cell-penetrating domain, or a marker domain.

**(a) CRISPR/Cas-like protein**

[0028] The fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof. CRISPR/Cas-like proteins are detailed above in section (I). The CRISPR/Cas-like protein can be located at the N-terminus, the C-terminus, or in an internal location of the fusion protein

[0029] In some embodiments, the CRISPR/Cas-like protein of the fusion protein can be derived from a Cas9 protein. The Cas9-derived protein can be wild type, modified, or a fragment thereof. In some embodiments, the Cas9-derived protein can be modified to contain only one functional nuclease domain (either a RuvC-like or a HNH-like nuclease domain). For example, the Cas9-derived protein can be modified such that one of the nuclease domains is deleted or mutated such that it is no longer functional (i.e., the nuclease activity is absent). In some embodiments in which one of the nuclease domains is inactive, the Cas9-derived protein is able to introduce a nick into a double-stranded nucleic acid (such protein is termed a “nickase”), but not cleave the double-stranded DNA. For example, an aspartate to alanine (D10A) conversion in a RuvC-like domain converts the Cas9-derived protein into a nickase. Likewise, a histidine to alanine (H840A or H839A) conversion in a HNH domain converts the Cas9-derived protein into a nickase. In other embodiments, both of the RuvC-like nuclease domain and the HNH-like nuclease domain can be modified or eliminated such that the Cas9-derived protein is unable to nick or cleave double stranded nucleic acid. In still other embodiments, all nuclease domains of the Cas9-derived protein can be modified or eliminated such that the Cas9-derived protein lacks all nuclease activity.

[0030] In any of the above-described embodiments, any or all of the nuclease domains can be inactivated by one or more deletion mutations, insertion mutations, and/or substitution mutations using well-known methods, such as site-directed mutagenesis, PCR-mediated mutagenesis, and total gene synthesis, as well as other methods known in the art. In an exemplary embodiment, the CRISPR/Cas-like protein of the fusion protein is derived from a Cas9 protein in which all the nuclease domains have been inactivated or deleted.

**(b) Effector domain**

[0031] The fusion protein also comprises an effector domain. The effector domain can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. The effector domain can be located at the N-terminus, the C-terminus, or in an internal location of the fusion protein.

(i) cleavage domain

[0032] In some embodiments, the effector domain is a cleavage domain. As used herein, a "cleavage domain" refers to a domain that cleaves DNA. The cleavage domain can be obtained from any endonuclease or exonuclease. Non-limiting examples of endonucleases from which a cleavage domain can be derived include, but are not limited to, restriction endonucleases and homing endonucleases. See, for example, New England Biolabs Catalog or Belfort et al. (1997) Nucleic Acids Res. 25:3379-3388. Additional enzymes that cleave DNA are known (e.g., S1 Nuclease; mung bean nuclease; pancreatic DNase I; micrococcal nuclease; yeast HO endonuclease). See also Linn et al. (eds.) Nucleases, Cold Spring Harbor Laboratory Press, 1993. One or more of these enzymes (or functional fragments thereof) can be used as a source of cleavage domains.

[0033] In some embodiments, the cleavage domain can be derived from a type II-S endonuclease. Type II-S endonucleases cleave DNA at sites that are typically several base pairs away the recognition site and, as such, have separable recognition and cleavage domains. These enzymes generally are monomers that transiently associate to form dimers to cleave each strand of DNA at staggered locations. Non-limiting examples of suitable type II-S endonucleases include BfiI, BpmI, BsaI, BsgI, BsmBI, BsmI, BspMI, FokI, MboII, and SapI. In exemplary embodiments, the cleavage domain of the fusion protein is a FokI cleavage domain or a derivative thereof.

[0034] In certain embodiments, the type II-S cleavage can be modified to facilitate dimerization of two different cleavage domains (each of which is attached to a CRISPR/Cas-like protein or fragment thereof). For example, the cleavage domain of FokI can be modified by mutating certain amino acid residues. By way of non-limiting example, amino acid residues at positions 446, 447, 479, 483, 484, 486, 487, 490, 491,

496, 498, 499, 500, 531, 534, 537, and 538 of FokI cleavage domains are targets for modification. For example, modified cleavage domains of FokI that form obligate heterodimers include a pair in which a first modified cleavage domain includes mutations at amino acid positions 490 and 538 and a second modified cleavage domain that includes mutations at amino acid positions 486 and 499 (Miller et al., 2007, Nat. Biotechnol, 25:778-785; Szczpek et al., 2007, Nat. Biotechnol, 25:786-793). For example, the Glu (E) at position 490 can be changed to Lys (K) and the Ile (I) at position 538 can be changed to K in one domain (E490K, I538K), and the Gln (Q) at position 486 can be changed to E and the I at position 499 can be changed to Leu (L) in another cleavage domain (Q486E, I499L). In other embodiments, modified FokI cleavage domains can include three amino acid changes (Doyon et al. 2011, Nat. Methods, 8:74-81). For example, one modified FokI domain (which is termed ELD) can comprise Q486E, I499L, N496D mutations and the other modified FokI domain (which is termed KKR) can comprise E490K, I538K, H537R mutations.

[0035] In exemplary embodiments, the effector domain of the fusion protein is a FokI cleavage domain or a modified FokI cleavage domain.

[0036] In embodiments wherein the effector domain is a cleavage domain and the CRISPR/Cas-like protein is derived from a Cas9 protein, the Cas9-derived can be modified as discussed herein such that its endonuclease activity is eliminated. For example, the Cas9-derived can be modified by mutating the RuvC and HNH domains such that they no longer possess nuclease activity.

(ii) epigenetic modification domain

[0037] In other embodiments, the effector domain of the fusion protein can be an epigenetic modification domain. In general, epigenetic modification domains alter histone structure and/or chromosomal structure without altering the DNA sequence. Changes histone and/or chromatin structure can lead to changes in gene expression. Examples of epigenetic modification include, without limit, acetylation or methylation of lysine residues in histone proteins, and methylation of cytosine residues in DNA. Non-limiting examples of suitable epigenetic modification domains include histone acetyltansferase domains, histone deacetylase domains, histone methyltransferase

domains, histone demethylase domains, DNA methyltransferase domains, and DNA demethylase domains.

[0038] In embodiments in which the effector domain is a histone acetyltansferase (HAT) domain, the HAT domain can be derived from EP300 (i.e., E1A binding protein p300), CREBBP (i.e., CREB-binding protein), CDY1, CDY2, CDYL1, CLOCK, ELP3, ESA1, GCN5 (KAT2A), HAT1,KAT2B, KAT5, MYST1, MYST2, MYST3, MYST4, NCOA1, NCOA2, NCOA3, NCOAT, P/CAF, Tip60, TAFII250, or TF3C4. In one such embodiment, the HAT domain is p300

[0039] In embodiments wherein the effector domain is an epigenetic modification domain and the CRISPR/Cas-like protein is derived from a Cas9 protein, the Cas9-derived can be modified as discussed herein such that its endonuclease activity is eliminated. For example, the Cas9-derived can be modified by mutating the RuvC and HNH domains such that they no longer possess nuclease activity.

(iii) transcriptional activation domain

[0040] In other embodiments, the effector domain of the fusion protein can be a transcriptional activation domain. In general, a transcriptional activation domain interacts with transcriptional control elements and/or transcriptional regulatory proteins (i.e., transcription factors, RNA polymerases, etc.) to increase and/or activate transcription of a gene. In some embodiments, the transcriptional activation domain can be, without limit, a herpes simplex virus VP16 activation domain, VP64 (which is a tetrameric derivative of VP16), a NFκB p65 activation domain, p53 activation domains 1 and 2, a CREB (cAMP response element binding protein) activation domain, an E2A activation domain, and an NFAT (nuclear factor of activated T-cells) activation domain. In other embodiments, the transcriptional activation domain can be Gal4, Gcn4, MLL, Rtg3, Gln3, Oaf1, Pip2, Pdr1, Pdr3, Pho4, and Leu3. The transcriptional activation domain may be wild type, or it may be a modified version of the original transcriptional activation domain. In some embodiments, the effector domain of the fusion protein is a VP16 or VP64 transcriptional activation domain.

[0041] In embodiments wherein the effector domain is a transcriptional activation domain and the CRISPR/Cas-like protein is derived from a Cas9 protein, the

Cas9-derived protein can be modified as discussed herein such that its endonuclease activity is eliminated. For example, the Cas9-derived can be modified by mutating the RuvC and HNH domains such that they no longer possess nuclease activity.

(iv) transcriptional repressor domain

[0042] In still other embodiments, the effector domain of the fusion protein can be a transcriptional repressor domain. In general, a transcriptional repressor domain interacts with transcriptional control elements and/or transcriptional regulatory proteins (i.e., transcription factors, RNA polymerases, etc.) to decrease and/or terminate transcription of a gene. Non-limiting examples of suitable transcriptional repressor domains include inducible cAMP early repressor (ICER) domains, Kruppel-associated box A (KRAB-A) repressor domains, YY1 glycine rich repressor domains, Sp1-like repressors, E(spl) repressors, IκB repressor, and MeCP2.

[0043] In embodiments wherein the effector domain is a transcriptional repressor domain and the CRISPR/Cas-like protein is derived from a Cas9 protein, the Cas9-derived protein can be modified as discussed herein such that its endonuclease activity is eliminated. For example, the cas9 can be modified by mutating the RuvC and HNH domains such that they no longer possess nuclease activity.

**(c) Additional domains**

[0044] In some embodiments, the fusion protein further comprises at least one additional domain. Non-limiting examples of suitable additional domains include nuclear localization signals, cell-penetrating or translocation domains, and marker domains. Non-limiting examples of suitable nuclear localization signals, cell-penetrating domains, and marker domains are presented above in section (I).

**(d) Fusion protein dimers**

[0045] In embodiments in which the effector domain of the fusion protein is a cleavage domain, a dimer comprising at least one fusion protein can form. The dimer can be a homodimer or a heterodimer. In some embodiments, the heterodimer

comprises two different fusion proteins. In other embodiments, the heterodimer comprises one fusion protein and an additional protein.

[0046] In some embodiments, the dimer is a homodimer in which the two fusion protein monomers are identical with respect to the primary amino acid sequence. In one embodiment where the dimer is a homodimer, the Cas9-derived proteins are modified such that their endonuclease activity is eliminated, i.e., such that they have no functional nuclease domains. In certain embodiments wherein the Cas9-derived proteins are modified such that their endonuclease activity is eliminated, each fusion protein monomer comprises an identical Cas9 like protein and an identical cleavage domain. The cleavage domain can be any cleavage domain, such as any of the exemplary cleavage domains provided herein. In one specific embodiment, the cleavage domain is a FokI cleavage domain or a modified FokI cleavage domain. In such embodiments, specific guide RNAs would direct the fusion protein monomers to different but closely adjacent sites such that, upon dimer formation, the nuclease domains of the two monomers would create a double stranded break in the target DNA.

[0047] In other embodiments, the dimer is a heterodimer of two different fusion proteins. For example, the CRISPR/Cas-like protein of each fusion protein can be derived from a different CRISPR/Cas protein or from an orthologous CRISPR/Cas protein from a different bacterial species. For example, each fusion protein can comprise a Cas9-like protein, which Cas9-like protein is derived from a different bacterial species. In these embodiments, each fusion protein would recognize a different target site (i.e., specified by the protospacer and/or PAM sequence). For example, the guide RNAs could position the heterodimer to different but closely adjacent sites such that their nuclease domains results in an effective double stranded break in the target DNA. The heterodimer can also have modified Cas9 proteins with nicking activity such that the nicking locations are different.

[0048] Alternatively, two fusion proteins of a heterodimer can have different effector domains. In embodiments in which the effector domain is a cleavage domain, each fusion protein can contain a different modified cleavage domain. For example, each fusion protein can contain a different modified FokI cleavage domain, as

detailed above in section (II)(b)(i). In these embodiments, the Cas-9 proteins can be modified such that their endonuclease activities are eliminated.

[0049] As will be appreciated by those skilled in the art, the two fusion proteins forming a heterodimer can differ in both the CRISPR/Cas-like protein domain and the effector domain.

[0050] In any of the above-described embodiments, the homodimer or heterodimer can comprise at least one additional domain chosen from nuclear localization signals (NLSs), cell-penetrating, translocation domains and marker domains, as detailed above.

[0051] In any of the above-described embodiments, one or both of the Cas9-derived proteins can be modified such that its endonuclease activity is eliminated or modified.

[0052] In still alternate embodiments, the heterodimer comprises one fusion protein and an additional protein. For example, the additional protein can be a nuclease. In one embodiment, the nuclease is a zinc finger nuclease. A zinc finger nuclease comprises a zinc finger DNA binding domain and a cleavage domain. A zinc finger recognizes and binds three (3) nucleotides. A zinc finger DNA binding domain can comprise from about three zinc fingers to about seven zinc fingers. The zinc finger DNA binding domain can be derived from a naturally occurring protein or it can be engineered. See, for example, Beerli et al. (2002) Nat. Biotechnol. 20:135-141; Pabo et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan et al. (2001) Nat. Biotechnol. 19:656-660; Segal et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al. (2000) Curr. Opin. Struct. Biol. 10:411-416; Zhang et al. (2000) J. Biol. Chem. 275(43):33850-33860; Doyon et al. (2008) Nat. Biotechnol. 26:702-708; and Santiago et al. (2008) Proc. Natl. Acad. Sci. USA 105:5809-5814. The cleavage domain of the zinc finger nuclease can be any cleavage domain detailed above in section (II)(b)(i). In exemplary embodiments, the cleavage domain of the zinc finger nuclease is a FokI cleavage domain or a modified FokI cleavage domain. Such a zinc finger nuclease will dimerize with a fusion protein comprising a FokI cleavage domain or a modified FokI cleavage domain.

[0053] In some embodiments, the zinc finger nuclease can comprise at least one additional domain chosen from nuclear localization signals, cell-penetrating or translocation domains, which are detailed above.

[0054] In certain embodiments, any of the fusion protein detailed above or a dimer comprising at least one fusion protein may be part of a protein-RNA complex comprising at least one guide RNA. A guide RNA interacts with the CRISPR-Cas0like protein of the fusion protein to direct the fusion protein to a specific target site, wherein the 5' end of the guide RNA base pairs with a specific protospacer sequence.

### *(III) Nucleic Acids Encoding RNA-Guided Endonucleases or Fusion Proteins*

[0055] Another aspect of the present disclosure provides nucleic acids encoding any of the RNA-guided endonucleases or fusion proteins described above in sections (I) and (II), respectively. The nucleic acid can be RNA or DNA. In one embodiment, the nucleic acid encoding the RNA-guided endonuclease or fusion protein is mRNA. The mRNA can be 5' capped and/or 3' polyadenylated. In another embodiment, the nucleic acid encoding the RNA-guided endonuclease or fusion protein is DNA. The DNA can be present in a vector (see below).

[0056] The nucleic acid encoding the RNA-guided endonuclease or fusion protein can be codon optimized for efficient translation into protein in the eukaryotic cell or animal of interest. For example, codons can be optimized for expression in humans, mice, rats, hamsters, cows, pigs, cats, dogs, fish, amphibians, plants, yeast, insects, and so forth (see Codon Usage Database at www.kazusa.or.jp/codon/). Programs for codon optimization are available as freeware (e.g., OPTIMIZER at genomes.urv.es/OPTIMIZER; OptimumGene™ from GenScript at www.genscript.com/codon_opt.html). Commercial codon optimization programs are also available.

[0057] In some embodiments, DNA encoding the RNA-guided endonuclease or fusion protein can be operably linked to at least one promoter control sequence. In some iterations, the DNA coding sequence can be operably linked to a promoter control sequence for expression in the eukaryotic cell or animal of interest. The promoter control sequence can be constitutive, regulated, or tissue-specific.

Suitable constitutive promoter control sequences include, but are not limited to, cytomegalovirus immediate early promoter (CMV), simian virus (SV40) promoter, adenovirus major late promoter, Rous sarcoma virus (RSV) promoter, mouse mammary tumor virus (MMTV) promoter, phosphoglycerate kinase (PGK) promoter, elongation factor (ED1)-alpha promoter, ubiquitin promoters, actin promoters, tubulin promoters, immunoglobulin promoters, fragments thereof, or combinations of any of the foregoing. Examples of suitable regulated promoter control sequences include without limit those regulated by heat shock, metals, steroids, antibiotics, or alcohol. Non-limiting examples of tissue-specific promoters include B29 promoter, CD14 promoter, CD43 promoter, CD45 promoter, CD68 promoter, desmin promoter, elastase-1 promoter, endoglin promoter, fibronectin promoter, Flt-1 promoter, GFAP promoter, GPIIb promoter, ICAM-2 promoter, INF-β promoter, Mb promoter, Nphsl promoter, OG-2 promoter, SP-B promoter, SYN1 promoter, and WASP promoter. The promoter sequence can be wild type or it can be modified for more efficient or efficacious expression. In one exemplary embodiment, the encoding DNA can be operably linked to a CMV promoter for constitutive expression in mammalian cells.

[0058] In certain embodiments, the sequence encoding the RNA-guided endonuclease or fusion protein can be operably linked to a promoter sequence that is recognized by a phage RNA polymerase for *in vitro* mRNA synthesis. In such embodiments, the *in vitro*-transcribed RNA can be purified for use in the methods detailed below in sections (IV) and (V). For example, the promoter sequence can be a T7, T3, or SP6 promoter sequence or a variation of a T7, T3, or SP6 promoter sequence. In an exemplary embodiment, the DNA encoding the fusion protein is operably linked to a T7 promoter for *in vitro* mRNA synthesis using T7 RNA polymerase.

[0059] In alternate embodiments, the sequence encoding the RNA-guided endonuclease or fusion protein can be operably linked to a promoter sequence for *in vitro* expression of the RNA-guided endonuclease or fusion protein in bacterial or eukaryotic cells. In such embodiments, the expressed protein can be purified for use in the methods detailed below in sections (IV) and (V). Suitable bacterial promoters include, without limit, T7 promoters, *lac* operon promoters, *trp* promoters, variations thereof, and combinations thereof. An exemplary bacterial promoter is *tac* which is a

hybrid of *trp* and *lac* promoters. Non-limiting examples of suitable eukaryotic promoters are listed above.

[0060] In additional aspects, the DNA encoding the RNA-guided endonuclease or fusion protein also can be linked to a polyadenylation signal (e.g., SV40 polyA signal, bovine growth hormone (BGH) polyA signal, etc.) and/or at least one transcriptional termination sequence. Additionally, the sequence encoding the RNA-guided endonuclease or fusion protein also can be linked to sequence encoding at least one nuclear localization signal, at least one cell-penetrating domain, and/or at least one marker domain, which are detailed above in section (I).

[0061] In various embodiments, the DNA encoding the RNA-guided endonuclease or fusion protein can be present in a vector. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors (e.g., lentiviral vectors, adeno-associated viral vectors, etc.). In one embodiment, the DNA encoding the RNA-guided endonuclease or fusion protein is present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (e.g., enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (e.g., antibiotic resistance genes), origins of replication, and the like. Additional information can be found in “Current Protocols in Molecular Biology" Ausubel et al., John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001.

[0062] In some embodiments, the expression vector comprising the sequence encoding the RNA-guided endonuclease or fusion protein can further comprise sequence encoding a guide RNA. The sequence encoding the guide RNA generally is operably linked to at least one transcriptional control sequence for expression of the guide RNA in the cell or embryo of interest. For example, DNA encoding the guide RNA can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6, U3, H1, and 7SL RNA promoters.

***(IV) Method for Modifying a Chromosomal Sequence Using an RNA-Guided Endonuclease***

[0063] Another aspect of the present disclosure encompasses a method for modifying a chromosomal sequence in a eukaryotic cell or embryo. The method comprises introducing into a eukaryotic cell or embryo (i) at least one RNA-guided endonuclease comprising at least one nuclear localization signal or nucleic acid encoding at least one RNA-guided endonuclease comprising at least one nuclear localization signal, (ii) at least one guide RNA or DNA encoding at least one guide RNA, and, optionally, (iii) at least one donor polynucleotide comprising a donor sequence. The method further comprises culturing the cell or embryo such that each guide RNA directs an RNA-guided endonuclease to a targeted site in the chromosomal sequence where the RNA-guided endonuclease introduces a double-stranded break in the targeted site, and the double-stranded break is repaired by a DNA repair process such that the chromosomal sequence is modified.

[0064] In some embodiments, the method can comprise introducing one RNA-guided endonuclease (or encoding nucleic acid) and one guide RNA (or encoding DNA) into a cell or embryo, wherein the RNA-guided endonuclease introduces one double-stranded break in the targeted chromosomal sequence. In embodiments in which the optional donor polynucleotide is not present, the double-stranded break in the chromosomal sequence can be repaired by a non-homologous end-joining (NHEJ) repair process. Because NHEJ is error-prone, deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. Accordingly, the targeted chromosomal sequence can be modified or inactivated. For example, a single nucleotide change (SNP) can give rise to an altered protein product, or a shift in the reading frame of a coding sequence can inactivate or “knock out” the sequence such that no protein product is made. In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence at the targeted site during repair of the double-stranded break. For example, in embodiments in which the donor

sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted site in the chromosomal sequence, the donor sequence can be exchanged with or integrated into the chromosomal sequence at the targeted site during repair mediated by homology-directed repair process. Alternatively, in embodiments in which the donor sequence is flanked by compatible overhangs (or the compatible overhangs are generated *in situ* by the RNA-guided endonuclease) the donor sequence can be ligated directly with the cleaved chromosomal sequence by a non-homologous repair process during repair of the double-stranded break. Exchange or integration of the donor sequence into the chromosomal sequence modifies the targeted chromosomal sequence or introduces an exogenous sequence into the chromosomal sequence of the cell or embryo.

[0065] In other embodiments, the method can comprise introducing two RNA-guided endonucleases (or encoding nucleic acid) and two guide RNAs (or encoding DNA) into a cell or embryo, wherein the RNA-guided endonucleases introduce two double-stranded breaks in the chromosomal sequence. See **FIG. 3B**. The two breaks can be within several base pairs, within tens of base pairs, or can be separated by many thousands of base pairs. In embodiments in which the optional donor polynucleotide is not present, the resultant double-stranded breaks can be repaired by a non-homologous repair process such that the sequence between the two cleavage sites is lost and/or deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break(s). In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence during repair of the double-stranded breaks by either a homology-based repair process (e.g., in embodiments in which the donor sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted sites in the chromosomal sequence) or a non-homologous repair process (e.g., in embodiments in which the donor sequence is flanked by compatible overhangs).

[0066] In still other embodiments, the method can comprise introducing one RNA-guided endonuclease modified to cleave one strand of a double-stranded sequence (or encoding nucleic acid) and two guide RNAs (or encoding DNA) into a cell or embryo, wherein each guide RNA directs the RNA-guided endonuclease to a specific target site, at which site the modified endonuclease cleaves one strand (i.e., nicks) of the double-stranded chromosomal sequence, and wherein the two nicks are in opposite stands and in close enough proximity to constitute a double-stranded break. See **FIG. 3A**. In embodiments in which the optional donor polynucleotide is not present, the resultant double-stranded break can be repaired by a non-homologous repair process such that deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence during repair of the double-stranded break by either a homology-based repair process (e.g., in embodiments in which the donor sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted sites in the chromosomal sequence) or a non-homologous repair process (e.g., in embodiments in which the donor sequence is flanked by compatible overhangs).

**(a) RNA-guided endonuclease**

[0067] The method comprises introducing into a cell or embryo at least one RNA-guided endonuclease comprising at least one nuclear localization signal or nucleic acid encoding at least one RNA-guided endonuclease comprising at least one nuclear localization signal. Such RNA-guided endonucleases and nucleic acids encoding RNA-guided endonucleases are described above in sections (I) and (III), respectively.

[0068] In some embodiments, the RNA-guided endonuclease can be introduced into the cell or embryo as an isolated protein. In such embodiments, the RNA-guided endonuclease can further comprise at least one cell-penetrating domain, which facilitates cellular uptake of the protein. In other embodiments, the RNA-guided

endonuclease can be introduced into the cell or embryo as an mRNA molecule. In still other embodiments, the RNA-guided endonuclease can be introduced into the cell or embryo as a DNA molecule. In general, DNA sequence encoding the fusion protein is operably linked to a promoter sequence that will function in the cell or embryo of interest. The DNA sequence can be linear, or the DNA sequence can be part of a vector. In still other embodiments, the fusion protein can be introduced into the cell or embryo as an RNA-protein complex comprising the fusion protein and the guide RNA.

[0069] In alternate embodiments, DNA encoding the RNA-guided endonuclease can further comprise sequence encoding a guide RNA. In general, each of the sequences encoding the RNA-guided endonuclease and the guide RNA is operably linked to appropriate promoter control sequence that allows expression of the RNA-guided endonuclease and the guide RNA, respectively, in the cell or embryo. The DNA sequence encoding the RNA-guided endonuclease and the guide RNA can further comprise additional expression control, regulatory, and/or processing sequence(s). The DNA sequence encoding the RNA-guided endonuclease and the guide RNA can be linear or can be part of a vector

**(b) Guide RNA**

[0070] The method also comprises introducing into a cell or embryo at least one guide RNA or DNA encoding at least one guide RNA. A guide RNA interacts with the RNA-guided endonuclease to direct the endonuclease to a specific target site, at which site the 5' end of the guide RNA base pairs with a specific protospacer sequence in the chromosomal sequence.

[0071] Each guide RNA comprises three regions: a first region at the 5' end that is complementary to the target site in the chromosomal sequence, a second internal region that forms a stem loop structure, and a third 3' region that remains essentially single-stranded. The first region of each guide RNA is different such that each guide RNA guides a fusion protein to a specific target site. The second and third regions of each guide RNA can be the same in all guide RNAs.

[0072] The first region of the guide RNA is complementary to sequence (i.e., protospacer sequence) at the target site in the chromosomal sequence such that

the first region of the guide RNA can base pair with the target site. In various embodiments, the first region of the guide RNA can comprise from about 10 nucleotides to more than about 25 nucleotides. For example, the region of base pairing between the first region of the guide RNA and the target site in the chromosomal sequence can be about 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, or more than 25 nucleotides in length. In an exemplary embodiment, the first region of the guide RNA is about 19, 20, or 21 nucleotides in length.

[0073] The guide RNA also comprises a second region that forms a secondary structure. In some embodiments, the secondary structure comprises a stem (or hairpin) and a loop. The length of the loop and the stem can vary. For example, the loop can range from about 3 to about 10 nucleotides in length, and the stem can range from about 6 to about 20 base pairs in length. The stem can comprise one or more bulges of 1 to about 10 nucleotides. Thus, the overall length of the second region can range from about 16 to about 60 nucleotides in length. In an exemplary embodiment, the loop is about 4 nucleotides in length and the stem comprises about 12 base pairs.

[0074] The guide RNA also comprises a third region at the 3' end that remains essentially single-stranded. Thus, the third region has no complementarity to any chromosomal sequence in the cell of interest and has no complementarity to the rest of the guide RNA. The length of the third region can vary. In general, the third region is more than about 4 nucleotides in length. For example, the length of the third region can range from about 5 to about 60 nucleotides in length.

[0075] The combined length of the second and third regions (also called the universal or scaffold region) of the guide RNA can range from about 30 to about 120 nucleotides in length. In one aspect, the combined length of the second and third regions of the guide RNA range from about 70 to about 100 nucleotides in length.

[0076] In some embodiments, the guide RNA comprises a single molecule comprising all three regions. In other embodiments, the guide RNA can comprise two separate molecules. The first RNA molecule can comprise the first region of the guide RNA and one half of the "stem" of the second region of the guide RNA. The second RNA molecule can comprise the other half of the "stem" of the second region of the guide RNA and the third region of the guide RNA. Thus, in this embodiment, the first

and second RNA molecules each contain a sequence of nucleotides that are complementary to one another. For example, in one embodiment, the first and second RNA molecules each comprise a sequence (of about 6 to about 20 nucleotides) that base pairs to the other sequence to form a functional guide RNA.

[0077] In some embodiments, the guide RNA can be introduced into the cell or embryo as a RNA molecule. The RNA molecule can be transcribed *in vitro*. Alternatively, the RNA molecule can be chemically synthesized.

[0078] In other embodiments, the guide RNA can be introduced into the cell or embryo as a DNA molecule. In such cases, the DNA encoding the guide RNA can be operably linked to promoter control sequence for expression of the guide RNA in the cell or embryo of interest. For example, the RNA coding sequence can be operably linked to a promoter sequence that is recognized by RNA polymerase III (Pol III). Examples of suitable Pol III promoters include, but are not limited to, mammalian U6 or H1 promoters. In exemplary embodiments, the RNA coding sequence is linked to a mouse or human U6 promoter. In other exemplary embodiments, the RNA coding sequence is linked to a mouse or human H1 promoter.

[0079] The DNA molecule encoding the guide RNA can be linear or circular. In some embodiments, the DNA sequence encoding the guide RNA can be part of a vector. Suitable vectors include plasmid vectors, phagemids, cosmids, artificial/mini-chromosomes, transposons, and viral vectors. In an exemplary embodiment, the DNA encoding the RNA-guided endonuclease is present in a plasmid vector. Non-limiting examples of suitable plasmid vectors include pUC, pBR322, pET, pBluescript, and variants thereof. The vector can comprise additional expression control sequences (e.g., enhancer sequences, Kozak sequences, polyadenylation sequences, transcriptional termination sequences, etc.), selectable marker sequences (e.g., antibiotic resistance genes), origins of replication, and the like.

[0080] In embodiments in which both the RNA-guided endonuclease and the guide RNA are introduced into the cell as DNA molecules, each can be part of a separate molecule (e.g., one vector containing fusion protein coding sequence and a second vector containing guide RNA coding sequence) or both can be part of the same

molecule (e.g., one vector containing coding (and regulatory) sequence for both the fusion protein and the guide RNA).

**(c) Target site**

[0081] An RNA-guided endonuclease in conjunction with a guide RNA is directed to a target site in the chromosomal sequence, wherein the RNA-guided endonuclease introduces a double-stranded break in the chromosomal sequence. The target site has no sequence limitation except that the sequence is immediately followed (downstream) by a consensus sequence. This consensus sequence is also known as a protospacer adjacent motif (PAM). Examples of PAM include, but are not limited to, NGG, NGGNG, and NNAGAAW (wherein N is defined as any nucleotide and W is defined as either A or T). As detailed above in section (IV)(b), the first region (at the 5' end) of the guide RNA is complementary to the protospacer of the target sequence. Typically, the first region of the guide RNA is about 19 to 21 nucleotides in length. Thus, in certain aspects, the sequence of the target site in the chromosomal sequence is 5'-$N_{19-21}$-*NGG*-3'. The PAM is in italics.

[0082] The target site can be in the coding region of a gene, in an intron of a gene, in a control region of a gene, in a non-coding region between genes, etc. The gene can be a protein coding gene or an RNA coding gene. The gene can be any gene of interest.

**(d) Optional donor polynucleotide**

[0083] In some embodiments, the method further comprises introducing at least one donor polynucleotide into the embryo. A donor polynucleotide comprises at least one donor sequence. In some aspects, a donor sequence of the donor polynucleotide corresponds to an endogenous or native chromosomal sequence. For example, the donor sequence can be essentially identical to a portion of the chromosomal sequence at or near the targeted site, but which comprises at least one nucleotide change. Thus, the donor sequence can comprise a modified version of the wild type sequence at the targeted site such that, upon integration or exchange with the native sequence, the sequence at the targeted chromosomal location comprises at least

one nucleotide change. For example, the change can be an insertion of one or more nucleotides, a deletion of one or more nucleotides, a substitution of one or more nucleotides, or combinations thereof. As a consequence of the integration of the modified sequence, the cell or embryo/animal can produce a modified gene product from the targeted chromosomal sequence.

[0084] In other aspects, the donor sequence of the donor polynucleotide corresponds to an exogenous sequence. As used herein, an "exogenous" sequence refers to a sequence that is not native to the cell or embryo, or a sequence whose native location in the genome of the cell or embryo is in a different location. For example, the exogenous sequence can comprise protein coding sequence, which can be operably linked to an exogenous promoter control sequence such that, upon integration into the genome, the cell or embryo/animal is able to express the protein coded by the integrated sequence. Alternatively, the exogenous sequence can be integrated into the chromosomal sequence such that its expression is regulated by an endogenous promoter control sequence. In other iterations, the exogenous sequence can be a transcriptional control sequence, another expression control sequence, an RNA coding sequence, and so forth. Integration of an exogenous sequence into a chromosomal sequence is termed a "knock in."

[0085] As can be appreciated by those skilled in the art, the length of the donor sequence can and will vary. For example, the donor sequence can vary in length from several nucleotides to hundreds of nucleotides to hundreds of thousands of nucleotides.

[0086] Donor polynucleotide comprising upstream and downstream sequences. In some embodiments, the donor sequence in the donor polynucleotide is flanked by an upstream sequence and a downstream sequence, which have substantial sequence identity to sequences located upstream and downstream, respectively, of the targeted site in the chromosomal sequence. Because of these sequence similarities, the upstream and downstream sequences of the donor polynucleotide permit homologous recombination between the donor polynucleotide and the targeted chromosomal sequence such that the donor sequence can be integrated into (or exchanged with) the chromosomal sequence.

[0087] The upstream sequence, as used herein, refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence upstream of the targeted site. Similarly, the downstream sequence refers to a nucleic acid sequence that shares substantial sequence identity with a chromosomal sequence downstream of the targeted site. As used herein, the phrase "substantial sequence identity" refers to sequences having at least about 75% sequence identity. Thus, the upstream and downstream sequences in the donor polynucleotide can have about 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with sequence upstream or downstream to the targeted site. In an exemplary embodiment, the upstream and downstream sequences in the donor polynucleotide can have about 95% or 100% sequence identity with chromosomal sequences upstream or downstream to the targeted site. In one embodiment, the upstream sequence shares substantial sequence identity with a chromosomal sequence located immediately upstream of the targeted site (i.e., adjacent to the targeted site). In other embodiments, the upstream sequence shares substantial sequence identity with a chromosomal sequence that is located within about one hundred (100) nucleotides upstream from the targeted site. Thus, for example, the upstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides upstream from the targeted site. In one embodiment, the downstream sequence shares substantial sequence identity with a chromosomal sequence located immediately downstream of the targeted site (i.e., adjacent to the targeted site). In other embodiments, the downstream sequence shares substantial sequence identity with a chromosomal sequence that is located within about one hundred (100) nucleotides downstream from the targeted site. Thus, for example, the downstream sequence can share substantial sequence identity with a chromosomal sequence that is located about 1 to about 20, about 21 to about 40, about 41 to about 60, about 61 to about 80, or about 81 to about 100 nucleotides downstream from the targeted site.

[0088] Each upstream or downstream sequence can range in length from about 20 nucleotides to about 5000 nucleotides. In some embodiments, upstream and

downstream sequences can comprise about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, or 5000 nucleotides. In exemplary embodiments, upstream and downstream sequences can range in length from about 50 to about 1500 nucleotides.

[0089] Donor polynucleotides comprising the upstream and downstream sequences with sequence similarity to the targeted chromosomal sequence can be linear or circular. In embodiments in which the donor polynucleotide is circular, it can be part of a vector. For example, the vector can be a plasmid vector.

[0090] Donor polynucleotide comprising targeted cleavage site(s). In other embodiments, the donor polynucleotide can additionally comprise at least one targeted cleavage site that is recognized by the RNA-guided endonuclease. The targeted cleavage site added to the donor polynucleotide can be placed upstream or downstream or both upstream and downstream of the donor sequence. For example, the donor sequence can be flanked by targeted cleavage sites such that, upon cleavage by the RNA-guided endonuclease, the donor sequence is flanked by overhangs that are compatible with those in the chromosomal sequence generated upon cleavage by the RNA-guided endonuclease. Accordingly, the donor sequence can be ligated with the cleaved chromosomal sequence during repair of the double stranded break by a non-homologous repair process. Generally, donor polynucleotides comprising the targeted cleavage site(s) will be circular (e.g., can be part of a plasmid vector).

[0091] Donor polynucleotide comprising a short donor sequence with optional overhangs. In still alternate embodiments, the donor polynucleotide can be a linear molecule comprising a short donor sequence with optional short overhangs that are compatible with the overhangs generated by the RNA-guided endonuclease. In such embodiments, the donor sequence can be ligated directly with the cleaved chromosomal sequence during repair of the double-stranded break. In some instances, the donor sequence can be less than about 1,000, less than about 500, less than about 250, or less than about 100 nucleotides. In certain cases, the donor polynucleotide can be a linear molecule comprising a short donor sequence with blunt ends. In other iterations, the donor polynucleotide can be a linear molecule comprising a short donor

sequence with 5' and/or 3' overhangs. The overhangs can comprise 1, 2, 3, 4, or 5 nucleotides.

[0092] Typically, the donor polynucleotide will be DNA. The DNA may be single-stranded or double-stranded and/or linear or circular. The donor polynucleotide may be a DNA plasmid, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), a viral vector, a linear piece of DNA, a PCR fragment, a naked nucleic acid, or a nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer. In certain embodiments, the donor polynucleotide comprising the donor sequence can be part of a plasmid vector. In any of these situations, the donor polynucleotide comprising the donor sequence can further comprise at least one additional sequence.

**(e) Introducing into the cell or embryo**

[0093] The RNA-targeted endonuclease(s) (or encoding nucleic acid), the guide RNA(s) (or encoding DNA), and the optional donor polynucleotide(s) can be introduced into a cell or embryo by a variety of means. In some embodiments, the cell or embryo is transfected. Suitable transfection methods include calcium phosphate-mediated transfection, nucleofection (or electroporation), cationic polymer transfection (e.g., DEAE-dextran or polyethylenimine), viral transduction, virosome transfection, virion transfection, liposome transfection, cationic liposome transfection, immunoliposome transfection, nonliposomal lipid transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, gene gun delivery, impalefection, sonoporation, optical transfection, and proprietary agent-enhanced uptake of nucleic acids. Transfection methods are well known in the art (*see,* e.g., "Current Protocols in Molecular Biology" Ausubel et al., John Wiley & Sons, New York, 2003 or "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001). In other embodiments, the molecules are introduced into the cell or embryo by microinjection. Typically, the embryo is a fertilized one-cell stage embryo of the species of interest. For example, the molecules can be injected into the pronuclei of one cell embryos.

[0094] The RNA-targeted endonuclease(s) (or encoding nucleic acid), the guide RNA(s) (or DNAs encoding the guide RNA), and the optional donor polynucleotide(s) can be introduced into the cell or embryo simultaneously or sequentially. The ratio of the RNA-targeted endonuclease(s) (or encoding nucleic acid) to the guide RNA(s) (or encoding DNA) generally will be about stoichiometric such that they can form an RNA-protein complex. In one embodiment, DNA encoding an RNA-targeted endonuclease and DNA encoding a guide RNA are delivered together within the plasmid vector.

**(f) Culturing the cell or embryo**

[0095] The method further comprises maintaining the cell or embryo under appropriate conditions such that the guide RNA(s) directs the RNA-guided endonuclease(s) to the targeted site(s) in the chromosomal sequence, and the RNA-guided endonuclease(s) introduce at least one double-stranded break in the chromosomal sequence. A double-stranded break can be repaired by a DNA repair process such that the chromosomal sequence is modified by a deletion of at least one nucleotide, an insertion of at least one nucleotide, a substitution of at least one nucleotide, or a combination thereof.

[0096] In embodiments in which no donor polynucleotide is introduced into the cell or embryo, the double-stranded break can be repaired via a non-homologous end-joining (NHEJ) repair process. Because NHEJ is error-prone, deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. Accordingly, the sequence at the chromosomal sequence can be modified such that the reading frame of a coding region can be shifted and that the chromosomal sequence is inactivated or "knocked out." An inactivated protein-coding chromosomal sequence does not give rise to the protein coded by the wild type chromosomal sequence.

[0097] In embodiments in which a donor polynucleotide comprising upstream and downstream sequences is introduced into the cell or embryo, the double-stranded break can be repaired by a homology-directed repair (HDR) process such that the donor sequence is integrated into the chromosomal sequence. Accordingly, an

exogenous sequence can be integrated into the genome of the cell or embryo, or the targeted chromosomal sequence can be modified by exchange of a modified sequence for the wild type chromosomal sequence.

[0098] In embodiments in which a donor polynucleotide comprising the targeted cleave site is introduced into the cell or embryo, the RNA-guided endonuclease can cleave both the targeted chromosomal sequence and the donor polynucleotide. The linearized donor polynucleotide can be integrated into the chromosomal sequence at the site of the double-stranded break by ligation between the donor polynucleotide and the cleaved chromosomal sequence via a NHEJ process.

[0099] In embodiments in which a linear donor polynucleotide comprising a short donor sequence is introduced into the cell or embryo, the short donor sequence can be integrated into the chromosomal sequence at the site of the double-stranded break via a NHEJ process. The integration can proceed via the ligation of blunt ends between the short donor sequence and the chromosomal sequence at the site of the double stranded break. Alternatively, the integration can proceed via the ligation of sticky ends (i.e., having 5' or 3' overhangs) between a short donor sequence that is flanked by overhangs that are compatible with those generated by the RNA-targeting endonuclease in the cleaved chromosomal sequence.

[0100] In general, the cell is maintained under conditions appropriate for cell growth and/or maintenance. Suitable cell culture conditions are well known in the art and are described, for example, in Santiago et al. (2008) PNAS 105:5809-5814; Moehle et al. (2007) PNAS 104:3055-3060; Urnov et al. (2005) Nature 435:646-651; and Lombardo et al (2007) Nat. Biotechnology 25:1298-1306. Those of skill in the art appreciate that methods for culturing cells are known in the art and can and will vary depending on the cell type. Routine optimization may be used, in all cases, to determine the best techniques for a particular cell type.

[0101] An embryo can be cultured *in vitro* (e.g., in cell culture). Typically, the embryo is cultured at an appropriate temperature and in appropriate media with the necessary $O_2/CO_2$ ratio to allow the expression of the RNA endonuclease and guide RNA, if necessary. Suitable non-limiting examples of media include M2, M16, KSOM, BMOC, and HTF media. A skilled artisan will appreciate that culture

conditions can and will vary depending on the species of embryo. Routine optimization may be used, in all cases, to determine the best culture conditions for a particular species of embryo. In some cases, a cell line may be derived from an in vitro-cultured embryo (e.g., an embryonic stem cell line).

[0102] Alternatively, an embryo may be cultured *in vivo* by transferring the embryo into the uterus of a female host. Generally speaking the female host is from the same or similar species as the embryo. Preferably, the female host is pseudo-pregnant. Methods of preparing pseudo-pregnant female hosts are known in the art. Additionally, methods of transferring an embryo into a female host are known. Culturing an embryo *in vivo* permits the embryo to develop and can result in a live birth of an animal derived from the embryo. Such an animal would comprise the modified chromosomal sequence in every cell of the body.

**(g) Cell and embryo types**

[0103] A variety of eukaryotic cells and embryos are suitable for use in the method. For example, the cell can be a human cell, a non-human mammalian cell, a non-mammalian vertebrate cell, an invertebrate cell, an insect cell, a plant cell, a yeast cell, or a single cell eukaryotic organism. In general, the embryo is non-human mammalian embryo. In specific embodiments, the embryos can be a one cell non-human mammalian embryo. Exemplary mammalian embryos, including one cell embryos, include without limit mouse, rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos. In still other embodiments, the cell can be a stem cell. Suitable stem cells include without limit embryonic stem cells, ES-like stem cells, fetal stem cells, adult stem cells, pluripotent stem cells, induced pluripotent stem cells, multipotent stem cells, oligopotent stem cells, unipotent stem cells and others. In exemplary embodiments, the cell is a mammalian cell.

[0104] Non-limiting examples of suitable mammalian cells include Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells; mouse myeloma NS0 cells, mouse embryonic fibroblast 3T3 cells (NIH3T3), mouse B lymphoma A20 cells; mouse melanoma B16 cells; mouse myoblast C2C12 cells; mouse myeloma SP2/0 cells; mouse embryonic mesenchymal C3H-10T1/2 cells; mouse carcinoma CT26 cells,

mouse prostate DuCuP cells; mouse breast EMT6 cells; mouse hepatoma Hepa1c1c7 cells; mouse myeloma J5582 cells; mouse epithelial MTD-1A cells; mouse myocardial MyEnd cells; mouse renal RenCa cells; mouse pancreatic RIN-5F cells; mouse melanoma X64 cells; mouse lymphoma YAC-1 cells; rat glioblastoma 9L cells; rat B lymphoma RBL cells; rat neuroblastoma B35 cells; rat hepatoma cells (HTC); buffalo rat liver BRL 3A cells; canine kidney cells (MDCK); canine mammary (CMT) cells; rat osteosarcoma D17 cells; rat monocyte/macrophage DH82 cells; monkey kidney SV-40 transformed fibroblast (COS7) cells; monkey kidney CVI-76 cells; African green monkey kidney (VERO-76) cells; human embryonic kidney cells (HEK293, HEK293T); human cervical carcinoma cells (HELA); human lung cells (W138); human liver cells (Hep G2); human U2-OS osteosarcoma cells, human A549 cells, human A-431 cells, and human K562 cells. An extensive list of mammalian cell lines may be found in the American Type Culture Collection catalog (ATCC, Mamassas, VA).

### *(V) Method for Using a Fusion Protein to Modify a Chromosomal Sequence or Regulate Expression of a Chromosomal Sequence*

[0105] Another aspect of the present disclosure encompasses a method for modifying a chromosomal sequence or regulating expression of a chromosomal sequence in a cell or embryo. The method comprises introducing into the cell or embryo (a) at least one fusion protein or nucleic acid encoding at least one fusion protein, wherein the fusion protein comprises a CRISPR/Cas-like protein or a fragment thereof and an effector domain, and (b) at least one guide RNA or DNA encoding the guide RNA, wherein the guide RNA guides the CRISPR/Cas-like protein of the fusion protein to a targeted site in the chromosomal sequence and the effector domain of the fusion protein modifies the chromosomal sequence or regulates expression of the chromosomal sequence.

[0106] Fusion proteins comprising a CRISPR/Cas-like protein or a fragment thereof and an effector domain are detailed above in section (II). In general, the fusion proteins disclosed herein further comprise at least one nuclear localization signal. Nucleic acids encoding fusion proteins are described above in section (III). In some embodiments, the fusion protein can be introduced into the cell or embryo as an

isolated protein (which can further comprise a cell-penetrating domain). Furthermore, the isolated fusion protein can be part of a protein-RNA complex comprising the guide RNA. In other embodiments, the fusion protein can be introduced into the cell or embryo as a RNA molecule (which can be capped and/or polyadenylated). In still other embodiments, the fusion protein can be introduced into the cell or embryo as a DNA molecule. For example, the fusion protein and the guide RNA can be introduced into the cell or embryo as discrete DNA molecules or as part of the same DNA molecule. Such DNA molecules can be plasmid vectors.

[0107] In some embodiments, the method further comprises introducing into the cell or embryo at least one zinc finger nuclease. Zinc finger nucleases are described above in section (II)(d). In still other embodiments, the method further comprises introducing into the cell or embryo at least one donor polynucleotide. Donor polynucleotides are detailed above in section (IV)(d). Means for introducing molecules into cells or embryos, as well as means for culturing cell or embryos are described above in sections (IV)(e) and (IV)(f), respectively. Suitable cells and embryos are described above in section (IV)(g).

[0108] In certain embodiments in which the effector domain of the fusion protein is a cleavage domain (e.g., a FokI cleavage domain or a modified FokI cleavage domain), the method can comprise introducing into the cell or embryo one fusion protein (or nucleic acid encoding one fusion protein) and two guide RNAs (or DNA encoding two guide RNAs). The two guide RNAs direct the fusion protein to two different target sites in the chromosomal sequence, wherein the fusion protein dimerizes (e.g., form a homodimer) such that the two cleavage domains can introduce a double stranded break into the chromosomal sequence. See **FIG. 1A**. In embodiments in which the optional donor polynucleotide is not present, the double-stranded break in the chromosomal sequence can be repaired by a non-homologous end-joining (NHEJ) repair process. Because NHEJ is error-prone, deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. Accordingly, the targeted chromosomal sequence can be modified or inactivated. For example, a single nucleotide change (SNP) can give rise to an altered protein product, or a shift in the reading frame of a coding

sequence can inactivate or "knock out" the sequence such that no protein product is made. In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence at the targeted site during repair of the double-stranded break. For example, in embodiments in which the donor sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted site in the chromosomal sequence, the donor sequence can be exchanged with or integrated into the chromosomal sequence at the targeted site during repair mediated by homology-directed repair process. Alternatively, in embodiments in which the donor sequence is flanked by compatible overhangs (or the compatible overhangs are generated *in situ* by the RNA-guided endonuclease) the donor sequence can be ligated directly with the cleaved chromosomal sequence by a non-homologous repair process during repair of the double-stranded break. Exchange or integration of the donor sequence into the chromosomal sequence modifies the targeted chromosomal sequence or introduces an exogenous sequence into the chromosomal sequence of the cell or embryo.

[0109] In other embodiments in which the effector domain of the fusion protein is a cleavage domain (e.g., a FokI cleavage domain or a modified FokI cleavage domain), the method can comprise introducing into the cell or embryo two different fusion proteins (or nucleic acid encoding two different fusion proteins) and two guide RNAs (or DNA encoding two guide RNAs). The fusion proteins can differ as detailed above in section (II). Each guide RNA directs a fusion protein to a specific target site in the chromosomal sequence, wherein the fusion proteins dimerize (e.g., form a heterodimer) such that the two cleavage domains can introduce a double stranded break into the chromosomal sequence. In embodiments in which the optional donor polynucleotide is not present, the resultant double-stranded breaks can be repaired by a non-homologous repair process such that deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence during repair of the

double-stranded break by either a homology-based repair process (e.g., in embodiments in which the donor sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted sites in the chromosomal sequence) or a non-homologous repair process (e.g., in embodiments in which the donor sequence is flanked by compatible overhangs).

[0110] In still other embodiments in which the effector domain of the fusion protein is a cleavage domain (e.g., a FokI cleavage domain or a modified FokI cleavage domain), the method can comprise introducing into the cell or embryo one fusion protein (or nucleic acid encoding one fusion protein), one guide RNA (or DNA encoding one guide RNA), and one zinc finger nuclease (or nucleic acid encoding the zinc finger nuclease), wherein the zinc finger nuclease comprises a FokI cleavage domain or a modified FokI cleavage domain. The guide RNA directs the fusion protein to a specific chromosomal sequence, and the zinc finger nuclease is directed to another chromosomal sequence, wherein the fusion protein and the zinc finger nuclease dimerize such that the cleavage domain of the fusion protein and the cleavage domain of the zinc finger nuclease can introduce a double stranded break into the chromosomal sequence. See FIG. **1B**. In embodiments in which the optional donor polynucleotide is not present, the resultant double-stranded breaks can be repaired by a non-homologous repair process such that deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. In embodiments in which the optional donor polynucleotide is present, the donor sequence in the donor polynucleotide can be exchanged with or integrated into the chromosomal sequence during repair of the double-stranded break by either a homology-based repair process (e.g., in embodiments in which the donor sequence is flanked by upstream and downstream sequences having substantial sequence identity with upstream and downstream sequences, respectively, of the targeted sites in the chromosomal sequence) or a non-homologous repair process (e.g., in embodiments in which the donor sequence is flanked by compatible overhangs).

[0111] In still other embodiments in which the effector domain of the fusion protein is a transcriptional activation domain or a transcriptional repressor domain, the method can comprise introducing into the cell or embryo one fusion protein (or nucleic acid encoding one fusion protein) and one guide RNA (or DNA encoding one guide RNA). The guide RNA directs the fusion protein to a specific chromosomal sequence, wherein the transcriptional activation domain or a transcriptional repressor domain activates or represses expression, respectively, of the targeted chromosomal sequence. See **FIG. 2A**.

[0112] In alternate embodiments in which the effector domain of the fusion protein is an epigenetic modification domain, the method can comprise introducing into the cell or embryo one fusion protein (or nucleic acid encoding one fusion protein) and one guide RNA (or DNA encoding one guide RNA). The guide RNA directs the fusion protein to a specific chromosomal sequence, wherein the epigenetic modification domain modifies the structure of the targeted the chromosomal sequence. See **FIG. 2A**. Epigenetic modifications include acetylation, methylation of histone proteins and/or nucleotide methylation. In some instances, structural modification of the chromosomal sequence leads to changes in expression of the chromosomal sequence.

### *(VI) Genetically Modified Cells and Animals*

[0113] The present disclosure encompasses genetically modified cells, non-human embryos, and non-human animals comprising at least one chromosomal sequence that has been modified using an RNA-guided endonuclease-mediated or fusion protein-mediated process, for example, using the methods described herein. The disclosure provides cells comprising at least one DNA or RNA molecule encoding an RNA-guided endonuclease or fusion protein targeted to a chromosomal sequence of interest or a fusion protein, at least one guide RNA, and optionally one or more donor polynucleotide(s). The disclosure also provides non-human embryos comprising at least one DNA or RNA molecule encoding an RNA-guided endonuclease or fusion protein targeted to a chromosomal sequence of interest, at least one guide RNA, and optionally one or more donor polynucleotide(s).

[0114] The present disclosure provides genetically modified non-human animals, non-human embryos, or animal cells comprising at least one modified chromosomal sequence. The modified chromosomal sequence may be modified such that it is (1) inactivated, (2) has an altered expression or produces an altered protein product, or (3) comprises an integrated sequence. The chromosomal sequence is modified with an RNA guided endonuclease-mediated or fusion protein-mediated process, using the methods described herein.

[0115] As discussed, one aspect of the present disclosure provides a genetically modified animal in which at least one chromosomal sequence has been modified. In one embodiment, the genetically modified animal comprises at least one inactivated chromosomal sequence. The modified chromosomal sequence may be inactivated such that the sequence is not transcribed and/or a functional protein product is not produced. Thus, a genetically modified animal comprising an inactivated chromosomal sequence may be termed a "knock out" or a "conditional knock out." The inactivated chromosomal sequence can include a deletion mutation (i.e., deletion of one or more nucleotides), an insertion mutation (i.e., insertion of one or more nucleotides), or a nonsense mutation (i.e., substitution of a single nucleotide for another nucleotide such that a stop codon is introduced). As a consequence of the mutation, the targeted chromosomal sequence is inactivated and a functional protein is not produced. The inactivated chromosomal sequence comprises no exogenously introduced sequence. Also included herein are genetically modified animals in which two, three, four, five, six, seven, eight, nine, or ten or more chromosomal sequences are inactivated.

[0116] In another embodiment, the modified chromosomal sequence can be altered such that it codes for a variant protein product. For example, a genetically modified animal comprising a modified chromosomal sequence can comprise a targeted point mutation(s) or other modification such that an altered protein product is produced. In one embodiment, the chromosomal sequence can be modified such that at least one nucleotide is changed and the expressed protein comprises one changed amino acid residue (missense mutation). In another embodiment, the chromosomal sequence can be modified to comprise more than one missense

mutation such that more than one amino acid is changed. Additionally, the chromosomal sequence can be modified to have a three nucleotide deletion or insertion such that the expressed protein comprises a single amino acid deletion or insertion. The altered or variant protein can have altered properties or activities compared to the wild type protein, such as altered substrate specificity, altered enzyme activity, altered kinetic rates, etc.

[0117] In another embodiment, the genetically modified animal can comprise at least one chromosomally integrated sequence. A genetically modified animal comprising an integrated sequence may be termed a "knock in" or a "conditional knock in." The chromosomally integrated sequence can, for example, encode an orthologous protein, an endogenous protein, or combinations of both. In one embodiment, a sequence encoding an orthologous protein or an endogenous protein can be integrated into a chromosomal sequence encoding a protein such that the chromosomal sequence is inactivated, but the exogenous sequence is expressed. In such a case, the sequence encoding the orthologous protein or endogenous protein may be operably linked to a promoter control sequence. Alternatively, a sequence encoding an orthologous protein or an endogenous protein may be integrated into a chromosomal sequence without affecting expression of a chromosomal sequence. For example, a sequence encoding a protein can be integrated into a "safe harbor" locus, such as the Rosa26 locus, HPRT locus, or AAV locus. The present disclosure also encompasses genetically modified animals in which two, three, four, five, six, seven, eight, nine, or ten or more sequences, including sequences encoding protein(s), are integrated into the genome.

[0118] The chromosomally integrated sequence encoding a protein can encode the wild type form of a protein of interest or can encode a protein comprising at least one modification such that an altered version of the protein is produced. For example, a chromosomally integrated sequence encoding a protein related to a disease or disorder can comprise at least one modification such that the altered version of the protein produced causes or potentiates the associated disorder. Alternatively, the chromosomally integrated sequence encoding a protein related to

a disease or disorder can comprise at least one modification such that the altered version of the protein protects against the development of the associated disorder.

[0119] In an additional embodiment, the genetically modified animal can be a "humanized" animal comprising at least one chromosomally integrated sequence encoding a functional human protein. The functional human protein can have no corresponding ortholog in the genetically modified animal. Alternatively, the wild type animal from which the genetically modified animal is derived may comprise an ortholog corresponding to the functional human protein. In this case, the orthologous sequence in the "humanized" animal is inactivated such that no functional protein is made and the "humanized" animal comprises at least one chromosomally integrated sequence encoding the human protein.

[0120] In yet another embodiment, the genetically modified animal can comprise at least one modified chromosomal sequence encoding a protein such that the expression pattern of the protein is altered. For example, regulatory regions controlling the expression of the protein, such as a promoter or a transcription factor binding site, can be altered such that the protein is over-produced, or the tissue-specific or temporal expression of the protein is altered, or a combination thereof. Alternatively, the expression pattern of the protein can be altered using a conditional knockout system. A non-limiting example of a conditional knockout system includes a Cre-lox recombination system. A Cre-lox recombination system comprises a Cre recombinase enzyme, a site-specific DNA recombinase that can catalyze the recombination of a nucleic acid sequence between specific sites (lox sites) in a nucleic acid molecule. Methods of using this system to produce temporal and tissue specific expression are known in the art. In general, a genetically modified animal is generated with lox sites flanking a chromosomal sequence. The genetically modified animal comprising the lox-flanked chromosomal sequence can then be crossed with another genetically modified animal expressing Cre recombinase. Progeny animals comprising the lox-flanked chromosomal sequence and the Cre recombinase are then produced, and the lox-flanked chromosomal sequence is recombined, leading to deletion or inversion of the chromosomal sequence encoding the protein. Expression

of Cre recombinase can be temporally and conditionally regulated to effect temporally and conditionally regulated recombination of the chromosomal sequence.

[0121] In any of these embodiments, the genetically modified animal disclosed herein can be heterozygous for the modified chromosomal sequence. Alternatively, the genetically modified animal can be homozygous for the modified chromosomal sequence.

[0122] The genetically modified animals disclosed herein can be crossbred to create animals comprising more than one modified chromosomal sequence or to create animals that are homozygous for one or more modified chromosomal sequences. For example, two animals comprising the same modified chromosomal sequence can be crossbred to create an animal homozygous for the modified chromosomal sequence. Alternatively, animals with different modified chromosomal sequences can be crossbred to create an animal comprising both modified chromosomal sequences.

[0123] For example, a first animal comprising an inactivated chromosomal sequence gene “x” can be crossed with a second animal comprising a chromosomally integrated sequence encoding a human gene “X” protein to give rise to "humanized" gene “X” offspring comprising both the inactivated gene “x” chromosomal sequence and the chromosomally integrated human gene “X” sequence. Also, a humanized gene “X” animal can be crossed with a humanized gene “Y” animal to create humanized gene X/gene Y offspring. Those of skill in the art will appreciate that many combinations are possible.

[0124] In other embodiments, an animal comprising a modified chromosomal sequence can be crossbred to combine the modified chromosomal sequence with other genetic backgrounds. By way of non-limiting example, other genetic backgrounds may include wild-type genetic backgrounds, genetic backgrounds with deletion mutations, genetic backgrounds with another targeted integration, and genetic backgrounds with non-targeted integrations.

[0125] The term "animal," as used herein, refers to a non-human animal. The animal may be an embryo, a juvenile, or an adult. Suitable animals include

vertebrates such as mammals, birds, reptiles, amphibians, shellfish, and fish. Examples of suitable mammals include without limit rodents, companion animals, livestock, and primates. Non-limiting examples of rodents include mice, rats, hamsters, gerbils, and guinea pigs. Suitable companion animals include but are not limited to cats, dogs, rabbits, hedgehogs, and ferrets. Non-limiting examples of livestock include horses, goats, sheep, swine, cattle, llamas, and alpacas. Suitable primates include but are not limited to capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. Non-limiting examples of birds include chickens, turkeys, ducks, and geese. Alternatively, the animal may be an invertebrate such as an insect, a nematode, and the like. Non-limiting examples of insects include Drosophila and mosquitoes. An exemplary animal is a rat. Non-limiting examples of suitable rat strains include Dahl Salt-Sensitive, Fischer 344, Lewis, Long Evans Hooded, Sprague-Dawley, and Wistar. In one embodiment, the animal is not a genetically modified mouse. In each of the foregoing iterations of suitable animals for the invention, the animal does not include exogenously introduced, randomly integrated transposon sequences.

[0126] A further aspect of the present disclosure provides genetically modified cells or cell lines comprising at least one modified chromosomal sequence. The genetically modified cell or cell line can be derived from any of the genetically modified animals disclosed herein. Alternatively, the chromosomal sequence can be modified in a cell as described herein above (in the paragraphs describing chromosomal sequence modifications in animals) using the methods descried herein. The disclosure also encompasses a lysate of said cells or cell lines.

[0127] In general, the cells are eukaryotic cells. Suitable host cells include fungi or yeast, such as *Pichia, Saccharomyces,* or *Schizosaccharomyces;* insect cells, such as SF9 cells from *Spodoptera frugiperda* or S2 cells from *Drosophila melanogaster;* and animal cells, such as mouse, rat, hamster, non-human primate, or human cells. Exemplary cells are mammalian. The mammalian cells can be primary cells. In general, any primary cell that is sensitive to double strand breaks may be

used. The cells may be of a variety of cell types, e.g., fibroblast, myoblast, T or B cell, macrophage, epithelial cell, and so forth.

[0128] When mammalian cell lines are used, the cell line can be any established cell line or a primary cell line that is not yet described. The cell line can be adherent or non-adherent, or the cell line can be grown under conditions that encourage adherent, non-adherent or organotypic growth using standard techniques known to individuals skilled in the art. Non-limiting examples of suitable mammalian cells and cell lines are provided herein in section (IV)(g). In still other embodiments, the cell can be a stem cell. Non-limiting examples of suitable stem cells are provided in section (IV)(g).

[0129] The present disclosure also provides a genetically modified non-human embryo comprising at least one modified chromosomal sequence. The chromosomal sequence can be modified in an embryo as described herein above (in the paragraphs describing chromosomal sequence modifications in animals) using the methods descried herein. In one embodiment, the embryo is a non-human fertilized one-cell stage embryo of the animal species of interest. Exemplary mammalian embryos, including one cell embryos, include without limit, mouse, rat, hamster, rodent, rabbit, feline, canine, ovine, porcine, bovine, equine, and primate embryos.

**DEFINITIONS**

[0130] Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton et al., Dictionary of Microbiology and Molecular Biology (2nd ed. 1994); The Cambridge Dictionary of Science and Technology (Walker ed., 1988); The Glossary of Genetics, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); and Hale & Marham, The Harper Collins Dictionary of Biology (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

[0131] When introducing elements of the present disclosure or the preferred embodiments(s) thereof, the articles "a", "an", "the" and "said" are intended to

mean that there are one or more of the elements. The terms "comprising", "including" and "having" are intended to be inclusive and mean that there may be additional elements other than the listed elements.

[0132] As used herein, the term "endogenous sequence" refers to a chromosomal sequence that is native to the cell.

[0133] The term “exogenous,” as used herein, refers to a sequence that is not native to the cell, or a chromosomal sequence whose native location in the genome of the cell is in a different chromosomal location.

[0134] A "gene," as used herein, refers to a DNA region (including exons and introns) 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 regions.

[0135] The term "heterologous" refers to an entity that is not endogenous or native to the cell of interest. For example, a heterologous protein refers to a protein that is derived from or was originally derived from an exogenous source, such as an exogenously introduced nucleic acid sequence. In some instances, the heterologous protein is not normally produced by the cell of interest.

[0136] The terms “nucleic acid” and “polynucleotide” refer to a deoxyribonucleotide or ribonucleotide polymer, in linear or circular conformation, and in either single- or double-stranded form. For the purposes of the present disclosure, these terms are not to be construed as limiting with respect to the length of a polymer. The terms can encompass known analogs of natural nucleotides, as well as nucleotides that are modified in the base, sugar and/or phosphate moieties (e.g., phosphorothioate backbones). In general, an analog of a particular nucleotide has the same base-pairing specificity; i.e., an analog of A will base-pair with T.

[0137] The term "nucleotide" refers to deoxyribonucleotides or ribonucleotides. The nucleotides may be standard nucleotides (i.e., adenosine,

guanosine, cytidine, thymidine, and uridine) or nucleotide analogs. A nucleotide analog refers to a nucleotide having a modified purine or pyrimidine base or a modified ribose moiety. A nucleotide analog may be a naturally occurring nucleotide (e.g., inosine) or a non-naturally occurring nucleotide. Non-limiting examples of modifications on the sugar or base moieties of a nucleotide include the addition (or removal) of acetyl groups, amino groups, carboxyl groups, carboxymethyl groups, hydroxyl groups, methyl groups, phosphoryl groups, and thiol groups, as well as the substitution of the carbon and nitrogen atoms of the bases with other atoms (e.g., 7-deaza purines). Nucleotide analogs also include dideoxy nucleotides, 2'-O-methyl nucleotides, locked nucleic acids (LNA), peptide nucleic acids (PNA), and morpholinos.

[0138] The terms "polypeptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues.

[0139] Techniques for determining nucleic acid and amino acid sequence identity are known in the art. Typically, such techniques include determining the nucleotide sequence of the mRNA for a gene and/or determining the amino acid sequence encoded thereby, and comparing these sequences to a second nucleotide or amino acid sequence. Genomic sequences can also be determined and compared in this fashion. In general, identity refers to an exact nucleotide-to-nucleotide or amino acid-to-amino acid correspondence of two polynucleotides or polypeptide sequences, respectively. Two or more sequences (polynucleotide or amino acid) can be compared by determining their percent identity. The percent identity of two sequences, whether nucleic acid or amino acid sequences, is the number of exact matches between two aligned sequences divided by the length of the shorter sequences and multiplied by 100. An approximate alignment for nucleic acid sequences is provided by the local homology algorithm of Smith and Waterman, Advances in Applied Mathematics 2:482-489 (1981). This algorithm can be applied to amino acid sequences by using the scoring matrix developed by Dayhoff, Atlas of Protein Sequences and Structure, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA, and normalized by Gribskov, Nucl. Acids Res. 14(6):6745-6763 (1986). An exemplary implementation of this algorithm to determine percent identity of a sequence is provided by the Genetics Computer Group (Madison, Wis.) in the

"BestFit" utility application. Other suitable programs for calculating the percent identity or similarity between sequences are generally known in the art, for example, another alignment program is BLAST, used with default parameters. For example, BLASTN and BLASTP can be used using the following default parameters: genetic code=standard; filter=none; strand=both; cutoff=60; expect=10; Matrix=BLOSUM62; Descriptions=50 sequences; sort by=HIGH SCORE; Databases=non-redundant, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss protein+Spupdate+PIR. Details of these programs can be found on the GenBank website.

[0140] As various changes could be made in the above-described cells and methods without departing from the scope of the invention, it is intended that all matter contained in the above description and in the examples given below, shall be interpreted as illustrative and not in a limiting sense.

## EXAMPLES

[0141] The following examples illustrate certain aspects of the invention.

### *Example 1: Modification of* Cas9 *Gene for Mammalian Expression*

[0142] A *Cas9* gene from *Streptococcus pyogenes* strain MGAS15252 (Accession number YP_005388840.1) was optimized with *Homo sapiens* codon preference to enhance its translation in mammalian cells. The *Cas9* gene also was modified by adding a nuclear localization signal PKKKRKV (SEQ ID NO:1) at the C terminus for targeting the protein into the nuclei of mammalian cells. Table 1 presents the modified Cas9 amino acid sequence, with the nuclear localization sequence underlined. Table 2 presents the codon optimized, modified *Cas9* DNA sequence.

| Table 1. Modified Cas9 Amino Acid Sequence |
|---|
| MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFGSGETAE<br>ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIF<br>GNIVDEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNS<br>DVDKLFIQLVQIYNQLFEENPINASRVDAKAILSARLSKSRRLENLIAQLPGEKRNGLFG<br>NLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDA |

ILLSDILRVNSEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
GYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELH
AILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEE
VVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPA
FLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGAYHDLL
KIIKDKDFLDNEENEDILEDIVLTLTLFEDRGMIEERLKTYAHLFDDKVMKQLKRRRYTG
WGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQG
HSLHEQIANLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTQKGQKNSR
ERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYD
VDHIVPQSFIKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRK
FDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVI
TLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKV
YDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDK
GRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGF
DSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL
IIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNE
QKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDPKKKRKV
(SEQ ID NO:9)

Table 2. Optimized *Cas9* DNA Sequence (5'-3')

ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGG
CCGTGATCACCGACGACTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAA
CACCGACCGGCACAGCATCAAGAAGAACCTGATCGGCGCCCTGCTGTTCGGCTCT
GGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACC
AGACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCA
AGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGA
TAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC
CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAGCTGGCCGACAGCACCG
ACAAGGCCGACCTGAGACTGATCTACCTGGCCCTGGCCCACATGATCAAGTTCCG
GGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAA
GCTGTTCATCCAGCTGGTGCAGATCTACAATCAGCTGTTCGAGGAAAACCCCATCA
ACGCCAGCAGAGTGGACGCCAAGGCCATCCTGAGCGCCAGACTGAGCAAGAGCA
GACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGCGGAATGGCCTGT
TCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTT
CGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGA
CCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCC
GCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACAGCG
AGATCACCAAGGCCCCCCTGTCCGCCTCTATGATCAAGAGATACGACGAGCACCA
CCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTAC
AAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATCGATGGCG
GAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGA
CGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCA
GCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCA

CGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAA
AAGATCGAGAAGATCCTGACCTTCAGAATCCCCTACTACGTGGGCCCTCTGGCCA
GGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCC
CCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCCAGCGCCCAGAGCTTCATCG
AGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCA
CAGCCTGCTGTACGAGTACTTCACCGTGTACAACGAGCTGACCAAAGTGAAATAC
GTGACCGAGGGAATGCGGAAGCCCGCCTTTCTGAGCGGCGAGCAGAAAAAGGCC
ATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAG
AGGACTACTTCAAGAAAATCGAGTGCTTCGACAGCGTGGAAATCAGCGGCGTGGA
AGATCGGTTCAACGCCTCCCTGGGCGCCTATCACGATCTGCTGAAAATTATCAAGG
ACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCT
GACCCTGACACTGTTTGAGGACCGGGGCATGATCGAGGAACGGCTGAAAACCTAT
GCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACC
GGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCC
GGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCAT
GCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAG
GTGTCCGGCCAGGGACACTCTCTGCACGAGCAGATCGCCAATCTGGCCGGATCC
CCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGATTGTGGACGAGCTCGTGA
AAGTGATGGGCCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGA
GGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACACC
CAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGT
ACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCACAT
TGTGCCCCAGTCCTTCATCAAGGACGACTCCATCGATAACAAAGTGCTGACTCGGA
GCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGA
AGATGAAGAACTACTGGCGCCAGCTGCTGAATGCCAAGCTGATTACCCAGAGGAA
GTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGC
CGGCTTCATTAAGCGGCAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCA
CAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAACGACAAACTGATCC
GGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGACTTCAGAAAGGA
TTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCCACGACGCCT
ACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAG
CGAGTTCGTGTACGGCGATTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAG
AGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGCAACATCA
TGAACTTTTTCAAGACCGAGATCACACTGGCCAACGGCGAGATCAGAAAGCGGCC
TCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGA
TTTTGCCACAGTGCGGAAAGTGCTGTCCATGCCCCAAGTGAATATCGTGAAAAAGA
CCGAGGTGCAGACCGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACTC
CGACAAGCTGATCGCCAGAAAGAAGGATTGGGACCCTAAGAAGTACGGCGGCTTT
GACAGCCCCACCGTGGCCTACTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGC
AAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGA
AGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAA
ACGGCCGGAAGCGGATGCTGGCTTCTGCCGGCGAACTGCAGAAGGGAAACGAGC
TGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAG
CTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACA

| |
|---|
| AGCACTACCTGGACGAGATCATCGAGCAGATTAGCGAGTTCTCCAAGCGCGTGAT<br>CCTGGCCGATGCCAACCTGGACAAGGTGCTGAGCGCCTACAACAAGCACCGGGA<br>TAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAACC<br>TGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTA<br>CACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGG<br>CCTGTACGAGACACGGATCGACCTGTCTCAGCTGGGAGGCGACCCCAAGAAAAAG<br>CGCAAAGTG (SEQ ID NO:10) |

[0143] The modified *Cas9* DNA sequence was placed under the control of cytomegalovirus (CMV) promoter for constituent expression in mammalian cells. The modified *Cas9* DNA sequence was also placed under the control T7 promoter for *in vitro* mRNA synthesis with T7 RNA polymerase. In vitro RNA transcription was performed by using MessageMAX T7 ARCA-Capped Message Transcription Kit and T7 mScript Standard mRNA Production System (Cellscript).

***Example 2: Targeting Cas9***

[0144] The adeno-associated virus integration site 1 (AAVS1) locus was used as a target for Cas9-mediated human genome modification. The human AAVS1 locus is located in intron 1 (4427 bp) of protein phosphatase 1, regulatory subunit 12C (PPP1R12C). Table 3 presents the first exon (shaded gray) and the first intron of PPP1R12C. The underlined sequence within the intron is the targeted modification site (i.e., AAVS1 locus).

| Table 3. First Exon and Intron of PPP1R12C (5'-3') |
|---|
| GCGGGCGGGCGGTGCGATGTCCGGAGAGGATGGCCCGGCGGCTGGCCCGGGG<br>GCGGCGGCGGCGGCTGCCCGGGAGCGGCGACGGGAGCAGCTGCGGCAGTGGG<br>GGGCGCGGGCGGGCGCCGAGCCTGGCCCCGGAGAGCGCCGCGCCCGCACCGT<br>CCGCTTCGAGCGCGCCGCCGAGTTCCTGGCGGCCTGTGCGGGCGGCGACCTGG<br>ACGAGGCGCGTCTGATGCTGCGCGCCGCCGACCCTGGCCCCGGCGCCGAGCTC<br>GACCCCGCCGCGCCGCCGCCCGCCCGCGCCGTGCTGGACTCCACCAACGCCGA<br>CGGTATCAGCGCCCTGCACCAGGTCAGCGCCCCCCGCCCGGCGTCTCCCGGGGC<br>CAGGTCCACCCTCTGCTGCGCCACCTGGGGCATCCTCCTTCCCCGTTGCCAGTCT<br>CGATCCGCCCCGTCGTTCCTGGCCCTGGGCTTTGCCACCCTATGCTGACACCCCG<br>TCCCAGTCCCCCTTACCATTCCCCTTCGACCACCCCACTTCCGAATTGGAGCCGCT<br>TCAACTGGCCCTGGGCTTAGCCACTCTGTGCTGACCACTCTGCCCCAGGCCTCCT<br>TACCATTCCCCTTCGACCTACTCTCTTCCGCATTGGAGTCGCTTTAACTGGCCCTG<br>GCTTTGGCAGCCTGTGCTGACCCATGCAGTCCTCCTTACCATCCCTCCCTCGACTT |

CCCCTCTTCCGATGTTGAGCCCCTCCAGCCGGTCCTGGACTTTGTCTCCTTCCCTG
CCCTGCCCTCTCCTGAACCTGAGCCAGCTCCCATAGCTCAGTCTGGTCTATCTGCC
TGGCCCTGGCCATTGTCACTTTGCGCTGCCCTCCTCTCGCCCCCGAGTGCCCTTG
CTGTGCCGCCGGAACTCTGCCCTCTAACGCTGCCGTCTCTCTCCTGAGTCCGGAC
CACTTTGAGCTCTACTGGCTTCTGCGCCGCCTCTGGCCCACTGTTTCCCCTTCCCA
GGCAGGTCCTGCTTTCTCTGACCTGCATTCTCTCCCCTGGGCCTGTGCCGCTTTCT
GTCTGCAGCTTGTGGCCTGGGTCACCTCTACGGCTGGCCCAGATCCTTCCCTGCC
GCCTCCTTCAGGTTCCGTCTTCCTCCACTCCCTCTTCCCCTTGCTCTCTGCTGTGT
TGCTGCCCAAGGATGCTCTTTCCGGAGCACTTCCTTCTCGGCGCTGCACCACGTG
ATGTCCTCTGAGCGGATCCTCCCCGTGTCTGGGTCCTCTCCGGGCATCTCTCCTC
CCTCACCCAACCCCATGCCGTCTTCACTCGCTGGGTTCCCTTTTCCTTCTCCTTCT
GGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCTTCTCCGACGGATGTCTCCCT
TGCGTCCCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACAACC
CCAAAGTACCCCGTCTCCCTGGCTTTAGCCACCTCTCCATCCTCTTGCTTTCTTTG
CCTGGACACCCCGTTCTCCTGTGGATTCGGGTCACCTCTCACTCCTTTCATTTGGG
CAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGT
CATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGG
CCCCTATGTCCACTTCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCC
GAGCTGGGACCACCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGT
ACTTTTATCTGTCCCCTCCACCCCACAGTGGGGCCACTAGGGACAGGATTGGTGA
CAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAA
CCCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACACACCCCCATTTCCTGGA
GCCATCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGG
ATCCTGGGAGGGAGAGCTTGGCAGGGGGTGGGAGGGAAGGGGGGGATGCGTGA
CCTGCCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCCAGAACCTGAGCTGC
TCTGACGCGGCCGTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCCTTACA
CTTCCCAAGAGGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAGTTGGTCCTGAG
TTCTAACTTTGGCTCTTCACCTTTCTAGTCCCCAATTTATATTGTTCCTCCGTGCGT
CAGTTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCCTGGCAGGGCTGTGG
TGAGGAGGGGGGTGTCCGTGTGGAAAACTCCCTTTGTGAGAATGGTGCGTCCTAG
GTGTTCACCAGGTCGTGGCCGCCTCTACTCCCTTTCTCTTTCTCCATCCTTCTTTCC
TTAAAGAGTCCCCAGTGCTATCTGGGACATATTCCTCCGCCCAGAGCAGGGTCCC
GCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGAGTCCTTGGCAAGCCCAGG
AGAGGCGCTCAGGCTTCCCTGTCCCCCTTCCTCGTCCACCATCTCATGCCCCTGG
CTCTCCTGCCCCTTCCCTACAGGGGTTCCTGGCTCTGCTCTTCAGACTGAGCCCC
GTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTTCCCCTGCATCCCCCAGAGGC
CCCAGGCCACCTACTTGGCCTGGACCCCACGAGAGGCCACCCCAGCCCTGTCTA
CCAGGCTGCCTTTTGGGTGGATTCTCCTCCAACTGTGGGGTGACTGCTTGGCAAA
CTCACTCTTCGGGGTATCCCAGGAGGCCTGGAGCATTGGGGTGGGCTGGGGTTC
AGAGAGGAGGGATTCCCTTCTCAGGTTACGTGGCCAAGAAGCAGGGGAGCTGGG
TTTGGGTCAGGTCTGGGTGTGGGGTGACCAGCTTATGCTGTTTGCCCAGGACAGC
CTAGTTTTAGCACTGAAACCCTCAGTCCTAGGAAAACAGGGATGGTTGGTCACTGT
CTCTGGGTGACTCTTGATTCCCGGCCAGTTTCTCCACCTGGGGCTGTGTTTCTCGT
CCTGCATCCTTCTCCAGGCAGGTCCCCAAGCATCGCCCCCCTGCTGTGGCTGTTC
CCAAGTTCTTAGGGTACCCCACGTGGGTTTATCAACCACTTGGTGAGGCTGGTACC
CTGCCCCCATTCCTGCACCCCAATTGCCTTAGTGGCTAGGGGGTTGGGGGCTAGA

GTAGGAGGGGCTGGAGCCAGGATTCTTAGGGCTGAACAGAGAAGAGCTGGGGGC
CTGGGCTCCTGGGTTTGAGAGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCCGA
GGGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCTGAGGGTGGAGGGACTGGGG
GCCTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCCTGGACTCGTGGGTC
TGAGGGAGGAGGGGCTGGGGGCCTGGACTTCTGGGTCTTAGGGAGGCGGGGCT
GGGCCTGGACCCCTGGGTCTGAATGGGGAGAGGCTGGGGGCCTGGACTCCTTCA
TCTGAGGGCGGAAGGGCTGGGGCCTGGCCTCCTGGGTTGAATGGGGAGGGGTTG
GGCCTGGACTCTGGAGTCCCTGGTGCCCAGGCCTCAGGCATCTTTCACAGGGATG
CCTGTACTGGGCAGGTCCTTGAAAGGGAAAGGCCCATTGCTCTCCTTGCCCCCCT
CCCCTATCGCCATGACAACTGGGTGGAAATAAACGAGCCGAGTTCATCCCGTTCC
CAGGGCACGTGCGGCCCCTTCACAGCCCGAGTTTCCATGACCTCATGCTCTTGGC
CCTCGTAGCTCCCTCCCGCCTCCTCCAGATGGGCAGCTTTGGAGAGGTGAGGGAC
TTGGGGGGTAATTTATCCCGTGGATCTAGGAGTTTAGCTTCACTCCTTCCTCAGCT
CCAGTTCAGGTCCCGGAGCCCACCCAGTGTCCACAAGGCCTGGGGCAAGTCCCT
CCTCCGACCCCCTGGACTTCGGCTTTTGTCCCCCCAAGTTTTGGACCCCTAAGGG
AAGAATGAGAAACGGTGGCCCGTGTCAGCCCCTGGCTGCAGGGCCCCGTGCAGA
GGGGGCCTCAGTGAACTGGAGTGTGACAGCCTGGGGCCCAGGCACACAGGTGTG
CAGCTGTCTCACCCCTCTGGGAGTCCCGCCCAGGCCCCTGAGTCTGTCCCAGCAC
AGGGTGGCCTTCCTCCACCCTGCATAGCCCTGGGCCCACGGCTTCGTTCCTGCAG
AGTATCTGCTGGGGTGGTTTCCGAGCTTGACCCTTGGAAGGACCTGGCTGGGTTT
AAGGCAGGAGGGGCTGGGGGCCAGGACTCCTGGCTCTGAAGGAGGAGGGGCTG
GAACCTCTTCCCTAGTCTGAGCACTGGAAGCGCCACCTGTGGGTGGTGACGGGG
GTTTTGCCGTGTCTAACAGGTACCATGTGGGGTTCCCGCACCCAGATGAGAAGCC
CCCTCCCTTCCCCGTTCACTTCCTGTTTGCAGATAGCCAGGAGTCCTTTCGTGGTT
TCCACTGAGCACTGAAGGCCTGGCCGGCCTGACCACTGGGCAACCAGGCGTATC
TTAAACAGCCAGTGGCCAGAGGCTGTTGGGTCATTTTCCCCACTGTCCTAGCACC
GTGTCCCTGGATCTGTTTTCGTGGCTCCCTCTGGAGTCCCGACTTGCTGGGACAC
CGTGGCTGGGGTAGGTGCGGCTGACGGCTGTTTCCCACCCCCAG (SEQ ID NO:11)

[0145] Cas9 guide RNAs were designed for targeting the human AAVS1 locus. A 42 nucleotide RNA (referred to herein as a "crRNA" sequence) comprising (5' to 3') a target recognition sequence (i.e., sequence complementary to the non-coding strand of the target sequence) and protospacer sequence; a 85 nucleotide RNA (referred to herein as a "tracrRNA" sequence) comprising 5' sequence with complementarity to the 3' sequence of the crRNA and additional hairpin sequence; and a chimeric RNA comprising nucleotides 1-32 of the crRNA, a GAAA loop, and nucleotides 19-45 of the tracrRNA were prepared. The crRNA was chemically synthesized by Sigma-Aldrich. The tracrRNA and chimeric RNA were synthesized by *in vitro* transcription with T7 RNA polymerase using T7-Scribe Standard RNA IVT Kit (Cellscript). The chimeric RNA coding sequence was also placed under the control of

human U6 promoter for *in vivo* transcription in human cells. Table 4 presents the sequences of the guide RNAs.

| Table 4. Guide RNAs | | |
|---|---|---|
| RNA | 5'-3'Sequence | SEQ ID NO: |
| AAVS1-crRNA | ACCCCACAGUGGGGCCACUAGUUUUAGAGCUAUGCUGUUUUG | 12 |
| tracrRNA | GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU | 13 |
| chimeric RNA | ACCCCACAGUGGGGCCACUAGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCG | 14 |

***Example 3: Preparation of Donor Polynucleotide to Monitor Genome Modification***

[0146] Targeted integration of a GFP protein into the N terminus of PPP1R12C was used to monitor Cas9-mediated genome modification. To mediate integration by homologous recombination a donor polynucleotide was prepared. The AAVS1-GFP DNA donor contained a 5' (1185 bp) AAVS1 locus homologous arm, an RNA splicing receptor, a turbo GFP coding sequence, a 3' transcription terminator, and a 3' (1217 bp) AAVS1 locus homologous arm. Table 5 presents the sequences of the RNA splicing receptor and the GFP coding sequence followed by the 3' transcription terminator. Plasmid DNA was prepared by using GenElute Endotoxin-Free Plasmid Maxiprep Kit (Sigma).

| Table 5. Sequences in the AAVS1-GFP DNA donor sequence | | |
|---|---|---|
| | 5'-3' Sequence | SEQ ID NO: |
| RNA splicing receptor | CTGACCTCTTCTCTTCCTCCCACAG | 15 |
| GFP coding sequence and transcription terminator | GCCACCATGGACTACAAAGACGATGACGACAAGGTCGACTCTAGAGCTGCAGAGAGCGACGAGAGCGGCCTGCCCGCCATGGAGATCGAGTGCCGCATCACCGGCACCCTGAACGGCGTGGAGTTCGAGCTGGTGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCATGACCAACAAGATGAAGAGCACCAAAGGCGCCCTGACCTTCAGCCCCTACCTGCTGAGCCACGTGATGGGCTACGGCTTCTACCACTTCGGCACCTACCCCAGCGGCT | 16 |

| | | |
|---|---|---|
| | ACGAGAACCCCTTCCTGCACGCCATCAACAACGGCGGCTA<br>CACCAACACCCGCATCGAGAAGTACGAGGACGGCGGCGT<br>GCTGCACGTGAGCTTCAGCTACCGCTACGAGGCCGGCCG<br>CGTGATCGGCGACTTCAAGGTGATGGGCACCGGCTTCCCC<br>GAGGACAGCGTGATCTTCACCGACAAGATCGTCCGCAGCA<br>ACGCCACCGTGGAGCACCTGCACCCCATGGGCGATAACG<br>ATCTGGATGGCAGCTTCACCCGCACCTTCAGCCTGCGCGA<br>CGGCGGCTACTACAGCTCCGTGGTGGACAGCCACATGCAC<br>TTCAAGAGCGCCATCCACCCCAGCATCCTGCAGAACGGGG<br>GCCCCATGTTCGCCTTCCGCCGCGTGGAGGAGGATCACA<br>GCAACACCGAGCTGGGCATCGTGGAGTACCAGCACGCCTT<br>CAAGACCCCGGATGCAGATGCCGGTGAAGAATGAAGATCT<br>CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCC<br>CCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTG<br>TCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGA<br>GTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG<br>ACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATG<br>CTGGGGATGCGGTGGGCTCTATGGACTCGAGGTTTAAACG<br>TCGACGCGGCCGCGT | |

[0147] Targeted gene integration will result in a fusion protein between the first 107 amino acids of the PPP1R12C and the turbo GFP. The expected fusion protein contains the first 107 amino acid residues of PPP1R12C (highlighted in grey) from RNA splicing between the first exon of PPP1R12C and the engineered splice receptor (see Table 6).

| Table 6. Predicted amino acid sequence of the PPP1R12C-GFP fusion protein. |
|---|
| MSGEDGPAAGPGAAAAAARERRREQLRQWGARAGAEPGPGERRARTVRFERAAEF<br>LAACAGGDLDEARLMLRAADPGPGAELDPAAPPPARAVLDSTNADGISALHQATMDY<br>KDDDDKVDSRAAESDESGLPAMEIECRITGTLNGVEFELVGGGEGTPEQGRMTNKMK<br>STKGALTFSPYLLSHVMGYGFYHFGTYPSGYENPFLHAINNGGYTNTRIEKYEDGGVL<br>HVSFSYRYEAGRVIGDFKVMGTGFPEDSVIFTDKIVRSNATVEHLHPMGDNDLDGSFT<br>RTFSLRDGGYYSSVVDSHMHFKSAIHPSILQNGGPMFAFRRVEEDHSNTELGIVEYQH<br>AFKTPDADAGEE (SEQ ID NO:17) |

***Example 4: Cas9-Mediated Targeted Integration***

[0148] Transfection was performed on human K562 cells. The K562 cell line was obtained from American Type Culture Collection (ATCC) and grown in Iscove's Modified Dulbecco's Medium, supplemented with 10% FBS and 2 mM L-glutamine. All media and supplements were obtained from Sigma-Aldrich. Cultures were split one day before transfection (at approximately 0.5 million cells per mL before transfection). Cells were transfected with Nucleofector Solution V (Lonza) on a Nucleofector (Lonza) with the T-016 program. Each nucleofection contained approximately 0.6 million cells. Transfection treatments are detailed in Table 7. Cells were grown at 37°C and 5% $CO_2$ immediately after nucleofection.

Table 7. Transfection Treatments.

| Treatment | Modified Cas9 | Guide RNA | Donor sequence |
|---|---|---|---|
| A | Cas9 mRNA transcribed with an Anti-Reverse Cap Analog (10 µg) | pre-annealed crRNA-tracrRNA duplex (0.3 nmol) | AAVS1-GFP plasmid DNA (10 µg) |
| B | Cas9 mRNA transcribed with an Anti-Reverse Cap Analog (10 µg) | chimeric RNA (0.3 nmol) | AAVS1-GFP plasmid DNA (10 µg) |
| C | Cas9 mRNA capped via post-transcription capping reaction (10 µg) | chimeric RNA (0.3 nmol) | AAVS1-GFP plasmid DNA (10 µg) |
| D | Cas9 plasmid DNA (10 µg) | U6-chimeric RNA plasmid DNA (5 µg) | AAVS1-GFP plasmid DNA (10 µg) |
| E | None | None | AAVS1-GFP plasmid DNA (10 µg) |
| F | None | None | None |

[0149] Fluorescence-activated cell sorting (FACS) was performed 4 days after transfection. FACS data are presented in **FIG. 4**. The percent GFP detected in each of the four experimental treatments (A-D) was greater than in the control treatments (E, F), confirming integration of the donor sequence and expression of the fusion protein.

***Example 5: PCR Confirmation of Targeted Integration***

[0150] Genomic DNA was extracted from transfected cells with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) 12 days after transfection. Genomic DNA was then PCR amplified with a forward primer located outside the 5' homologous arm of the AAVS1-GFP plasmid donor and a reverse primer located at the 5' region of the GFP. The forward primer was 5'- CCACTCTGTGCTGACCACTCT-3' (SEQ ID NO:18) and reverse primer was 5'- GCGGCACTCGATCTCCA-3' (SEQ ID NO:19). The expected fragment size from the junction PCR was 1388 bp. The amplification was carried out with JumpStart Taq ReadyMix (Sigma), using the following cycling conditions: 98℃ for 2 minutes for initial denaturation; 35 cycles of 98℃ for 15 seconds, 62℃ for 30 seconds, and 72℃ for 1minutes and 30 seconds; and a final extension at 72℃ for 5 minutes. PCR products were resolved on 1% agarose gel.

[0151] Cells transfected with 10 µg of Cas9 mRNA transcribed with an Anti-Reverse Cap Analog, 0.3 nmol of pre-annealed crRNA-tracrRNA duplex, and 10 µg of AAVS1-GFP plasmid DNA displayed a PCR product of the expected size (see lane A, **FIG. 5**).

***Example 6: Cas9-Based Genome Editing in Mouse Embryos***

[0152] The mouse Rosa26 locus can be targeted for genome modifications. Table 8 presents a portion of the mouse Rosa26 sequence in which potential target sites are shown in bold. Each target site comprises a protospacer.

Table 8. Mouse Rosa26 Sequence

GAGCGGCTGCGGGGCGGGTGCAAGCACGTTTCCGACTTGAGTTGCCTCAAGAGG
GGCGTGCTGAGCCAGACCTCCATCGCGCACTCCGGGGAGTGGAGGGAAGGAGC
GAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCCAAAGTCGC
TCTGAGTTGTTATCAGTAAGGGAGCTGCAGTGGAGTAGGCGGGGAGAAGGCCGC
ACCCTTCTCCGGAGGGGGGAGGGGAGTGTTGCAATACCTTTCTGGGAGTTCTCTG
CTGCCTCCTGGCTTCTGAGGACCGCCCTGGGCCTGGGAGAATCCCTTCCCCCTCT
TCCCTCGTGATCTGCAA**CTCCAGTCTTTCTAGAAGAT**GGGCGGGAGTCTTCTGGG
CAGGCTTAAAGGCTAACCTGGTGTGTGGGCGTTGTCCTGCAGGGGAAT**TGAACAG**
**GTGTAAAATTGGA**GGGACAAGACTTCCCACAGATTTTCGGTTT**TGTCGGGAAGTTT**
**TTTAATA**GGGGCAAATAAGGAAAATGGGAGGATAGGTAGTCATCTGGGGTTTTATG

| |
|---|
| CAGCAAAACTACAGGTTATTATTGCTTGTGATCCGCCTCGGAGTATTTTCCATCGA GGTAGATTAAAGACATGCTCACCCGAGTTTTATACTCTCCTGCTTGAGATCCTTACT ACAGTATGAAATTACAGTGTCGCGAGTTAGACTATGTAAGCAGAATTTTA (SEQ ID NO:20) |

[0153] Guide RNAs were designed to target each of the target sites in the mouse Rosa26 locus. The sequences are shown in Table 9, each is 42 nucleotides in length and the 5' region is complementary to the strand that is not presented in Table 8 (i.e., the strand that is complementary to the strand shown in Table 8).

Table 9. Mouse Rosa26 Guide RNAs

| RNA | 5'-3'Sequence | SEQ ID NO: |
|---|---|---|
| mRosa26-crRNA-1 | CUCCAGUCUUUCUAGAAGAUGUUUUAGAGCUAU GCUGUUUUG | 21 |
| mRosa26-crRNA-2 | UGAACAGGUGUAAAAUUGGAGUUUUAGAGCUAU GCUGUUUUG | 22 |
| mRosa26-crRNA-3 | UGUCGGGAAGUUUUUUAAUAGUUUUAGAGCUAU GCUGUUUUG | 23 |

[0154] The crRNAs were chemically synthesized and pre-annealed to the tracrRNA (SEQ ID NO:13; see Example 2). Pre-annealed crRNA / tracrRNA and *in vitro* transcribed mRNA encoding modified Cas9 protein (SEQ ID NO. 9; see Example 1) can be microinjected into the pronuclei of fertilized mouse embryos. Upon guidance to the target set by the crRNA, the Cas9 protein cleaves the target site, and the resultant double-stranded break can be repaired via a non-homologous end-joining (NHEJ) repair process. The injected embryos can be either incubated at 37℃, 5% $CO_2$ overnight or for up to 4 days, followed by genotyping analysis, or the injected embryos can be implanted into recipient female mice such that live born animals can be genotyped. The *in vitro*-incubated embryos or tissues from live born animals can be screened for the presence of Cas9-induced mutation at the Rosa locus using standard methods. For example, the embryos or tissues from fetus or live-born animals can be harvested for DNA extraction and analysis. DNA can be isolated using standard procedures. The targeted region of the Rosa26 locus can be PCR amplified using appropriate primers.

Because NHEJ is error-prone, deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. Mutations can be detected using PCR-based genotyping methods, such as Cel-I mismatch assays and DNA sequencing.

***Example 7: Cas9-Based Genome Modification in Mouse Embryos***

[0155] The Rosa26 locus can be modified in mouse embryos by co-injecting a donor polynucleotide, as detailed above in section (IV)(d), along with the pre-annealed crRNA / tracrRNA and mRNA encoding modified Cas9 as described above in Example 6. *In vitro*-incubated embryos or tissues from live born animals (as described in Example 6) can be screened for a modified Rosa26 locus using PCR-based genotyping methods, such as RFLP assays, junction PCR, and DNA sequencing.

***Example 8: Cas9-Based Genome Editing in Rat Embryos***

[0156] The rat Rosa26 locus can be targeted for genome modifications. Table 10 presents a portion of the rat sequence in which potential target sites are shown in bold. Each target site comprises a protospacer.

Table 10. Rat Rosa26 Sequence

GGGATTCCTCCTTGAGTTGTGGCACTGAGGAACGTGCTGAACAAGACCTACATTG
CACTCCAGGGAGTGGATGAAGGAGTTGGGGCTCAGTCGGGTTGTATTGGAGACAA
GAAGCACTTGCTCTCCAAAAGTCGGTTTGAGTTATCATTAAGGGAGCTGCAGTGGA
GTAGGCGGAGAAAAGGCCGCACCCTTCTCAGGACGGGGGAGGGGAGTGTTGCAA
TACCTTTCTGGGAGTTCTCTGCTGCCTCCTGTCTTCTGAGGACCGCCCTGGGCC**T**
**GGAAGATTCCCTTCCCCCT**TCTTCCCTCGTGA**TCTGCAACTGGAGTCTTTCT**GGAA
GAT**AGGCGGGAGTCTTCTGGGCA**GGCTTAAAGGCTAACCTGGTGCGTGGGGCGT
TGTCCTGCAGAGGAATTGAACAGGTGTAAAATTGGAGGGGCAAGACTTCCCACAG
ATTTTCGATTGTGTTGTTAAGTATTGTAATAGGGGCAAATAAGGGAAATAGACTAGG
CACTCACCTGGGGTTTTATGCAGCAAAACTACAGGTTATTATTGCTTGTGATCCGC
CCTGGAGAATTTTTCACCGAGGTAGATTGAAGACATGCCCACCCAAATTTTAATATT
CTTCCACTTGCGATCCTTGCTACAGTATGAAA (SEQ ID NO:24)

[0157] Guide RNAs were designed to target each of the target sites in the rat Rosa26 locus. The sequences are shown in Table 11, each is 42 nucleotides in

length and the 5' region is complementary to the strand that is not presented in Table 10 (i.e., the strand that is complementary to the strand shown in Table 10).

| Table 11. Rat Rosa26 Guide RNAs | | |
|---|---|---|
| RNA | 5'-3'Sequence | SEQ ID NO: |
| rRosa26-crRNA-1 | AGGGGGAAGGGAAUCUUCCAGUUUUAGAGCUA UGCUGUUUUG | 25 |
| rRosa26-crRNA-2 | UCUGCAACUGGAGUCUUUCUGUUUUAGAGCUA UGCUGUUUUG | 26 |
| rRosa26-crRNA-3 | AGGCGGGAGUCUUCUGGGCAGUUUUAGAGCUA UGCUGUUUUG | 27 |

[0158] The crRNAs were chemically synthesized and pre-annealed to the tracrRNA (SEQ ID NO:13; see Example 2). Pre-annealed crRNA / tracrRNA and *in vitro* transcribed mRNA encoding modified Cas9 protein (SEQ ID NO. 9; see Example 1) can be microinjected into the pronuclei of fertilized rat embryos. Upon guidance to the target site by the crRNA, the Cas9 protein cleaves the target site, and the resultant double-stranded break can be repaired via a non-homologous end-joining (NHEJ) repair process. The injected embryos can be either incubated at 37°C, 5% $CO_2$ overnight or for up to 4 days, followed by genotyping analysis, or the injected embryos can be implanted into recipient female mice such that live born animals can be genotyped. The *in vitro*-incubated embryos or tissues from live born animals can be screened for the presence of Cas9-induced mutation at the Rosa locus using standard methods. For example, the embryos or tissues from fetus or live-born animals can be harvested for DNA extraction and analysis. DNA can be isolated using standard procedures. The targeted region of the Rosa26 locus can be PCR amplified using appropriate primers. Because NHEJ is error-prone, deletions of at least one nucleotide, insertions of at least one nucleotide, substitutions of at least one nucleotide, or combinations thereof can occur during the repair of the break. Mutations can be detected using PCR-based genotyping methods, such as Cel-I mismatch assays and DNA sequencing.

***Example 9: Cas9-Based Genome Modification in Rat Embryos***

[0159] The Rosa26 locus can be modified in rat embryos by co-injecting a donor polynucleotide, as detailed above in section (IV)(d), along with the pre-annealed crRNA / tracrRNA and mRNA encoding modified Cas9 as described above in Example 8. *In vitro*-incubated embryos or tissues from live born rats (as described in Example 8) can be screened for a modified Rosa26 locus using PCR-based genotyping methods, such as RFLP assays, junction PCR, and DNA sequencing.

***Related Applications***

This application is a divisional of AU2018229489, AU 2017204031 and AU 2013355214, which is the Australian national phase of International Patent Application No. PCT/US2013/073307 (WO 2014/089290), each of which is incorporated herein by reference in its entirety, and claims priority to US 61/794,422 filed 15 March 2013, US 61/761,046 filed 5 February 2013, US 61/758,624 filed 30 January 2013, and US 61/734,256 filed 6 December 2012, each of which is incorporated herein by reference in its entirety.

**The claims defining the invention are as follows:**

1. Vectors comprising:

(a) a DNA coding sequence encoding at least one guide RNA operably linked to a promoter control sequence for expression of the at least one guide RNA in a eukaryotic cell, each guide RNA comprising

(i) a first region complementary to a target site in a eukaryotic chromosomal sequence that can base pair with the target site,

(ii) a second region that forms a stem and loop structure, and

(iii) a third region which is essentially single stranded,

wherein (i), (ii) and (iii) are arranged in the 5' to 3' direction,

(b) a DNA coding sequence encoding an engineered RNA-guided endonuclease operably linked to a promoter control sequence for expression in a eukaryotic cell, wherein the RNA-guided endonuclease is a type II CRISPR/Cas9 endonuclease comprising at least one nuclear localization signal,

wherein (a) and (b) are located on the same or different vectors,

whereby the RNA-guided endonuclease is directed to specific nucleic acid sequences by the at least one guide RNA and the RNA-guided endonuclease cleaves the specific nucleic acid sequences, whereby the nucleic acid sequence is modified by a deletion of at least one nucleotide, an insertion of at least one nucleotide, a substitution of at least one nucleotide or a combination thereof.

2. The vectors of claim 1, wherein the guide RNA comprises two separate molecules.

3. The vectors of claim 2, wherein the first molecule of the guide RNA comprises the first region of the guide RNA and one half of the stem of the second region of the guide RNA.

4. The vectors of claim 2, wherein the second molecule of the guide RNA comprises the other half of the stem of the second region of the guide RNA and the third region of the guide RNA.

5. The vectors of any of claims 1 to 4, wherein the first region of the guide RNA comprises from about 10 nucleotides to more than about 25 nucleotides.

6. The vectors of any of claims 1 to 5, wherein the second region of the guide RNA is about 16 to about 60 nucleotides in length.

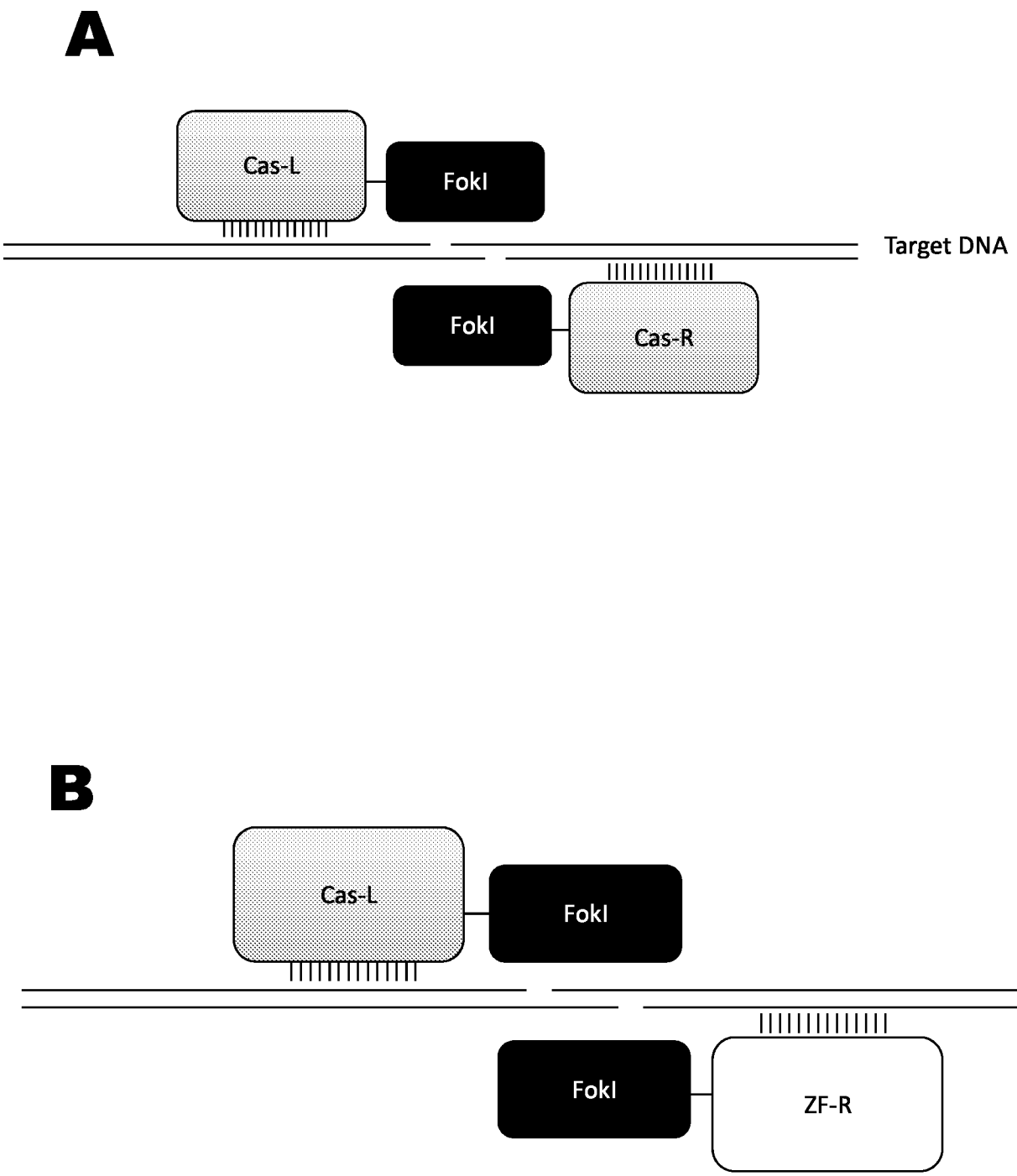


FIG. 1

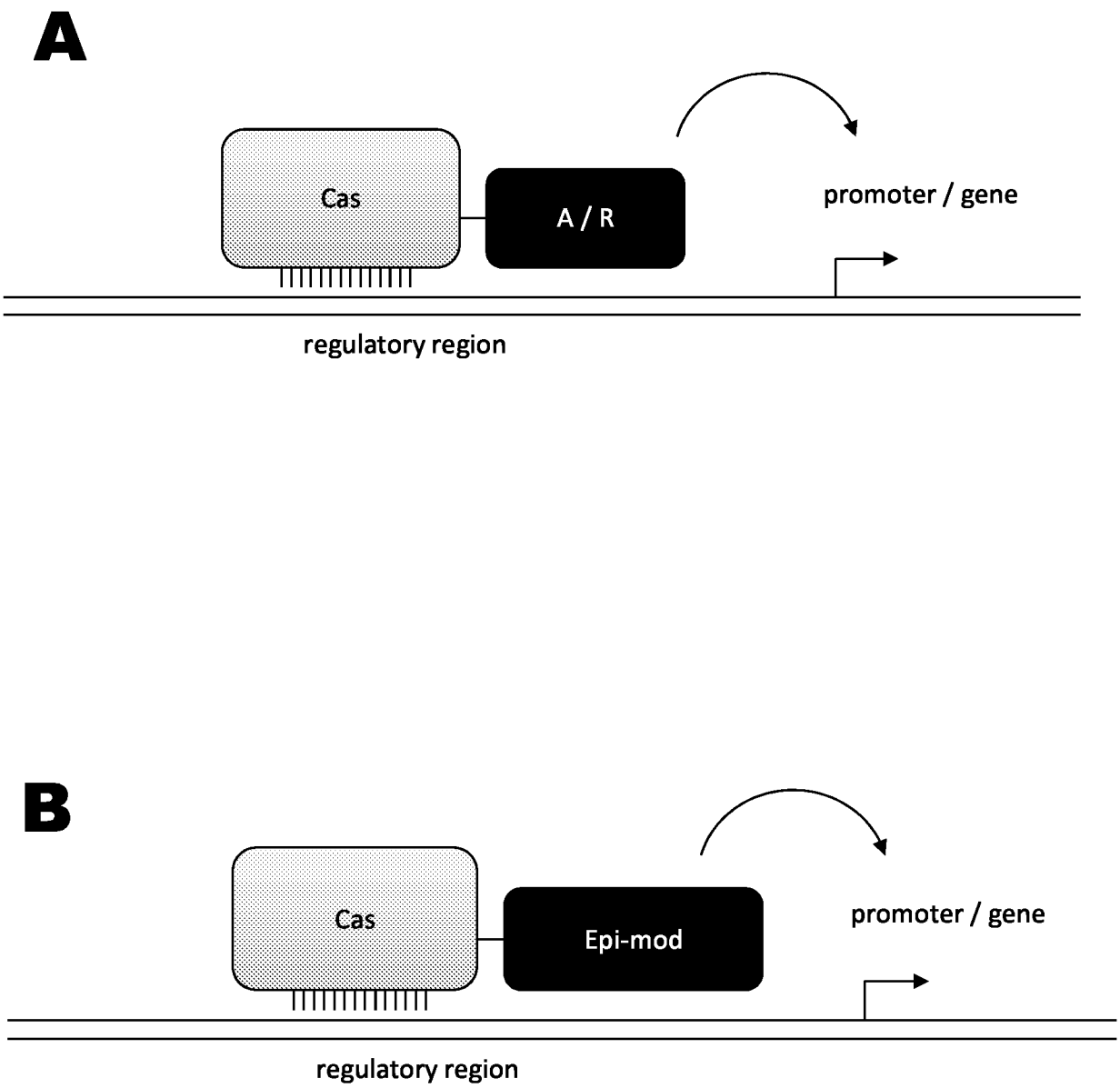


FIG. 2

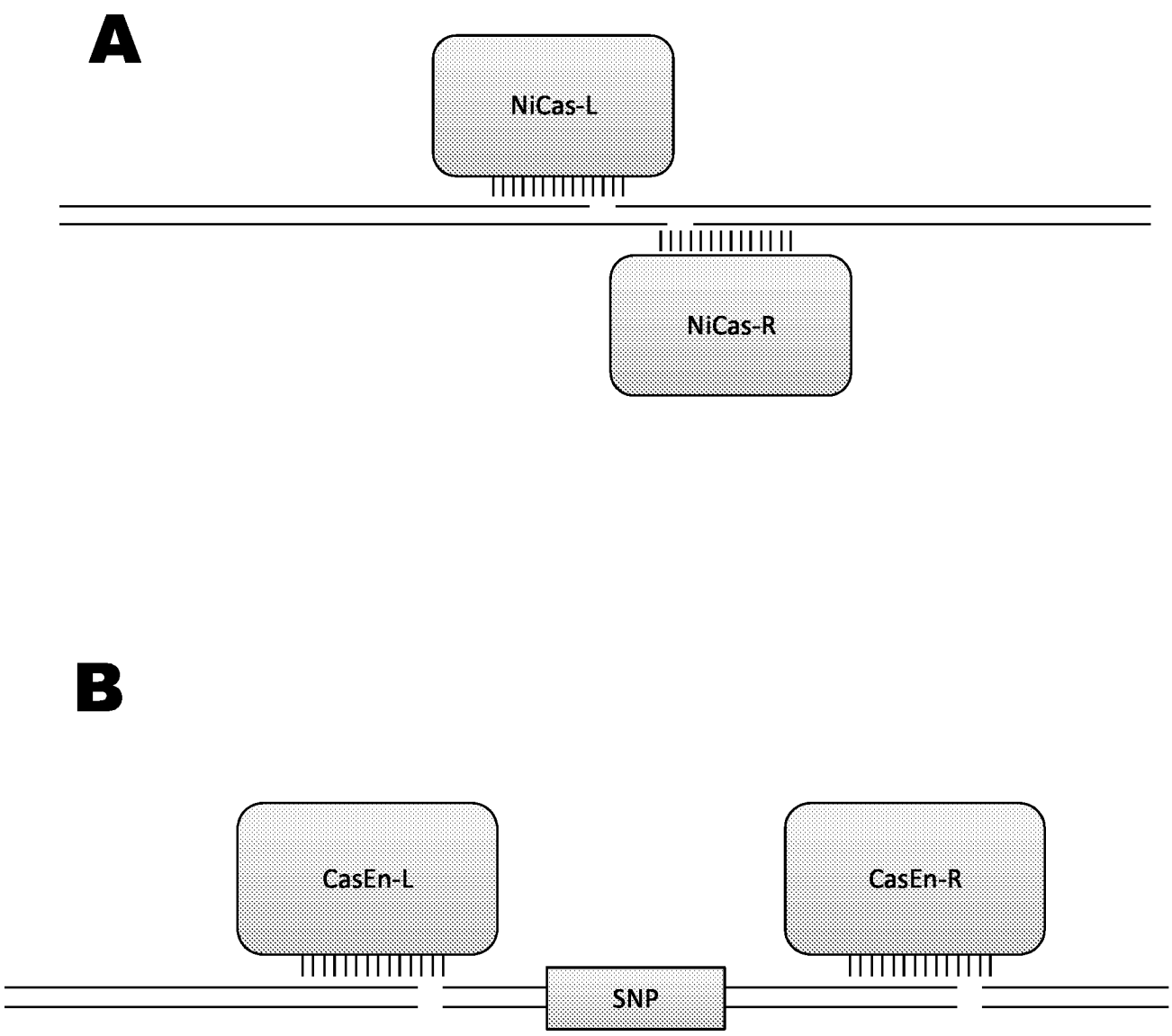


FIG. 3

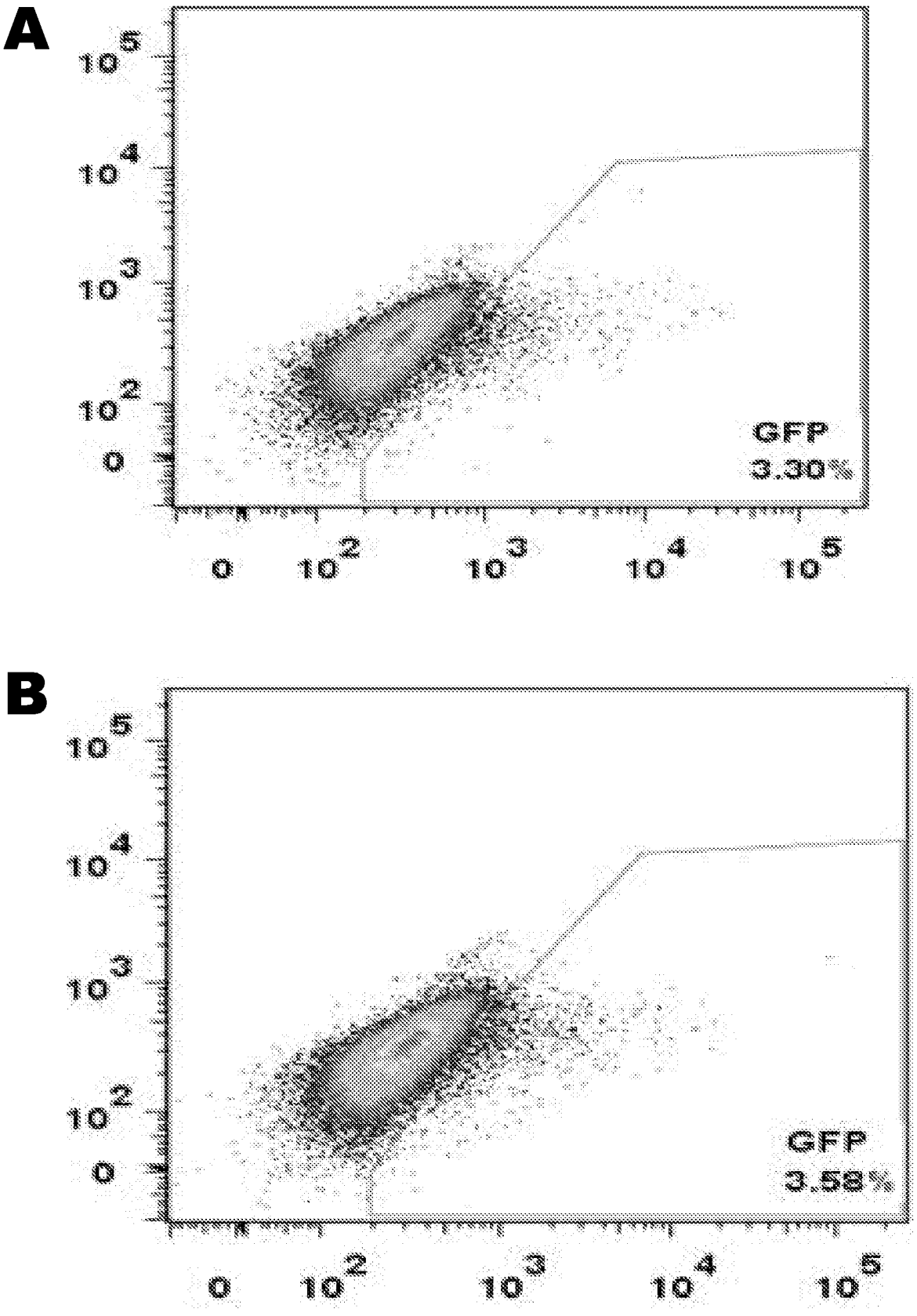


FIG. 4

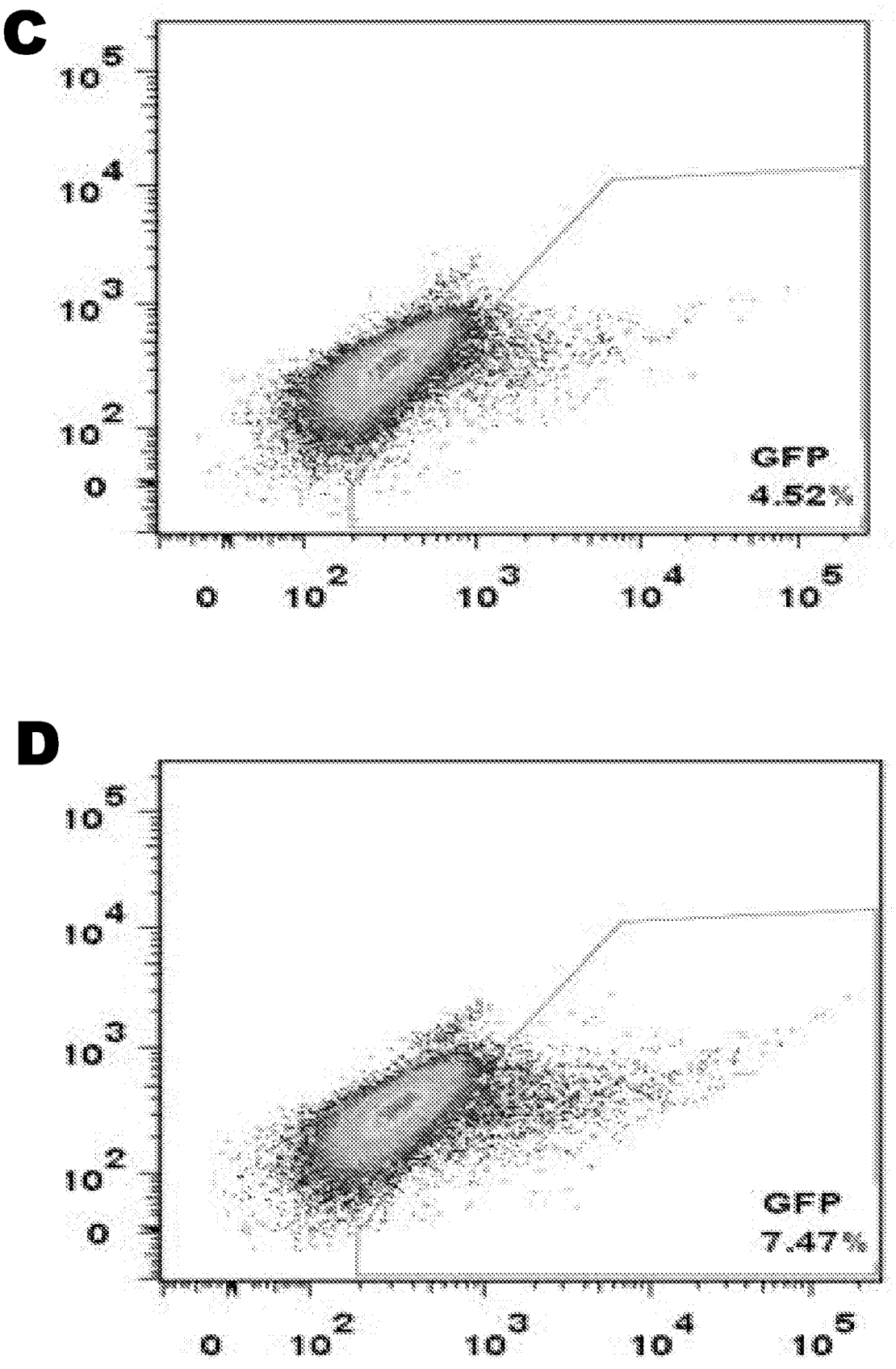


FIG. 4

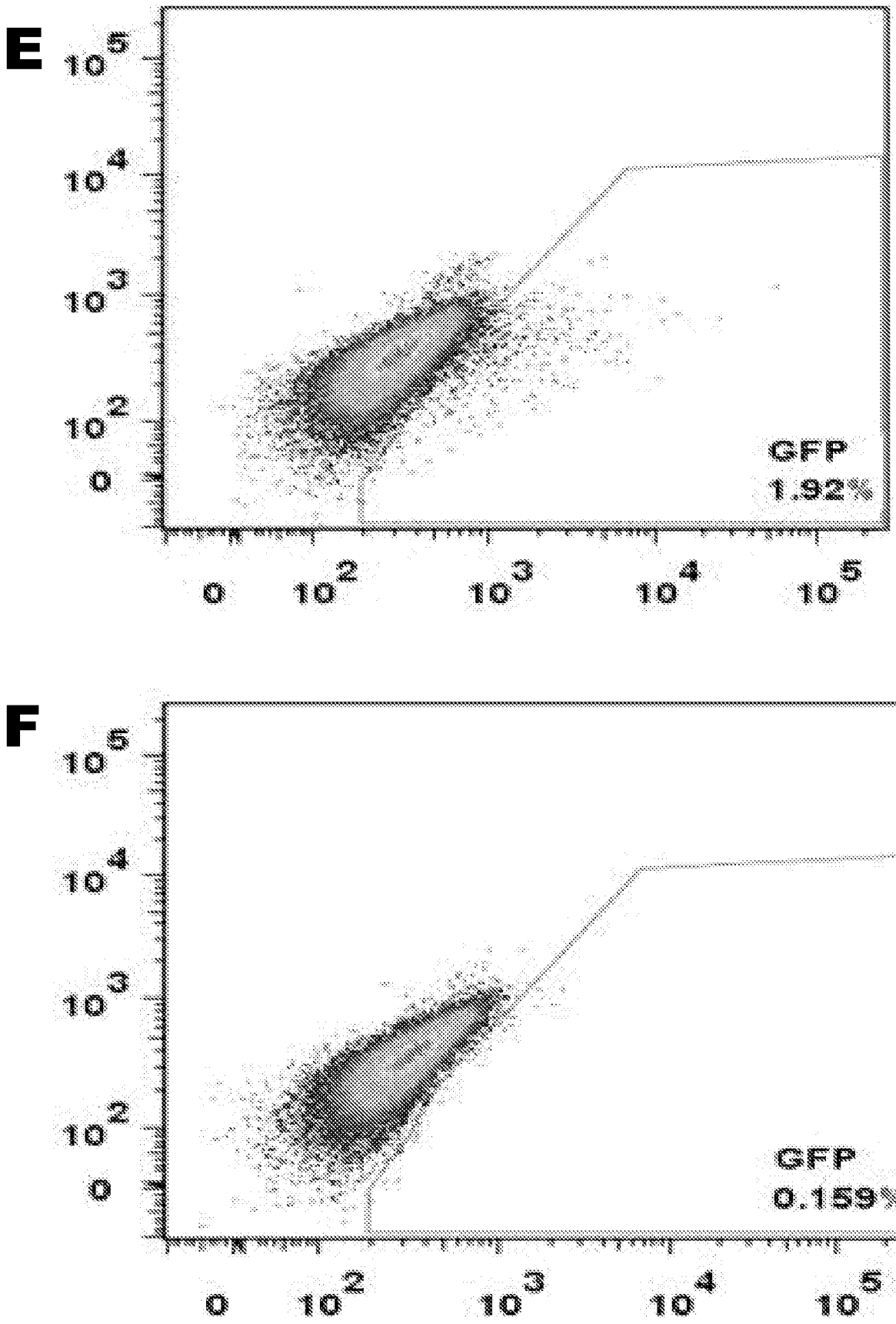


FIG. 4

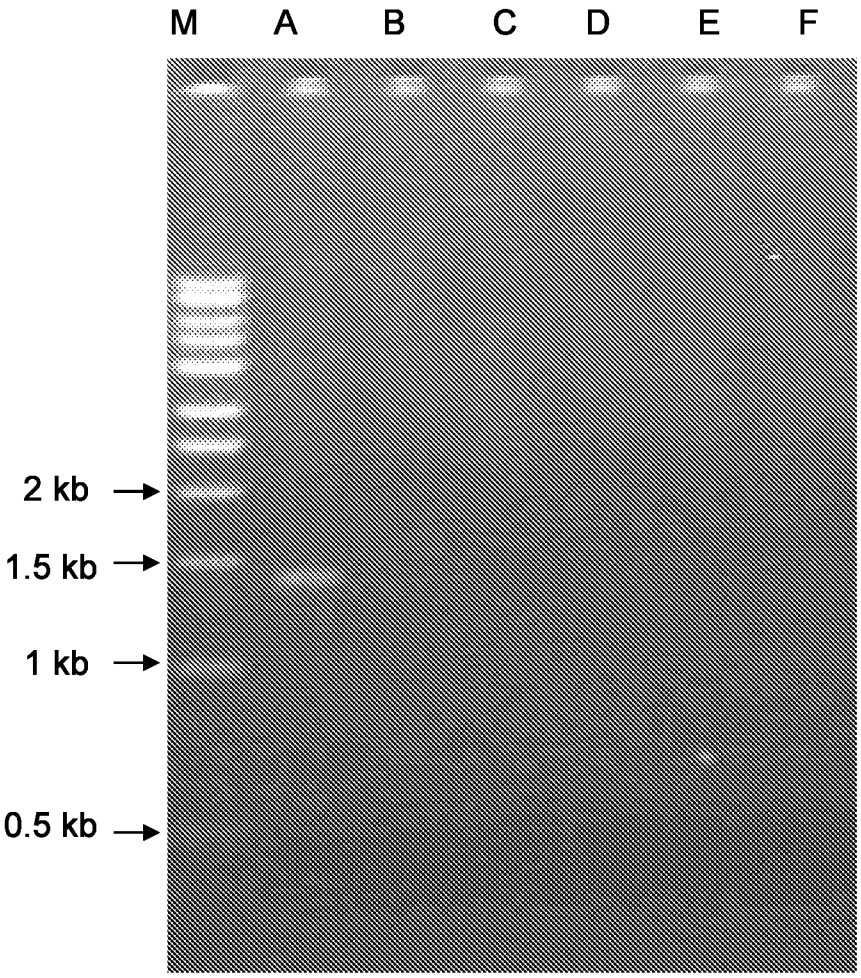


FIG. 5

SEQUENCE LISTING

<110> SIGMA-ALDRICH CO. LLC
CHEN, Fuqiang
DAVIS, Gregory D.
KANG, Qiaohua
KNIGHT, Scott W.

<120> CRISPR-BASED GENOME MODIFICATION AND REGULATION

<130> 047497-465606

<150> US 61/734,256
<151> 2012-12-06

<150> US 61/758,624
<151> 2013-01-30

<150> US 61/761,046
<151> 2013-02-05

<150> US 61/794,422
<151> 2013-03-15

<160> 27

<170> PatentIn version 3.5

<210> 1
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 1

Pro Lys Lys Lys Arg Lys Val
1 5

<210> 2
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 2

Pro Lys Lys Lys Arg Arg Val

1 5

<210> 3
<211> 16
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 3

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys
1 5 10 15

<210> 4
<211> 20
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 4

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys
1 5 10 15

Lys Arg Lys Val
20

<210> 5
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 5

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys
1 5 10 15

Arg Lys Val

<210> 6
<211> 24
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 6

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly
1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val
20

<210> 7
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 7

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly
1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val
20 25

<210> 8
<211> 21
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 8

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys
1 5 10 15

Lys Lys Arg Lys Val
20

<210> 9
<211> 1374
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 9

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45

Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

```
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
            180                 185                 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala
        195                 200                 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Arg Asn Gly Leu Phe Gly Asn
225                 230                 235                 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300

Ile Leu Arg Val Asn Ser Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320

Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335

Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350

Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365

Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380
```

```
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400

Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415

Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
```

```
Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620

Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
            660                 665                 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
        675                 680                 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
    690                 695                 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
705                 710                 715                 720

His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
                725                 730                 735

Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
            740                 745                 750

His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr
        755                 760                 765

Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile Glu
    770                 775                 780

Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro Val
785                 790                 795                 800

Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu Gln
```

```
                805                 810                 815

Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg Leu
            820                 825                 830

Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Ile Lys Asp
        835                 840                 845

Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg Gly
    850                 855                 860

Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys Asn
865                 870                 875                 880

Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys Phe
                885                 890                 895

Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp Lys
            900                 905                 910

Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr Lys
        915                 920                 925

His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp Glu
    930                 935                 940

Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser Lys
945                 950                 955                 960

Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg Glu
                965                 970                 975

Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val Val
            980                 985                 990

Gly Thr Ala Leu Ile Lys Lys Tyr  Pro Lys Leu Glu Ser  Glu Phe Val
        995                 1000                1005

Tyr Gly  Asp Tyr Lys Val Tyr  Asp Val Arg Lys Met  Ile Ala Lys
    1010                1015                1020
```

```
Ser Glu  Gln Glu Ile Gly Lys  Ala Thr Ala Lys Tyr  Phe Phe Tyr
    1025                1030                1035

Ser Asn  Ile Met Asn Phe Phe  Lys Thr Glu Ile Thr  Leu Ala Asn
    1040                1045                1050

Gly Glu  Ile Arg Lys Arg Pro  Leu Ile Glu Thr Asn  Gly Glu Thr
    1055                1060                1065

Gly Glu  Ile Val Trp Asp Lys  Gly Arg Asp Phe Ala  Thr Val Arg
    1070                1075                1080

Lys Val  Leu Ser Met Pro Gln  Val Asn Ile Val Lys  Lys Thr Glu
    1085                1090                1095

Val Gln  Thr Gly Gly Phe Ser  Lys Glu Ser Ile Leu  Pro Lys Arg
    1100                1105                1110

Asn Ser  Asp Lys Leu Ile Ala  Arg Lys Lys Asp Trp  Asp Pro Lys
    1115                1120                1125

Lys Tyr  Gly Gly Phe Asp Ser  Pro Thr Val Ala Tyr  Ser Val Leu
    1130                1135                1140

Val Val  Ala Lys Val Glu Lys  Gly Lys Ser Lys Lys  Leu Lys Ser
    1145                1150                1155

Val Lys  Glu Leu Leu Gly Ile  Thr Ile Met Glu Arg  Ser Ser Phe
    1160                1165                1170

Glu Lys  Asn Pro Ile Asp Phe  Leu Glu Ala Lys Gly  Tyr Lys Glu
    1175                1180                1185

Val Lys  Lys Asp Leu Ile Ile  Lys Leu Pro Lys Tyr  Ser Leu Phe
    1190                1195                1200

Glu Leu  Glu Asn Gly Arg Lys  Arg Met Leu Ala Ser  Ala Gly Glu
    1205                1210                1215
```

Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
1220 1225 1230

Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro
1235 1240 1245

Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
1250 1255 1260

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg
1265 1270 1275

Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
1280 1285 1290

Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
1295 1300 1305

Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
1310 1315 1320

Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr
1325 1330 1335

Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1340 1345 1350

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Pro
1355 1360 1365

Lys Lys Lys Arg Lys Val
1370

<210> 10
<211> 4122
<212> DNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 10

atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg 60

atcaccgacg actacaaggt gcccagcaag aaattcaagg tgctgggcaa caccgaccgg 120

cacagcatca agaagaacct gatcggcgcc ctgctgttcg gctctggcga aacagccgag 180

gccacccggc tgaagagaac cgccagaaga agatacacca gacggaagaa ccggatctgc 240

tatctgcaag agatcttcag caacgagatg gccaaggtgg acgacagctt cttccacaga 300

ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgagaaag 420

aagctggccg acagcaccga caaggccgac ctgagactga tctacctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540

gtggacaagc tgttcatcca gctggtgcag atctacaatc agctgttcga ggaaaacccc 600

atcaacgcca gcagagtgga cgccaaggcc atcctgagcg ccagactgag caagagcaga 660

cggctggaaa atctgatcgc ccagctgccc ggcgagaagc ggaatggcct gttcggcaac 720

ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780

gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacagc gagatcacca aggcccccct gtccgcctct 960

atgatcaaga gatacgacga gcaccaccag gacctgaccc tgctgaaagc tctcgtgcgg 1020

cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080

ggctacatcg atggcggagc cagccaggaa gagttctaca agttcatcaa gcccatcctg 1140

gaaaagatgg acggcaccga ggaactgctc gtgaagctga acagagagga cctgctgcgg 1200

aagcagcgga ccttcgacaa cggcagcatc ccccaccaga tccacctggg agagctgcac 1260

gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgaccttcag aatcccctac tacgtgggcc ctctggccag gggaaacagc 1380

agattcgcct ggatgaccag aaagagcgag gaaaccatca ccccctggaa cttcgaggaa 1440

gtggtggaca agggcgccag cgcccagagc ttcatcgagc ggatgaccaa cttcgataag 1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg 1560

tacaacgagc tgaccaaagt gaaatacgtg accgagggaa tgcggaagcc cgcctttctg 1620

agcggcgagc agaaaaaggc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680

gtgaagcagc tgaaagagga ctacttcaag aaaatcgagt gcttcgacag cgtggaaatc 1740

agcggcgtgg aagatcggtt caacgcctcc ctgggcgcct atcacgatct gctgaaaatt 1800

atcaaggaca aggacttcct ggacaatgag gaaaacgagg acattctgga agatatcgtg 1860

ctgaccctga cactgtttga ggaccggggc atgatcgagg aacggctgaa aacctatgcc 1920

cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980

aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040

gatttcctga agtccgacgg cttcgccaac agaaacttca tgcagctgat ccacgacgac 2100

agcctgacct ttaaagagga catccagaaa gcccaggtgt ccggccaggg acactctctg 2160

cacgagcaga tcgccaatct ggccggatcc cccgccatta agaagggcat cctgcagaca 2220

gtgaagattg tggacgagct cgtgaaagtg atgggccaca agcccgagaa catcgtgatc 2280

gaaatggcca gagagaacca gaccacccag aagggacaga agaacagccg cgagagaatg 2340

aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acaccccgtg 2400

gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctgcagaa tgggcgggat 2460

atgtacgtgg accaggaact ggacatcaac cggctgtccg actacgatgt ggaccacatt 2520

gtgccccagt ccttcatcaa ggacgactcc atcgataaca aagtgctgac tcggagcgac 2580

aagaaccggg gcaagagcga caacgtgccc tccgaagagg tcgtgaagaa gatgaagaac 2640

tactggcgcc agctgctgaa tgccaagctg attacccaga ggaagttcga caatctgacc 2700

aaggccgaga gaggcggcct gagcgaactg gataaggccg gcttcattaa gcggcagctg 2760

gtggaaaccc ggcagatcac aaagcacgtg gcacagatcc tggactcccg gatgaacact 2820

aagtacgacg agaacgacaa actgatccgg gaagtgaaag tgatcaccct gaagtccaag 2880

ctggtgtccg acttcagaaa ggatttccag ttttacaaag tgcgcgagat caacaactac 2940

caccacgccc acgacgccta cctgaacgcc gtcgtgggaa ccgccctgat caaaaagtac 3000

cctaagctgg aaagcgagtt cgtgtacggc gattacaagg tgtacgacgt gcggaagatg 3060

atcgccaaga gcgagcagga aatcggcaag gctaccgcca agtacttctt ctacagcaac 3120

atcatgaact ttttcaagac cgagatcaca ctggccaacg gcgagatcag aaagcggcct 3180

ctgatcgaga caaacggcga aaccggggag atcgtgtggg ataagggccg ggattttgcc 3240

acagtgcgga aagtgctgtc catgccccaa gtgaatatcg tgaaaaagac cgaggtgcag 3300

accggcggct tcagcaaaga gtctatcctg cccaagagga actccgacaa gctgatcgcc 3360

agaaagaagg attgggaccc taagaagtac ggcggctttg acagccccac cgtggcctac 3420

tctgtgctgg tggtggccaa agtggaaaag ggcaagtcca agaaactgaa gagtgtgaaa 3480

gagctgctgg ggatcaccat catggaaaga agcagcttcg agaagaatcc catcgacttt 3540

ctggaagcca agggctacaa agaagtgaaa aaggacctga tcatcaagct gcctaagtac 3600

tccctgttcg agctggaaaa cggccggaag cggatgctgg cttctgccgg cgaactgcag 3660

aagggaaacg agctggccct gccctccaaa tatgtgaact tcctgtacct ggccagccac 3720

tatgagaagc tgaagggctc ccccgaggat aatgagcaga aacagctgtt tgtggaacag 3780

cacaagcact acctggacga gatcatcgag cagattagcg agttctccaa gcgcgtgatc 3840

ctggccgatg ccaacctgga caaggtgctg agcgcctaca acaagcaccg ggataagccc 3900

atcagagagc aggccgagaa tatcatccac ctgtttaccc tgaccaacct gggagcccct 3960

gccgccttca agtactttga caccaccatc gaccggaaga ggtacaccag caccaaagag 4020

gtgctggacg ccaccctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac 4080

ctgtctcagc tgggaggcga ccccaagaaa aagcgcaaag tg 4122

<210> 11
<211> 4764
<212> DNA
<213> Homo sapiens

<400> 11
gcgggcgggc ggtgcgatgt ccggagagga tggcccggcg gctggcccgg gggcggcggc 60

ggcggctgcc cgggagcggc gacgggagca gctgcggcag tgggggggcgc gggcgggcgc 120

cgagcctggc cccggagagc gccgcgcccg caccgtccgc ttcgagcgcg ccgccgagtt 180

cctggcggcc tgtgcgggcg gcgacctgga cgaggcgcgt ctgatgctgc gcgccgccga 240

ccctggcccc ggcgccgagc tcgaccccgc cgcgccgccg cccgcccgcg ccgtgctgga 300

ctccaccaac gccgacggta tcagcgccct gcaccaggtc agcgcccccc gcccggcgtc 360

tcccggggcc aggtccaccc tctgctgcgc cacctggggc atcctccttc cccgttgcca 420

gtctcgatcc gccccgtcgt tcctggccct gggctttgcc accctatgct gacaccccgt 480

cccagtcccc cttaccattc cccttcgacc accccacttc cgaattggag ccgcttcaac 540

tggccctggg cttagccact ctgtgctgac cactctgccc caggcctcct taccattccc 600

cttcgaccta ctctcttccg cattggagtc gctttaactg gccctggctt tggcagcctg 660

tgctgaccca tgcagtcctc cttaccatcc ctccctcgac ttcccctctt ccgatgttga 720

gcccctccag ccggtcctgg actttgtctc cttccctgcc ctgccctctc ctgaacctga 780

gccagctccc atagctcagt ctggtctatc tgcctggccc tggccattgt cactttgcgc 840

tgccctcctc tcgcccccga gtgcccttgc tgtgccgccg gaactctgcc ctctaacgct 900

gccgtctctc tcctgagtcc ggaccacttt gagctctact ggcttctgcg ccgcctctgg 960

cccactgttt ccccttccca ggcaggtcct gctttctctg acctgcattc tctccctgg 1020

gcctgtgccg ctttctgtct gcagcttgtg gcctgggtca cctctacggc tggcccagat 1080

ccttccctgc cgcctccttc aggttccgtc ttcctccact ccctcttccc cttgctctct 1140

gctgtgttgc tgcccaagga tgctctttcc ggagcacttc cttctcggcg ctgcaccacg 1200

tgatgtcctc tgagcggatc ctccccgtgt ctgggtcctc tccgggcatc tctcctccct 1260

cacccaaccc catgccgtct tcactcgctg ggttcccttt tccttctcct tctggggcct 1320

gtgccatctc tcgtttctta ggatggcctt ctccgacgga tgtctccctt gcgtcccgcc 1380

tccccttctt gtaggcctgc atcatcaccg tttttctgga caaccccaaa gtaccccgtc 1440

tccctggctt tagccacctc tccatcctct tgctttcttt gcctggacac cccgttctcc 1500

tgtggattcg ggtcacctct cactcctttc atttgggcag ctcccctacc ccccttacct 1560

ctctagtctg tgctagctct tccagccccc tgtcatggca tcttccaggg gtccgagagc 1620

tcagctagtc ttcttcctcc aacccgggcc cctatgtcca cttcaggaca gcatgtttgc 1680

tgcctccagg gatcctgtgt ccccgagctg ggaccacctt atattcccag ggccggttaa 1740

tgtggctctg gttctgggta cttttatctg tcccctccac cccacagtgg ggccactagg 1800

gacaggattg gtgacagaaa agccccatcc ttaggcctcc tccttcctag tctcctgata 1860

ttgggtctaa cccccacctc ctgttaggca gattccttat ctggtgacac acccccattt 1920

cctggagcca tctctctcct tgccagaacc tctaaggttt gcttacgatg gagccagaga 1980

ggatcctggg agggagagct tggcaggggg tgggagggaa ggggggggatg cgtgacctgc 2040

ccggttctca gtggccaccc tgcgctaccc tctcccagaa cctgagctgc tctgacgcgg 2100

ccgtctggtg cgtttcactg atcctggtgc tgcagcttcc ttacacttcc caagaggaga 2160

agcagtttgg aaaaacaaaa tcagaataag ttggtcctga gttctaactt tggctcttca 2220

cctttctagt ccccaattta tattgttcct ccgtgcgtca gttttacctg tgagataagg 2280

ccagtagcca gccccgtcct ggcagggctg tggtgaggag gggggtgtcc gtgtggaaaa 2340

ctccctttgt gagaatggtg cgtcctaggt gttcaccagg tcgtggccgc ctctactccc 2400

tttctctttc tccatccttc tttccttaaa gagtccccag tgctatctgg gacatattcc 2460

tccgcccaga gcagggtccc gcttccctaa ggccctgctc tgggcttctg ggtttgagtc 2520

cttggcaagc ccaggagagg cgctcaggct tccctgtccc ccttcctcgt ccaccatctc 2580

atgcccctgg ctctcctgcc ccttccctac aggggttcct ggctctgctc ttcagactga 2640

gccccgttcc cctgcatccc cgttcccctg catccccctt cccctgcatc ccccagaggc 2700

cccaggccac ctacttggcc tggaccccac gagaggccac cccagccctg tctaccaggc 2760

tgccttttgg gtggattctc ctccaactgt ggggtgactg cttggcaaac tcactcttcg 2820

gggtatccca ggaggcctgg agcattgggg tgggctgggg ttcagagagg agggattccc 2880

ttctcaggtt acgtggccaa gaagcagggg agctgggttt gggtcaggtc tgggtgtggg 2940

gtgaccagct tatgctgttt gcccaggaca gcctagtttt agcactgaaa ccctcagtcc 3000

taggaaaaca gggatggttg gtcactgtct ctgggtgact cttgattccc ggccagtttc 3060

tccacctggg gctgtgtttc tcgtcctgca tccttctcca ggcaggtccc caagcatcgc 3120

ccccctgctg tggctgttcc caagttctta gggtacccca cgtgggttta tcaaccactt 3180

ggtgaggctg gtaccctgcc cccattcctg caccccaatt gccttagtgg ctaggggttt 3240

gggggctaga gtaggagggg ctggagccag gattcttagg gctgaacaga gaagagctgg 3300

gggcctgggc tcctgggttt gagagaggag gggctggggc ctggactcct gggtccgagg 3360

gaggaggggc tggggcctgg actcctgggt ctgagggtgg agggactggg ggcctggact 3420

cctgggtccg agggaggagg ggctggggcc tggactcgtg ggtctgaggg aggaggggct 3480

gggggcctgg acttctgggt cttagggagg cggggctggg cctggacccc tgggtctgaa 3540

tggggagagg ctgggggcct ggactccttc atctgagggc ggaagggctg gggcctggcc 3600

tcctgggttg aatggggagg ggttgggcct ggactctgga gtccctggtg cccaggcctc 3660

aggcatcttt cacagggatg cctgtactgg gcaggtcctt gaaagggaaa ggcccattgc 3720

tctccttgcc cccctcccct atcgccatga caactgggtg gaaataaacg agccgagttc 3780

atcccgttcc cagggcacgt gcggcccctt cacagcccga gtttccatga cctcatgctc 3840

ttggccctcg tagctccctc ccgcctcctc cagatgggca gctttggaga ggtgagggac 3900

ttgggggta atttatcccg tggatctagg agtttagctt cactccttcc tcagctccag 3960

ttcaggtccc ggagcccacc cagtgtccac aaggcctggg gcaagtccct cctccgaccc 4020

cctggacttc ggcttttgtc cccccaagtt ttggacccct aagggaagaa tgagaaacgg 4080

tggcccgtgt cagcccctgg ctgcagggcc ccgtgcagag ggggcctcag tgaactggag 4140

tgtgacagcc tggggcccag gcacacaggt gtgcagctgt ctcacccctc tgggagtccc 4200

gcccaggccc ctgagtctgt cccagcacag ggtggccttc ctccaccctg catagccctg 4260

ggcccacggc ttcgttcctg cagagtatct gctggggtgg tttccgagct tgacccttgg 4320

aaggacctgg ctgggtttaa ggcaggaggg gctgggggcc aggactcctg gctctgaagg 4380

aggaggggct ggaacctctt ccctagtctg agcactggaa gcgccacctg tgggtggtga 4440

cgggggtttt gccgtgtcta acaggtacca tgtggggttc ccgcacccag atgagaagcc 4500

ccctcccttc cccgttcact tcctgtttgc agatagccag gagtcctttc gtggtttcca 4560

ctgagcactg aaggcctggc cggcctgacc actgggcaac caggcgtatc ttaaacagcc 4620

agtggccaga ggctgttggg tcattttccc cactgtccta gcaccgtgtc cctggatctg 4680

ttttcgtggc tccctctgga gtcccgactt gctgggacac cgtggctggg gtaggtgcgg 4740

ctgacggctg tttcccaccc ccag 4764

<210> 12
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 12
accccacagu ggggccacua guuuuagagc uaugcuguuu ug 42

<210> 13
<211> 86
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 13
ggaaccauuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa 60

aguggcaccg agucggugcu uuuuuu 86

<210> 14
<211> 62
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 14
accccacagu ggggccacua guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cg 62

<210> 15
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 15
ctgacctctt ctcttcctcc cacag 25

<210> 16
<211> 1009
<212> DNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 16
gccaccatgg actacaaaga cgatgacgac aaggtcgact ctagagctgc agagagcgac 60

gagagcggcc tgcccgccat ggagatcgag tgccgcatca ccggcaccct gaacggcgtg 120

gagttcgagc tggtgggcgg cggagagggc acccccgagc agggccgcat gaccaacaag 180

atgaagagca ccaaaggcgc cctgaccttc agcccctacc tgctgagcca cgtgatgggc 240

tacggcttct accacttcgg cacctacccc agcggctacg agaacccctt cctgcacgcc 300

atcaacaacg gcggctacac caacacccgc atcgagaagt acgaggacgg cggcgtgctg 360

cacgtgagct tcagctaccg ctacgaggcc ggccgcgtga tcggcgactt caaggtgatg 420

ggcaccggct tccccgagga cagcgtgatc ttcaccgaca agatcgtccg cagcaacgcc 480

accgtggagc acctgcaccc catgggcgat aacgatctgg atggcagctt cacccgcacc 540

ttcagcctgc gcgacggcgg ctactacagc tccgtggtgg acagccacat gcacttcaag 600

agcgccatcc accccagcat cctgcagaac gggggcccca tgttcgcctt ccgccgcgtg 660

gaggaggatc acagcaacac cgagctgggc atcgtggagt accagcacgc cttcaagacc 720

ccggatgcag atgccggtga agaatgaaga tctctgtgcc ttctagttgc cagccatctg 780

ttgtttgccc ctccccccgtg ccttccttga ccctggaagg tgccactccc actgtccttt 840

cctaataaaa tgaggaaatt gcatcgcatt gtctgagtag gtgtcattct attctggggg 900

gtggggtggg gcaggacagc aaggggggagg attgggaaga caatagcagg catgctgggg 960

atgcggtggg ctctatggac tcgaggttta aacgtcgacg cggccgcgt 1009

<210> 17
<211> 355
<212> PRT
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 17

Met Ser Gly Glu Asp Gly Pro Ala Ala Gly Pro Gly Ala Ala Ala Ala
1 5 10 15

Ala Ala Arg Glu Arg Arg Arg Glu Gln Leu Arg Gln Trp Gly Ala Arg
20 25 30

Ala Gly Ala Glu Pro Gly Pro Gly Glu Arg Arg Ala Arg Thr Val Arg
35 40 45

```
Phe Glu Arg Ala Ala Glu Phe Leu Ala Ala Cys Ala Gly Gly Asp Leu
    50                  55                  60

Asp Glu Ala Arg Leu Met Leu Arg Ala Ala Asp Pro Gly Pro Gly Ala
65                  70                  75                  80

Glu Leu Asp Pro Ala Ala Pro Pro Pro Ala Arg Ala Val Leu Asp Ser
                85                  90                  95

Thr Asn Ala Asp Gly Ile Ser Ala Leu His Gln Ala Thr Met Asp Tyr
            100                 105                 110

Lys Asp Asp Asp Asp Lys Val Asp Ser Arg Ala Ala Glu Ser Asp Glu
        115                 120                 125

Ser Gly Leu Pro Ala Met Glu Ile Glu Cys Arg Ile Thr Gly Thr Leu
    130                 135                 140

Asn Gly Val Glu Phe Glu Leu Val Gly Gly Gly Glu Gly Thr Pro Glu
145                 150                 155                 160

Gln Gly Arg Met Thr Asn Lys Met Lys Ser Thr Lys Gly Ala Leu Thr
                165                 170                 175

Phe Ser Pro Tyr Leu Leu Ser His Val Met Gly Tyr Gly Phe Tyr His
            180                 185                 190

Phe Gly Thr Tyr Pro Ser Gly Tyr Glu Asn Pro Phe Leu His Ala Ile
        195                 200                 205

Asn Asn Gly Gly Tyr Thr Asn Thr Arg Ile Glu Lys Tyr Glu Asp Gly
    210                 215                 220

Gly Val Leu His Val Ser Phe Ser Tyr Arg Tyr Glu Ala Gly Arg Val
225                 230                 235                 240

Ile Gly Asp Phe Lys Val Met Gly Thr Gly Phe Pro Glu Asp Ser Val
                245                 250                 255
```

Ile Phe Thr Asp Lys Ile Val Arg Ser Asn Ala Thr Val Glu His Leu
260 265 270

His Pro Met Gly Asp Asn Asp Leu Asp Gly Ser Phe Thr Arg Thr Phe
275 280 285

Ser Leu Arg Asp Gly Gly Tyr Tyr Ser Ser Val Val Asp Ser His Met
290 295 300

His Phe Lys Ser Ala Ile His Pro Ser Ile Leu Gln Asn Gly Gly Pro
305 310 315 320

Met Phe Ala Phe Arg Arg Val Glu Glu Asp His Ser Asn Thr Glu Leu
325 330 335

Gly Ile Val Glu Tyr Gln His Ala Phe Lys Thr Pro Asp Ala Asp Ala
340 345 350

Gly Glu Glu
355

<210> 18
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 18
ccactctgtg ctgaccactc t 21

<210> 19
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 19
gcggcactcg atctcca 17

<210> 20

<211> 711
<212> DNA
<213> Mus musculus

<400> 20
gagcggctgc ggggcgggtg caagcacgtt tccgacttga gttgcctcaa gaggggcgtg 60

ctgagccaga cctccatcgc gcactccggg gagtggaggg aaggagcgag ggctcagttg 120

ggctgttttg gaggcaggaa gcacttgctc tcccaaagtc gctctgagtt gttatcagta 180

agggagctgc agtggagtag gcggggagaa ggccgcaccc ttctccggag gggggagggg 240

agtgttgcaa tacctttctg ggagttctct gctgcctcct ggcttctgag gaccgccctg 300

ggcctgggag aatcccttcc ccctcttccc tcgtgatctg caactccagt ctttctagaa 360

gatgggcggg agtcttctgg gcaggcttaa aggctaacct ggtgtgtggg cgttgtcctg 420

caggggaatt gaacaggtgt aaaattggag ggacaagact tcccacagat tttcggtttt 480

gtcgggaagt tttttaatag gggcaaataa ggaaaatggg aggataggta gtcatctggg 540

gttttatgca gcaaaactac aggttattat tgcttgtgat ccgcctcgga gtattttcca 600

tcgaggtaga ttaaagacat gctcacccga gttttatact ctcctgcttg agatccttac 660

tacagtatga aattacagtg tcgcgagtta gactatgtaa gcagaatttt a 711

<210> 21
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 21
cuccagucuu ucuagaagau guuuuagagc uaugcuguuu ug 42

<210> 22
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 22
ugaacaggug uaaaauugga guuuuagagc uaugcuguuu ug 42

<210> 23
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 23
ugucgggaag uuuuuuaaua guuuuagagc uaugcuguuu ug 42

<210> 24
<211> 642
<212> DNA
<213> Rattus rattus

<400> 24
gggattcctc cttgagttgt ggcactgagg aacgtgctga acaagaccta cattgcactc 60

cagggagtgg atgaaggagt tggggctcag tcgggttgta ttggagacaa gaagcacttg 120

ctctccaaaa gtcggtttga gttatcatta agggagctgc agtggagtag gcggagaaaa 180

ggccgcaccc ttctcaggac gggggagggg agtgttgcaa tacctttctg ggagttctct 240

gctgcctcct gtcttctgag gaccgccctg ggcctggaag attcccttcc cccttcttcc 300

ctcgtgatct gcaactggag tctttctgga agataggcgg gagtcttctg ggcaggctta 360

aaggctaacc tggtgcgtgg ggcgttgtcc tgcagaggaa ttgaacaggt gtaaaattgg 420

aggggcaaga cttcccacag attttcgatt gtgttgttaa gtattgtaat aggggcaaat 480

aagggaaata gactaggcac tcacctgggg ttttatgcag caaaactaca ggttattatt 540

gcttgtgatc cgccctggag aatttttcac cgaggtagat tgaagacatg cccacccaaa 600

ttttaatatt cttccacttg cgatccttgc tacagtatga aa 642

<210> 25
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 25
agggggaagg gaaucuucca guuuuagagc uaugcuguuu ug 42

<210> 26
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 26
ucugcaacug gagucuuucu guuuuagagc uaugcuguuu ug 42

<210> 27
<211> 42
<212> RNA
<213> Artificial Sequence

<220>
<223> SYNTHESIZED

<400> 27
aggcgggagu cuucugggca guuuuagagc uaugcuguuu ug 42