

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

(19) World Intellectual Property

Organization

International Bureau

(43) International Publication Date

07 May 2020 (07.05.2020)



(10) International Publication Number

WO 2020/093025 A1

(51) International Patent Classification:

A01H 1/00 (2006.01) C12N 15/00 (2006.01)

C12N 5/00 (2006.01) C12N 15/85 (2006.01)

C12N 5/10 (2006.01)

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

(21) International Application Number:

PCT/US2019/059557

(22) International Filing Date:

01 November 2019 (01.11.2019)

Published:

— with international search report (Art. 21(3))

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/754,451 01 November 2018 (01.11.2018) US

62/754,382 01 November 2018 (01.11.2018) US

(71) Applicant: SYNTHEGO CORPORATION [US/US];

3696-A Haven Avenue, Redwood City, California 94063 (US).

(72) Inventor: MAURES, Travis; 220 A Olympian Way, Pacifica, California 94044 (US).

(74) Agent: ROHATGI, Kyanne R.; WILSON SONSINI GOODRICH & ROSATI, 650 Page Mill Road, Palo Alto, California 94304-1050 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available):

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,

(54) Title: METHODS FOR KNOCK-OUT OF A TARGET SEQUENCE THROUGH INTRODUCTION OF A PREMATURE STOP CODON

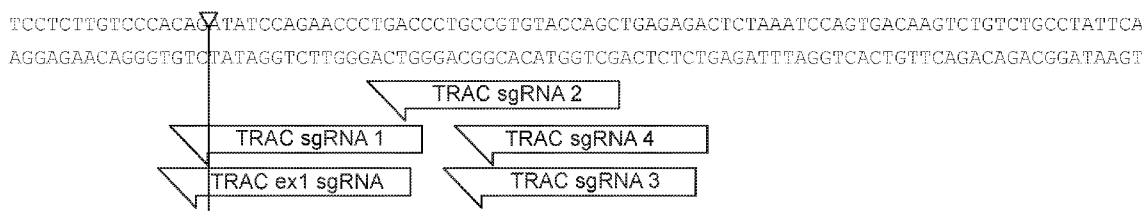


FIG. 1

(57) Abstract: The present disclosure provides a method for using homology directed repair to introduce an insert into a gene in order to induce knock-out of the gene. The insert can comprise a stop codon in each reading frame of the target sequence, and optionally, a restriction endonuclease recognition sequence.



WO 2020/093025 A1

## METHODS FOR KNOCK-OUT OF A TARGET SEQUENCE THROUGH INTRODUCTION OF A PREMATURE STOP CODON

### CROSS-REFERENCE

**[0001]** This application claims the benefit of U.S. Provisional Application No. 62/754,451, filed November 1, 2018, and U.S. Provisional Application No. 62/754,382, filed November 1, 2018, which are incorporated herein by reference in their entirety. The subject matter of this application relates to the U.S. Patent Application titled “Methods for Analyzing Nucleic Acid Sequences,” Attorney Docket No. 54108-718.201, filed November 1, 2019, which is incorporated herein by reference in its entirety.

### BACKGROUND

**[0002]** Engineered nuclease technologies designed to target and manipulate specific DNA sequences are rapidly being adopted as useful techniques for a number of different applications, including genetic manipulation of cells and whole organisms, targeted gene deletion, replacement and repair, and insertion of exogenous sequences (transgenes) into the genome. Genome editing techniques can include the use of zinc finger nucleases, transcription activator-like effector (TALE) nucleases, and clustered regularly interspaced short palindromic repeats CRISPR-associated (Cas) nucleases. Following a cut by one of the aforementioned nucleases at a targeted genomic site, an insert can be introduced into the targeted site through homology directed repair (HDR) mechanisms. This insert can be used to induce knockout of a gene, which can be useful for both research and therapeutic purposes.

### SUMMARY

**[0003]** Disclosed herein, in certain embodiments, are methods for introducing at least three stop codons in at least three different reading frames of a gene in a cell, comprising: (a) introducing into the cell: (i) a guide RNA (gRNA) capable of hybridizing to the gene; (ii) a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein; and (iii) a repair template comprising: (a) a first homology region comprising homology to a first portion of the gene; (b) an insert comprising: (i) a first stop codon in a first reading frame; (ii) a second stop codon in a second reading frame; and (iii) a third stop codon in a third reading frame, wherein the first reading frame, the second reading frame, and the third reading frame are different; and (c) a second homology region comprising homology to a second portion of the gene; wherein the repair template comprises a nucleotide sequence of less than 100 bases in length; and (b) incorporating by homology directed repair the insert at a cut site introduced by the Cas protein in

the gene, wherein the incorporating results in knock-out of the gene in the cell. In some embodiments, the incorporating is in an exon of the gene. In some embodiments, the first stop codon, the second stop codon, and the third stop codon are identical. In some embodiments, the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA.

**[0004]** In some embodiments, the insert comprises a nucleotide sequence of a length that is not an integer of 3. In some embodiments, the insert comprises a nucleotide sequence of a length of no more than 20 bases. In some embodiments, the insert comprises a nucleotide sequence of a length of 14 bases. In some embodiments, the insert has a GC content from 40% to 50%. In some embodiments, the insert comprises a nucleic acid sequence selected from any one of SEQ ID NO: 1- 18. In some embodiments, the insert comprises SEQ ID NO:1. In some embodiments, the insert comprises SEQ ID NO: 2.

**[0005]** In some embodiments, the insert comprises an identifier sequence. In some embodiments, the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof. In some embodiments, the restriction endonuclease recognition sequence is recognized by a restriction endonuclease. In some embodiments, the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHIII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some embodiments, the restriction endonuclease is BcII.

**[0006]** In some embodiments, the first homology region comprises a nucleotide sequence of a length of 30 bases or less. In some embodiments, the second homology region comprises a nucleotide sequence of a length of 30 bases or less. In some embodiments, the repair template comprises a nucleotide sequence of a length of 74 bases.

**[0007]** In some embodiments, the Cas protein is Cas9. In some embodiments, the gRNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). In some embodiments, the introducing comprises electroporation.

**[0008]** In some embodiments, the cell is from a cell line. In some embodiments, the cell line is selected from the group consisting of: Chinese hamster ovary (CHO) cells; HEK293 cells; Caco2

cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; DG44 cells; K-562 cells, U-937 cells; MC5 cells; MCF-7 cells; Saos-2 cells; P3 cells; Sf9 cells; IMR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HL-60 cells; HT-1080 cells; HCT-116 cells; Hu-h7 cells; Huvec cells; and Molt 4 cells. In some embodiments, the cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. In some embodiments, the human cell is a T cell. In some embodiments, the gene is a T cell receptor alpha (TRAC).

**[0009]** In some embodiments, the gene causes a condition in a human. In some embodiments, the condition is caused by a gain-of-function mutation in the gene. In some embodiments, the condition is selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy.

**[0010]** In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 10% of the population of cells. In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 40% of the population of cells. In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 50% of the population of cells. In some embodiments, introducing the nuclease and the repair template into a population of cells results in an incorporating of the insert at the cut site in about 10% - 60% of the population of cells.

**[0011]** Disclosed herein, in certain embodiments, are compositions for knock-out of a gene comprising a repair template comprising: (i) a first homology region comprising homology to a first portion of the gene; (ii) an insert comprising: (a) a first stop codon in a first reading frame; (b) a second stop codon in a second reading frame; and (c) a third stop codon in a third reading frame, wherein the first reading frame, the second reading frame, and the third reading frame are different; and (iii) a second homology region comprising homology to a second portion of the gene; wherein the repair template comprises a nucleotide sequence length of less than 100 bases. In some embodiments, the first stop codon, the second stop codon, and the third stop codon are identical. In some embodiments, the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA.

**[0012]** In some embodiments, the insert comprises a nucleotide sequence of a length that is not an integer of 3. In some embodiments, the insert comprises a nucleotide sequence of a length of

no more than 20 bases. In some embodiments, the insert comprises a nucleotide sequence of a length of 14 bases. In some embodiments, the insert has a GC content from 40% to 50%. In some embodiments, the insert comprises a nucleic acid sequence selected from any one of SEQ ID NO: 1-18. In some embodiments, the insert comprises SEQ ID NO: 1. In some embodiments, the insert comprises SEQ ID NO: 2.

**[0013]** In some embodiments, the insert comprises an identifier sequence. In some embodiments, the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof. In some embodiments, the restriction endonuclease recognition sequence is recognized by a restriction endonuclease. In some embodiments, the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfoI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some embodiments, the restriction endonuclease is BcII.

**[0014]** In some embodiments, the first homology region comprises a nucleotide sequence of a length of 30 bases or less. In some embodiments, the second homology region comprises a nucleotide sequence of a length of 30 bases or less. In some embodiments, the repair template comprises a nucleotide sequence of a length of 74 bases. In some embodiments, the composition is in the form of a powder or a solution. In some embodiments, the powder is a lyophilized powder.

**[0015]** Disclosed herein, in certain embodiments, are kits comprising a composition described herein and instructions. In some embodiments, the kit further comprises a guide RNA capable of hybridizing to a gene. In some embodiments, the guide RNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). In some embodiments, the kit further comprises a clustered, regularly interspersed, short palindromic repeats (CRISPR) associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein. In some embodiments, the Cas protein is Cas9.

**[0016]** Disclosed herein, in certain embodiments, are methods of treating a subject comprising a gain-of-function mutation in a gene, the method comprising administering to the subject a composition described herein. In some embodiments, the method further comprises

administering to the subject a guide RNA capable of hybridizing to the gene. In some embodiments, the guide RNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). In some embodiments, the method further comprises administering to the subject a clustered, regularly interspersed, short palindromic repeats (CRISPR) associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein. In some embodiments, the Cas protein is Cas9.

**[0017]** In some embodiments, the gain-of-function mutation results in a condition. In some embodiments, the condition is selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy. In some embodiments, the subject is a mammal. In some embodiments, the mammal is a human.

**[0018]** Disclosed herein, in certain embodiments, are engineered cells comprising a gene engineered to comprise an insert in a , wherein the insert comprises a first stop codon in a first reading frame, a second stop codon in a second reading frame, and a third stop codon in a third reading frame, wherein the insert is a non-naturally occurring sequence, and wherein the insert comprises a nucleic acid sequence of 14 bases in length. In some embodiments, the insert is in an exon of the gene. In some embodiments, the first stop codon, the second stop codon, and the third stop codon are identical. In some embodiments, the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA. In some embodiments, the insert has a GC content from 40% to 50%. In some embodiments, the insert comprises SEQ ID NO: 1. In some embodiments, the insert comprises SEQ ID NO: 2.

**[0019]** In some embodiments, the insert comprises an identifier sequence. In some embodiments, the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof. In some embodiments, the restriction endonuclease recognition sequence is recognized by a restriction endonuclease. In some embodiments, the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BclII, BglIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHIII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI,

SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some embodiments, the restriction endonuclease is BcII. In some embodiments, the engineered cell is a mammalian cell. In some embodiments, the mammalian cell is a human cell. Disclosed herein, in certain embodiments, are kits comprising the engineered cell described herein and instructions.

**[0020]** Disclosed herein, in certain embodiments, are methods for knock-out of a gene in a cell, comprising: (a) introducing into the cell (i) a guide RNA (gRNA) capable of hybridizing to the gene; (ii) a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein; and (iii) a repair template comprising (a) a first stop codon in a first reading frame, (b) a second stop codon in a second reading frame, (c) a third stop codon in a third reading frame, and (d) an identifier sequence; and (b) incorporating by homology directed repair the insert at a cut site introduced by the Cas protein in the gene, wherein the incorporating results in knock-out of the gene in the cell.

**[0021]** In some embodiments, the insert is incorporated into an exon of the gene. In some embodiments, the first stop codon, the second stop codon, and the third stop codon are identical. In some embodiments, the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA. In some embodiments, a region of the repair template comprising the first stop codon, the second stop codon, the third stop codon, and the target site for the restriction endonuclease comprises a nucleotide sequence of a length that is not an integer of 3. In some embodiments, the repair template comprises SEQ ID NO. 1.

**[0022]** In some embodiments, the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof. In some embodiments, the restriction endonuclease recognition sequence is recognized by a restriction endonuclease. In some embodiments, the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHIII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some embodiments, the restriction endonuclease is BcII.

**[0023]** In some embodiments, the repair template further comprises comprising a first homology region with homology to the gene on a first side of the cut site and a second homology

region with homology to the gene on a second side of the cut site. In some embodiments, the Cas protein is Cas9. In some embodiments, the gRNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). In some embodiments, the introducing comprises electroporation. In some embodiments, the cell is a human cell. In some embodiments, the human cell is a T cell. In some embodiments, the gene is a T cell receptor alpha (TCRA).

**[0024]** In some embodiments, the subject is a human. In some embodiments, the gene causes a condition in the human. In some embodiments, the condition is caused by a gain-of-function mutation in the gene. In some embodiments, the condition is selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy.

**[0025]** In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 10% of the population of cells. In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 40% of the population of cells. In some embodiments, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 50% of the population of cells. In some embodiments, introducing the nuclease and the repair template into a population of cells results in an incorporating of the insert at the cut site in about 10% - 60% of the population of cells.

**[0026]** Disclosed herein, in certain embodiments, are compositions for knock-out of a gene comprising: (a) a guide RNA (gRNA) capable of hybridizing to the gene; (b) a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein; and (c) a repair template comprising (a) a first stop codon in a first reading frame, (b) a second stop codon in a second reading frame, (c) a third stop codon in a third reading frame, and (d) an identifier sequence.

**[0027]** In some embodiments, the first stop codon, the second stop codon, and the third stop codon are identical. In some embodiments, the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA. In some embodiments, a region of the repair template comprising the first stop codon, the second stop codon, the third stop codon, and the target site for the restriction endonuclease comprises a nucleotide sequence of a length that is not an integer of 3. In some embodiments, the repair template comprises SEQ ID NO. 1. In some



embodiments, the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof. In some embodiments, the restriction endonuclease recognition sequence is recognized by a restriction endonuclease. In some embodiments, the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, BmeI580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHIII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmlI, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some embodiments, the restriction endonuclease is BcII. In some embodiments, the repair template further comprises comprising a first homology region with homology to the gene on a first side of the cut site and a second homology region with homology to the gene on a second side of the cut site. In some embodiments, the Cas protein is Cas9. In some embodiments, the gRNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA).

### INCORPORATION BY REFERENCE

[0028] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0029] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0030] **FIG. 1** illustrates location of five sgRNAs targeting Exon 1 of the human T cell receptor alpha (TRAC) gene. The black vertical line indicates the start of the TRAC gene.

[0031] **FIG. 2** illustrates the inferred sequences present in the edited population and their relative proportions as determined by a software to predict genomic edits of most likely genomic outcomes. The cut site is represented by the vertical dashed line and the wild-type sequence is marked by a “+” symbol on the far left. The KO-Score indicates the proportion of indels that

indicates a frameshift or are +21bp in length, and assumes all edits are in a coding region of a gene.

[0032] **FIG. 3** shows the inferred distribution of indels in the entire edited population of genomes. Each indel size represented can not necessarily occur in the same sequence. The percentages of different indel sizes in the cell population are not the same as KO scores determined by the software.

[0033] **FIG. 4** illustrates a discordance plot showing the level of alignment per base between the wild type (control) and the edited sample in the inference window (the region around the cut site), i.e. it shows the average amount of signal that disagrees with the reference sequence derived from the control trace file. Both lines should be close together before the cut site, with a typical CRISPR edit resulting in a jump in the discordance near the cut site and continuing to remain far apart after the cut site (representing a high level of sequence discordance).

[0034] **FIGS. 5A-5B** illustrate plots identifying CD3+ PAN T-cells. **FIG. 5A** shows a scatter plot of forward versus side scatter (FSC vs SSC) to identify T-cells. **FIG. 5B** shows the wild-type T-cells found to be CD3+ using an FITC conjugated anti-CD3 antibody.

[0035] **FIGS. 6A-6B** illustrate plots identifying CD3+ T-cells from the TCRA KO T-cell pool. **FIG. 6A** shows a scatter plot of forward versus side scatter (FSC vs SSC) to identify T-cells. **FIG. 6B** shows the cells from the KO pool found to be CD3+ using an FITC conjugated anti-CD3 antibody.

[0036] **FIGS. 7A-7C** show conditions and results for nucleofection experiments of a repair template into HEK293 cells. **FIG. 7A** shows nucleofection experimental setups. **FIG. 7B** shows Sanger sequencing analysis by a software predict genomic edits of most likely genetic outcomes, including knock-in % (KI%). **FIG. 7C** shows results from restriction enzyme cleavage analysis.

[0037] **FIGS. 8A-8C** show a schematic of insertion of a repair template comprising an insert into exon 6 of the MAST3 gene. **FIG. 8A** shows the location of the cut site (vertical line) and the location of the gRNA binding site. **FIG. 8B** shows the repair template. **FIG. 8C** shows exon 6 of the MAST3 gene following incorporation of the repair template. The insert of the repair template is illustrated between the two vertical lines.

[0038] **FIG. 9** shows a computer system that is programmed or otherwise configured to implement methods, or portions of the methods, provided herein.

### DETAILED DESCRIPTION

[0039] The generation of cells or organisms with a knock-out of a gene can provide a way to study the function of the targeted gene or can be used therapeutically to target gain-of-function

mutations. The disclosure presented herein describes methods and compositions for knock-out of a gene.

**[0040]** The terminology used herein is for the purpose of describing particular cases only and is not intended to be limiting. The below terms are discussed to illustrate meanings of the terms as used in this specification, in addition to the understanding of these terms by those of skill in the art. As used herein and in the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the context clearly dictates otherwise. It is further noted that the claims can be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as “solely,” “only,” and the like in connection with the recitation of claim elements, or use of a “negative” limitation.

**[0041]** Certain ranges are presented herein with numerical values being preceded by the term “about.” The term “about” can be used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating un-recited number can be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number. Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, can be encompassed within the methods and compositions described herein are. The upper and lower limits of these smaller ranges can independently be included in the smaller ranges and are also encompassed within the methods and compositions described herein, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits can also be included in the methods and compositions described herein.

**[0042]** The term “CRISPR/Cas,” as used herein, can refer to a ribonucleoprotein complex, e.g., a two part component ribonucleoprotein complex, with guide RNA (gRNA) and a CRISPR-associated (Cas) endonuclease. In some cases, CRISPR/Cas comprises more than two components. The term “CRISPR” can refer to the Clustered Regularly Interspaced Short Palindromic Repeats and the related system thereof. CRISPR can be used as an adaptive defense system that enables bacteria and archaea to detect and silence foreign nucleic acids (e.g., from viruses or plasmids). CRISPR can be adapted for use in a variety of cell types to allow for polynucleotide editing in a sequence-specific manner. In some cases, one or more elements of a CRISPR system can be derived from a type I, type II, or type III CRISPR system. In the CRISPR

type II system, the guide RNA can interact with Cas and direct the nuclease activity of the Cas enzyme to a target region. The target region can comprise a “protospacer” and a “protospacer adjacent motif” (PAM), and both domains can be used for a Cas enzyme mediated activity (e.g., cleavage). The protospacer can be referred to as a target site (or a genomic target site). The gRNA can pair with (or hybridize) the opposite strand of the protospacer (binding site) to direct the Cas enzyme to the target region. The PAM site can refer to a short sequence recognized by the Cas enzyme and, in some cases, required for the Cas enzyme activity. The sequence and number of nucleotides for the PAM site can differ depending on the type of the Cas enzyme.

**[0043]** The term “Cas,” as used herein, generally refers to a wild type Cas protein, a fragment thereof, or a mutant or variant thereof. The term “Cas,” “enzyme Cas,” “enzyme CRISPR,” “protein CRISPR,” or “protein Cas” can be used interchangeably throughout the present disclosure.

**[0044]** A Cas protein can comprise a protein of or derived from a CRISPR/Cas type I, type II, or type III system, which has an RNA-guided polynucleotide-binding or nuclease activity. Examples of suitable Cas proteins include CasX, Cas3, Cas4, Cas5, Cas5e (or CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (also known as Csn1 and Csx12), 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, Cu1966, homologues thereof, and modified versions thereof. In some cases, a Cas protein can comprise a protein of or derived from a CRISPR/Cas type V or type VI system, such as Cpf1, C2c1, C2c2, homologues thereof, and modified versions thereof. In some cases, a Cas protein can be a catalytically dead or inactive Cas (dCas). In some cases, a Cas protein can have reduced or minimal nuclease activity (i.e., deactivated Cas, or dCas). In some cases, a Cas protein can be operatively coupled to one or more additional proteins, such as a nucleic acid polymerase. In an example, a Cas protein can be a dCas that is fused to a reverse transcriptase.

**[0045]** The term “guide RNA” or “gRNA,” as used herein, can refer to an RNA molecule (or a group of RNA molecules collectively) that can bind to a Cas protein and aid in targeting the Cas protein to a specific location within a target polynucleotide (e.g., a DNA). A guide RNA can comprise a CRISPR RNA (crRNA) segment and a trans-activating crRNA (tracrRNA) segment. The term “crRNA” or “crRNA segment,” as used herein, can refer to an RNA molecule or portion thereof that includes a polynucleotide-targeting guide sequence, a stem sequence, and, optionally, a 5'-overhang sequence. The term “tracrRNA” or “tracrRNA segment,” can refer to

an RNA molecule or portion thereof that includes a protein-binding segment (e.g., the protein-binding segment can be capable of interacting with a CRISPR-associated protein, such as a Cas9). The term “guide RNA” can encompass a single guide RNA (sgRNA), where the crRNA segment and the tracrRNA segment are located in the same RNA molecule. The term “guide RNA” can also encompass, collectively, a group of two or more RNA molecules, where the crRNA segment and the tracrRNA segment can be located in separate RNA molecules. In some cases, the gRNA comprises nucleotides other than ribonucleotides.

**[0046]** The term “codon,” as used herein, can refer to any group of three consecutive nucleotide bases in a given messenger RNA molecule, or coding strand of DNA that specifies a particular amino-acid. In some cases, the codon can be a starting or stopping signal for translation. The term codon also can refer to base triplets in a DNA strand.

**[0047]** The term “nucleic acid editing tool,” “gene editing tool,” or “genome editing tool,” as used interchangeably herein, can refer to a nuclease or nuclease system that can induce a cleavage (e.g., a targeted cleavage, targeted break, or targeted cut) in a nucleic acid, e.g., a gene. The nuclease or nuclease system can comprise a polypeptide (e.g., an enzyme) and/or a ribozyme. Thus, the nucleic acid editing tool can generate a cleavage site, break site, or cut site in the nucleic acid, e.g., gene. In an example, the cleavage can be a targeted single strand break (SSB). In another example, the cleavage can be a targeted double-strand break (DSB). The nucleic acid editing tool can be, e.g., an endonuclease. Examples of the nucleic acid editing tool include meganucleases (MN), zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALEN), clustered regularly interspaced short palindromic repeat (CRISPR)-associated nucleases (e.g., CRISPR/Cas), one or more components of an RNA-induced silencing complex (RISC) (e.g., Argonaute), homologues thereof, and modified versions thereof.

**[0048]** The nucleic acid editing tool can be useful in any in vitro or in vivo application in which it is desirable to modify a nucleic acid (e.g., DNA) in a site-specific (targeted) way, for example gene knock-out (KO), gene knock-in (KI), gene editing, gene tagging, etc., as used in, for example, gene therapy. Examples of uses of nucleic acid editing can include gene therapies for antiviral, antipathogenic, and anticancer therapeutic; the production of genetically modified organisms in agriculture; the production (e.g., large scale production) of proteins by cells for therapeutic, diagnostic, or research purposes; the induction of induced pluripotent stem cells (iPS cells or iPSCs); and the targeting of genes of pathogens for deletion or replacement.

**[0049]** The term “gene,” as used herein, can refer to a nucleotide sequence that acts as a physical or functional unit of heredity. In some cases, a gene encodes a polypeptide (e.g.,

protein). In some cases, a gene does not encode a polypeptide. The gene can comprise DNA, RNA, or other nucleotides. A gene can comprise from about 100 to about 2 million bases.

**[0050]** The term “polynucleotide” or “nucleic acid,” as used interchangeably herein, can refer to a polymeric form of nucleotides (e.g., ribonucleotides or deoxyribonucleotides) of any length. Thus, this term includes single-, double-, or multi-stranded DNA or RNA, genomic DNA, complementary DNA (cDNA), guide RNA (gRNA), messenger RNA (mRNA), DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The term “oligonucleotide,” as used herein, can refer to a polynucleotide of between about 5 and about 100 nucleotides of single- or double-stranded DNA or RNA. The length of a nucleic acid can be referred to in reference of the number of bases in the nucleic acid sequence. For example, a sequence of 100 nucleotides can be referred to as being 100 bases in length. However, for the purposes of this disclosure, there can be no upper limit to the length of an oligonucleotide. In some cases, oligonucleotides can be known as “oligomers” or “oligos” and can be isolated from genes, or chemically synthesized by methods known in the art. The terms “polynucleotide” and “nucleic acid” can include single-stranded (such as sense or antisense) and double-stranded polynucleotides. Examples of nucleotides for DNA can include cytosine (C), guanine (G), adenine (A), thymine (T), or modifications thereof. Examples of nucleotides for RNA can include C, G, A, uracil (U), or modifications thereof.

**[0051]** The term “hybridization” or “hybridizing,” as used herein, can refer to a process where completely or partially complementary polynucleotide strands come together under suitable hybridization conditions to form a double-stranded structure or region in which the two constituent strands are joined by hydrogen bonds. The term “partial hybridization,” as used herein, can refer to a double-stranded structure or region containing one or more bulges or mismatches.

**[0052]** The term “cleavage” or “cleaving,” as used herein, can refer to breaking of a covalent phosphodiester linkage in the ribosyl phosphodiester backbone of a polynucleotide. The term “cleavage” or “cleaving” can encompass both single-stranded breaks and double-stranded breaks. In some cases, a cleavage can result in the production of either blunt ends or staggered (or sticky) ends.

**[0053]** The term “Sanger sequencing,” as used here, can refer to a method of DNA sequencing based on selective incorporation of labeled chain-terminating dideoxynucleotides (ddNTPs) during in vitro DNA replication. Sequence information can be obtained using cycles of template denaturation, primer annealing and primer extension. Each round of primer extension can be

stochastically terminated by incorporation of labeled ddNTPs. In the resulting mixture of end-labeled extension products, the label on the terminating ddNTP of any given fragment can correspond to the nucleotide identity of its terminal position. Sequence can be determined by high-resolution electrophoretic separation of the single-stranded, end-labeled extension products in a capillary-based polymer gel (e.g., capillary electrophoresis). Laser excitation of fluorescent labels as fragments of discrete lengths exit the capillary, coupled to four-color detection of emission spectra, can provide the readout that is represented in a Sanger sequencing “trace.” Software can translate these traces into DNA sequences, while also generating error probabilities for one or more base-calls (e.g., one or more nucleobase calls). In some cases, Sanger sequencing can sequence a single DNA fragment; thus, the Sanger sequencing trace can have a single DNA fragment trace.

**[0054]** The term “regression analysis,” as used herein, can refer to a method of using one or more mathematical/statistical analyses to model a relationship between a dependent variable and one or more independent variables, in which the dependent variable is dependent on the one or more independent variants. A regression analysis can utilize a linear model or a non-linear model. Examples of a regression analysis include, but are not limited to, linear regression, polynomial regression, logistic regression, quantile regression, ridge regression, least absolute shrinkage and selection operator (Lasso) regression, elastic net regression, principal components regression, partial least squares (PLS) regression, support vector regression, ordinal regression, Poisson regression, negative binomial regression, quasi Poisson regression, Cox regression, and Tobit regression. Other examples of a regression analysis can include Bayesian methods, percentage regression, least absolute deviations, nonparametric regression, and distance metric learning. In some cases, the regression analysis can utilize regularization, which generally refers to a process to solve over-fitting problem of the statistical model by constraining one or more model parameters. Examples of the regularization-based regression analysis include ridge regression and Lasso regression.

**[0055]** In some cases, the CRISPR/Cas activity can be useful in any *in vitro* or *in vivo* application in which it is desirable to modify DNA in a site-specific (targeted) way, for example gene knock-out (KO), gene knock-in (KI), gene editing, gene tagging, etc., as used in, for example, gene therapy. Examples of gene therapy include treating a disease or as an antiviral, antipathogenic, or anticancer therapeutic; the production of genetically modified organisms in agriculture; the large scale production of proteins by cells for therapeutic, diagnostic, or research purposes; the induction of induced pluripotent stem cells (iPS cells or iPSCs); and the targeting of genes of pathogens for deletion or replacement. In some cases, the Cas can be a catalytically

dead or inactive Cas (dCas), and the resulting CRISPR/dCas system can be useful for sequence-specific repression (CRISPR interference) or activation (CRISPR activation) of gene expression.

**[0056]** As used herein, a “knock-out” of a gene can describe disruption in the sequence of the gene that results in a decrease or elimination of a function of the gene. The knock-out can result in expression of the gene that is undetectable or insignificant. The disruption can be due to introduction of a frameshift mutation or a premature stop codon into the sequence of the gene, which can occur through the use gene editing technologies, such as described herein.

**[0057]** The term “premature stop codon,” as used herein, can refer to a codon generated by a mutation in a gene that results in a truncated, incomplete, and/or nonfunctional polypeptide product. In some cases, the premature stop codon can be a nonsense mutation. The term “nonsense mutation,” as used herein, can refer to a point mutation that changes a codon corresponding to an amino acid to a stop codon. In some cases, the premature stop codon cannot be a nonsense mutation, but rather a new codon that is within an insertion mutation.

**[0058]** A “subject” disclosed herein includes any living organism. Thus, in some embodiments, subjects are mammals, avians, reptiles, amphibians, fish, plants, fungi, or bacteria. Mammalian subjects include but are not limited to humans, non-human primates (e.g., gorilla, monkey, baboon, and chimpanzee, etc.), dogs, cats, goats, horses, pigs, cattle, sheep, and the like, and laboratory animals (e.g., rats, guinea pigs, mice, gerbils, hamsters, and the like). Avian subjects include but are not limited to chickens, ducks, turkeys, geese, quail, pheasants, and birds kept as pets. In some embodiments, suitable subjects include both males and females and subjects of any age, including embryonic (e.g., *in-utero* or *in-ovo*), infant, juvenile, adolescent, adult and geriatric subjects. In some embodiments, a subject is a human.

**[0059]** “Treating” or “treatment” can refer to both therapeutic treatment and prophylactic or preventative measures, wherein the object is to prevent or slow down (lessen) a targeted pathologic condition or disorder. Those in need of treatment can include those already with the disorder, as well as those prone to have the disorder, or those in whom the disorder is to be prevented. For example, a subject can be successfully “treated” for a disease caused by a gain-of-function mutation, if, after receiving a therapeutic amount of a composition according to the methods of the present disclosure, the subject shows observable and/or measurable reduction in or absence of one or more of the following: relief to some extent of one or more of the symptoms associated with the specific disease; reduced morbidity and/or mortality, and improvement in quality of life issues.

**[0060]** Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the methods and



compositions described herein belong. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the methods and compositions described herein, representative illustrative methods and materials are now described.

### **Compositions for knock-out of a gene**

**[0061]** Disclosed herein, in certain embodiments, are compositions for knock-out of a gene. The compositions described herein can comprise a guide RNA (gRNA), a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein, a repair template, or any combination thereof. The gRNA can hybridize to the gene at a gRNA binding site. The repair template can comprise a nucleic acid sequence of no more than 100 bases in length and an insert comprising at least a first stop codon in a first reading frame. The insert can further comprise a second stop codon in a second reading frame. The insert can further comprise a third stop codon in a third reading frame. The first reading frame, the second reading frame, and the third reading frame can be different reading frames. The first stop codon, second stop codon, and third stop codon can be premature stop codons. In some embodiments, the insert further comprises a fourth stop codon in the first reading frame, a fifth stop codon in the second reading frame, a sixth stop codon in the third reading frame, or a combination thereof. In some embodiments, the insert does not comprise a stop codon.

**[0062]** The insert can be incorporated into one or both alleles of the gene. In some cases, knock-down of the gene results in reduced expression of the gene relative to a gene that has not been knocked-down. In some cases, knock-down of the gene results in no expression of the gene.

**[0063]** The composition can be in the form of a powder. The powder can be a lyophilized powder. The composition can be in the form of a solution. The composition can be kept at room temperature. The composition can be frozen. The composition can be frozen at about -20°C or -80°C.

### **Targeted Endonuclease**

**[0064]** The methods of the present disclosure can comprise introducing into the cell a targeted endonuclease. The targeted endonuclease can be a zinc finger nuclease, a transcription activator-like effector nuclease (TALEN), or a clustered regularly interspaced short palindromic repeats CRISPR-associated (Cas) nuclease, also referred to herein as a Cas protein or a Cas enzyme. The targeted endonuclease can be a deactivated endonuclease. The targeted endonuclease can

produce a break in the gene, at a location in the gene referred to as a cut site. The break can be a double stranded break. The break can be a single stranded break. The break can be a blunt break. The break can produce overhangs at the cut site. The cut site can be located in an exon of the gene. The targeted endonuclease can be directed to the cut site with the help of a guide RNA, as further discussed herein.

**[0065]** The targeted endonuclease can be a clustered regularly interspaced short palindromic repeats CRISPR-associated (Cas) nuclease. The Cas protein in the CRISPR/Cas system can be a Class 1 or Class 2 Cas. The Cas protein can be a Class 2 Cas. The Cas protein in the CRISPR/Cas system can be type I, type II, type III, type IV, or type V Cas. In some cases, the CRISPR Cas system is a type II CRISPR/Cas system. In some cases, the CRISPR Cas system is a type V CRISPR/Cas system. The type II Cas protein can be Cas9. The type V Cas protein can be Cpf1. In some cases, the CRISPR/Cas system mediated edit occurs at a cut site in an oligonucleotide sequence. The mediated edit can be a double stranded break in a DNA. The double stranded break can produce a blunt cut, such as when Cas9 is used, or a cut with a 5' overhang, such as when Cpf1 is used.

**[0066]** In some cases, the cut site can be adjacent to a PAM site for a nuclease. The nuclease can be selected from the group consisting of: Cas9, C2c1, C2c3, Cpf1, Cas13b, or Cas13c. In one example, the Cas nuclease is Cas9 from *Streptococcus pyogenes* (SpCas9), and the plurality of cut sites include all nucleotide sequences adjacent to the PAM site of SpCas9 (NGG, where "N" is any nucleotide). In another example, the Cas nuclease is Cas9 from *Neisseria meningitidis* (NmCas9), and the plurality of cut sites include all nucleotide sequences adjacent to the PAM site of NmCas9 (GATT). To be directed to such cut sites, one or more of the nucleases (e.g., Cas9, C2c1, C2c3, Cpf1, Cas13b, Cas13c, etc.) can be coupled to at least one gRNA. The at least one gRNA can be designed to hybridize at least one binding site that is an opposite strand of the cut site (protospacer).

### **Guide RNA**

**[0067]** The methods of the present disclosure can comprise introducing into the cell a guide RNA (gRNA). The guide RNA can be a single guide RNA (sgRNA). In some cases, the sgRNA can be a single polynucleotide chain. The sgRNA can comprise a hybridizing polynucleotide sequence and a second polynucleotide sequence. The hybridizing polynucleotide sequence can hybridize a portion of the target expressible gene, also referred to herein as the gRNA binding site. The gRNA binding site can comprise the sequence which will be cut by the Cas protein (the cut site). The hybridizing polynucleotide sequence of the sgRNA can range between 17 to 23

nucleotides. The hybridizing polynucleotide sequence of the sgRNA can be at least 17, 18, 19, 20, 21, 22, 23, or more nucleotides. The hybridizing polynucleotide sequence of the sgRNA can be at most 23, 22, 21, 20, 19, 18, 17, or less nucleotides. In an example, the hybridizing polynucleotide sequence of the gRNA is 20 nucleotides. The second polynucleotide sequence of the single polynucleotide chain sgRNA can interact (bind) with the Cas enzyme. The second polynucleotide sequence can be about 80 nucleotides. The second polynucleotide sequence can be 80 nucleotides. The second polynucleotide sequence can be at least 80, or more nucleotides. The second polynucleotide sequence can be at most 80, or less nucleotides. Overall, the single polynucleotide chain sgRNA can range between 97 to 103 nucleotides. The single polynucleotide chain sgRNA can be at least 97, 98, 99, 100, 101, 102, 103, or more nucleotides. The single polynucleotide chain sgRNA can be at most 103, 102, 101, 100, 99, 98, 97, or less nucleotides. In an example, the single polynucleotide chain sgRNA can be 100 nucleotides. In some cases, the hybridizing polynucleotide sequence and the second polynucleotide sequence are joined by a linker. In some embodiments, the hybridizing polynucleotide is a crRNA and the second polynucleotide sequence is a tracrRNA.

**[0068]** In some cases, the guide RNA can be a complex (e.g., via hydrogen bonds) of a CRISPR RNA (crRNA) segment and a trans-activating crRNA (tracrRNA) segment. The crRNA can comprise a hybridizing polynucleotide sequence and a tracrRNA-binding polynucleotide sequence. The hybridizing polynucleotide sequence can hybridize the portion of the gene (e.g., the selected exon of the selected transcript of the plurality of transcripts of the gene). The hybridizing polynucleotide sequence of the crRNA can range between 17 to 23 nucleotides. The hybridizing polynucleotide sequence of the crRNA can be at least 17, 18, 19, 20, 21, 22, 23, or more nucleotides. The hybridizing polynucleotide sequence of the crRNA can be at most 23, 22, 21, 20, 19, 18, 17, or less nucleotides. In an example, the hybridizing polynucleotide sequence of the crRNA is 20 nucleotides. The tracrRNA-binding polynucleotide sequence of the crRNA can be 22 nucleotides. The tracrRNA-binding polynucleotide sequence of the crRNA can be at least 22, or more nucleotides. The tracrRNA-binding polynucleotide sequence of the crRNA can be at most 22, or less nucleotides. Overall, the crRNA can range between 39 to 45 nucleotides. The crRNA can be at least 39, 40, 41, 42, 43, 44, 45, or more nucleotides. The crRNA can be at most 45, 44, 43, 42, 41, 40, 39, or less nucleotides. The tracrRNA can be 72 nucleotides. The tracrRNA can be at least 72, or more nucleotides. The tracrRNA can be at most 72, or less nucleotides. In an example, the hybridizing polynucleotide sequence of the crRNA is 20 nucleotides, the crRNA is 43 nucleotides, and the respective tracrRNA is 72 nucleotides. In one example, the sgRNA can be administered to the cell in a single vector. In another example, the

crRNA and the tracrRNA can be administered to the cell separately, for example delivery of the crRNA in a first vector and delivery of the tracrRNA in a second vector.

**[0069]** In some cases, one gRNA is introduced into the cell. In other cases, more than one gRNA is introduced into the cell. For example, two, three, four, five, six, seven, eight, nine, ten, or more than ten gRNAs can be introduced into the cell. In some instances, at least two of the more than one gRNAs target different target sequences. Different target sequences can be different locations within an identical gene or in different genes.

**[0070]** In some cases, the gRNA comprises a modification of at least one nucleotide of the gRNA. The modification can be a modification which (a) improves target specificity; (b) reduced effective concentration of the CRISPR/Cas complex; (c) improves stability of the gRNA (e.g., resistance to ribonuclease (RNases) and/or deoxyribonucleases (DNases)); (d) decreases immunogenicity, or (e) any combination thereof. Examples of the modification of the at least one nucleotide can include: (a) end modifications, including 5' end modifications or 3' end modifications; (b) nucleobase (or "base") modifications, including replacement or removal of bases; (c) sugar modifications, including modifications at the 2', 3', and/or 4' positions; and (d) backbone modifications, including modification or replacement of the phosphodiester linkages. In one example, the modification is a 2'-O-methyl nucleotide.

**[0071]** In some cases, a nucleotide sugar modification incorporated into the guide RNA is selected from the group consisting of 2'-O-C1-4alkyl such as 2'-O-methyl (2'-OMe), 2'-deoxy (2'-H), 2'-O-C1-3alkyl-O-C1-3alkyl such as 2'-methoxyethyl ("2'-MOE"), 2'-fluoro ("2'-F"), 2'-amino ("2'-NH<sub>2</sub>"), 2'-arabinosyl ("2'-arabino") nucleotide, 2'-F-arabinosyl ("2'-F-arabino") nucleotide, 2'-locked nucleic acid ("LNA") nucleotide, 2'-unlocked nucleic acid ("ULNA") nucleotide, a sugar in L form ("L-sugar"), and 4'-thioribosyl nucleotide. In some cases, an internucleotide linkage modification incorporated into the guide RNA is selected from the group consisting of: phosphorothioate "P(S)" (P(S)), phosphonocarboxylate (P(CH<sub>2</sub>)<sub>n</sub>COOR) such as phosphonoacetate "PACE" (P(CH<sub>2</sub>COO<sup>-</sup>)), thiophosphonocarboxylate ((S)P(CH<sub>2</sub>)<sub>n</sub>COOR) such as thiophosphonoacetate "thioPACE" ((S)P(CH<sub>2</sub>)<sub>n</sub>COO<sup>-</sup>), alkylphosphonate (P(C1-3alkyl) such as methylphosphonate -P(CH<sub>3</sub>), boranophosphonate (P(BH<sub>3</sub>)), and phosphorodithioate (P(S)<sub>2</sub>).

**[0072]** In some cases, a nucleobase ("base") modification incorporated into the guide RNA is selected from the group consisting of: 2-thiouracil ("2-thioU"), 2-thiocytosine ("2-thioC"), 4-thiouracil ("4-thioU"), 6-thioguanine ("6-thioG"), 2-aminoadenine ("2-aminoA"), 2-aminopurine, pseudouracil, hypoxanthine, 7-deazaguanine, 7-deaza-8-azaguanine, 7-deazaadenine, 7-deaza-8-azaadenine, 5-methylcytosine ("5-methylC"), 5-methyluracil ("5-

methylU”), 5-hydroxymethylcytosine, 5-hydroxymethyluracil, 5,6-dehydrouracil, 5-propynylcytosine, 5-propynyluracil, 5-ethynylcytosine, 5-ethynyluracil, 5-allyluracil (“5-allylU”), 5-allylcytosine (“5-allylC”), 5-aminoallyluracil (“5-aminoallylU”), 5-aminoallylcytosine (“5-aminoallylC”), an abasic nucleotide, Z base, P base, Unstructured Nucleic Acid (“UNA”), isoguanine (“isoG”), isocytosine (“isoC”), 5-methyl-2-pyrimidine, x(A,G,C,T), and y(A,G,C,T).

**[0073]** In some cases, one or more isotopic modifications are introduced on the nucleotide sugar, the nucleobase, the phosphodiester linkage and/or the nucleotide phosphates. Such modifications include nucleotides comprising one or more <sup>15</sup>N, <sup>13</sup>C, <sup>14</sup>C, Deuterium, <sup>3</sup>H, <sup>32</sup>P, <sup>125</sup>I, <sup>131</sup>I atoms or other atoms or elements used as tracers.

**[0074]** In some cases, an “end” modification incorporated into the guide RNA is selected from the group consisting of: PEG (polyethyleneglycol), hydrocarbon linkers (including: heteroatom (O,S,N)-substituted hydrocarbon spacers; halo-substituted hydrocarbon spacers; keto-, carboxyl-, amido-, thionyl-, carbamoyl-, thionocarbamaoyl-containing hydrocarbon spacers), spermine linkers, dyes including fluorescent dyes (for example fluoresceins, rhodamines, cyanines) attached to linkers such as for example 6-fluorescein-hexyl, quenchers (for example dabcy1, BHQ) and other labels (for example biotin, digoxigenin, acridine, streptavidin, avidin, peptides and/or proteins). In some cases, an “end” modification comprises a conjugation (or ligation) of the guide RNA to another molecule comprising an oligonucleotide (comprising deoxynucleotides and/or ribonucleotides), a peptide, a protein, a sugar, an oligosaccharide, a steroid, a lipid, a folic acid, a vitamin and/or other molecule. In some cases, an “end” modification incorporated into the guide RNA is located internally in the guide RNA sequence via a linker such as for example 2-(4-butylamidofluorescein)propane-1,3-diol bis(phosphodiester) linker, which is incorporated as a phosphodiester linkage and can be incorporated anywhere between two nucleotides in the guide RNA.

### **Repair template**

**[0075]** The methods of the present disclosure can comprise introduction of a repair template into the cell. The repair template can comprise an insert. The insert can be incorporated at the cut site introduced by a nuclease. The nuclease can be a Cas protein. The Cas protein can be Cas9 or Cpf1. In some cases, the incorporating of the insert at the cut site results in knock-out of the gene. The repair template can be less than 500, 450, 400, 350, 300, 250, 200, 150, 100, 150, 125, 100, 95, 90, 85, 80, or 75 bases in length. The repair template can be from 70 to 80 bases, from 70 to 90 bases, from 70 to 100 bases in length, from 70 to 300 bases in length, or from 70 to 500

bases in length. The repair template can be less than 75 bases in length. The repair template can be 74 bases in length. The repair template can be introduced via a vector comprising an expression cassette comprising the repair template and at least one promoter operably linked to the at least one sequence to be expressed.

**[0076]** The insert can comprise a first stop codon in a first reading frame. The insert can further comprise a second stop codon in a second reading frame. The insert can further comprise a third stop codon in a third reading frame. In some cases, the first reading frame, the second reading frame, and the third reading frame are different. The first stop codon, the second stop codon, the third stop codon, or any combination thereof can be selected from the following stop codons: TGA, TAA, and TAG. The first stop codon, the second stop codon, the third stop codon, or any combination thereof can comprise TGA. In some cases, the first stop codon, the second stop codon, the third stop codon are identical. For example, the first stop codon, the second stop codon, the third stop codon can all be TGA. In some cases, at least one of the first stop codon, the second stop codon, and the third stop codon is TAG. In some cases, the first stop codon, the second stop codon, the third stop codon are different.

**[0077]** The insert can have a length that is not an integer of 3. The insert can be selected from any insert in **TABLE 1**. The insert can be TGATTGATCAATGA (SEQ ID NO. 1). The insert can be TGA CTGATCACTGA (SEQ ID NO. 2). The insert can comprise a nucleic acid of Formula I, wherein Formula I is represented by the sequence TGAXTGAXXXXTGA (SEQ ID NO: 28), wherein X is any nucleotide. The insert can be 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 bases in length. The insert can be less than 20, 19, 18, 17, 16, 15, 14 bases in length. The insert can be 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 bases in length. The guanine-cytosine (GC)-content of the insert can be from 20% to 80%, from 40% to 60%, or from 40% to 50%. The guanine-cytosine (GC)-content of the insert can be greater than 20%, 25%, 30%, 35%, or 40%. The guanine-cytosine (GC)-content of the insert can about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, or about 60%. In some embodiments, the insert does not comprise homology with the gene.

**TABLE 1. Exemplary inserts. Stop codons are indicated with bold text.**

<b>Insert sequence</b>	<b>SEQ ID NO.</b>
<b>TGATTGATCAATGA</b>	1
<b>TGA CTGATCACTGA</b>	2
<b>TAGTTAGTCAATAG</b>	3
<b>TAGCTAGTCACTAG</b>	4
<b>TAATTAATCAATAA</b>	5
<b>TAACTAATCACTGA</b>	6
<b>TGATTGATCAATAG</b>	7

<b>TGATTAGTCAATGA</b>	8
<b>TAGTTGATCAATGA</b>	9
<b>TAGTTAGTCAATGA</b>	10
<b>TAGTTGATCAATAG</b>	11
<b>TGATTAGTCAATAG</b>	12
<b>TGACTGATCACTAG</b>	13
<b>TGACTAGTCACTGA</b>	14
<b>TAGCTGATCACTGA</b>	15
<b>TAGCTAGTCACTGA</b>	16
<b>TAGCTGATCACTAG</b>	17
<b>TGACTAGTCACTAG</b>	18

---

**[0078]** The insert can comprise the identifier sequence. In some embodiments, the insert does not comprise an identifier sequence. In some embodiments, the insert does not comprise homology with the gene. The identifier sequence can be a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, or a combination thereof. The identifier sequence can be a restriction endonuclease recognition sequence. In some cases, the insert comprises a first stop codon in a first reading frame, a second stop codon in a second reading frame, a third stop codon in a third reading frame, and the identifier sequence. The identifier sequence can provide a means to provide confirmation (i.e. identification) of successful incorporation of the insert into the gene. For example, if the identifier sequence is a restriction endonuclease recognition sequence, digestion of the sequence can be informative as to whether the insert was successfully incorporated into the gene. In some embodiments, the identifier sequence is part of the insert. For example, an insert of TGATTTGATCAATGA (SEQ ID NO:1) comprises a BcII restriction site as illustrated by the underlined sequence, which can serve as the identifier sequence.

**[0079]** The restriction endonuclease recognition sequence can be recognized by a restriction endonuclease. The restriction endonuclease recognition sequence is recognized by a restriction endonuclease. The restriction endonuclease can be selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspAII, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI. In some instances, the restriction endonuclease is BcII.

**[0080]** The repair template can further comprise a first homology region with homology to a first portion of the gene and a second homology region with homology to a second portion of gene. The first portion of the gene and the second portion of the gene can be contiguous. The first portion of the gene can be located in an intron, exon, UTR, or a combination thereof of the gene. The second portion of the gene can be located in an intron, exon, UTR, or a combination thereof of the gene. In some cases, homology to the gene indicate that the first homology region, the second homology region, or the combination thereof is identical to the gene (100% homology). In some cases, homology to the gene is 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% homologous of the first homology region, the second homology region, or the combination thereof with the gene. The repair template can comprise the formula H1-I-H2, wherein H1 comprises or consists of the first homology region, I comprises or consists of the insert, and H2 comprises or consists of the second homology region, wherein the insert optionally comprises an identifier sequence.

**[0081]** The percent homology can be calculated by determining the number of positions at which the identical nucleic acid base occurs in two sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions, which can include additions or deletions, and multiplying the result by 100 to yield the percentage of sequence identity. Percent homology, also referred to as percent sequence identity, can be determined by aligning each sequence in any suitable sequence alignment program, such as Clustal Omega, Multiple Sequence Comparison by Log-Expectation (MUSCLE), Multiple Alignment using Fast Fourier Transform (MAFFT), MegAlign, and Basic Local Alignment Search Tool (BLAST).

**[0082]** The first homology region can have homology to the gene on a first side of the cut site. The second homology region can have homology to the gene on a second site of the cut site. The first homology region can be on a first side of the insert and the second homology region can be on a second side of the insert. The first homology region can be 30 bases in length. The first homology region can be 93 bases in length. The first homology region or the combination thereof can be at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, or 110 bases in length. The first homology region or the combination thereof can be no more than 10, 15, 20, 25, 30, 35, 40, 45, or 50 bases in length. The second homology region can be 30 bases in length. The second homology region can be 93 bases in length. The second homology region can be at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, or 110 bases in length. The second homology region or the combination thereof can be no more than 10, 15, 20, 25, 30, 35, 40, 45, or 50 bases in length. In some cases, the



repair template is at least 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, or 230 bases in length, wherein the repair template comprises the first homology region, the insert, and the second homology. In some cases, the repair template does not comprise a PAM site.

**[0083]** Further disclosed, in certain embodiments, are kits comprising a repair template described herein. The repair template can comprise: a first homology region comprising homology to a first portion of the gene; an insert comprising: (i) a first stop codon in a first reading frame; (ii) a second stop codon in a second reading frame; and (iii) a third stop codon in a third reading frame; and a second homology region comprising homology to a second portion of the gene. The repair template can comprise a nucleotide sequence of less than 100 bases in length. The kit can further comprise a guide RNA (gRNA) capable of hybridizing to a gene. The kit can further comprise a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein. Also described herein, in certain embodiments, are kits comprising cells comprising at least one premature stop codon in a gene. The at least one premature stop codon can be caused by incorporation of an insert in the genomic sequence encoding the gene. The at least one premature stop codon can result in knock-out of the gene.

**[0084]** In some embodiments, the kit comprises a carrier, package, or container that is compartmentalized to receive one or more containers such as vials, tubes, and the like, each of the container(s) comprising one of the separate elements to be used in a method described herein. Suitable containers include, for example, bottles, vials, syringes, and test tubes. In some embodiments, the container is formed from a variety of materials such as glass or plastic.

**[0085]** In some embodiments, a kit comprises one or more additional containers, each with one or more of various materials (such as reagents, optionally in concentrated form, and/or devices) desirable from a commercial and user standpoint for use of described herein. Non-limiting examples of such materials include, but not limited to, buffers, primers, enzymes, diluents, filters, carrier, package, container, vial and/or tube labels listing contents and/or instructions for use and package inserts with instructions for use. In some cases, a set of instructions is included. In some cases, a label is on or associated with the container. The label can be on a container when letters, numbers or other characters forming the label are attached, molded or etched into the container itself. The label can be associated with a container when it is present within a receptacle or carrier that also holds the container, e.g., as a package insert. The label can be used to indicate that the contents are to be used for a specific therapeutic application. The label can indicate directions for use of the contents, such as in the methods described herein.

**Methods for knock-out of a gene**

**[0086]** Disclosed herein, in certain embodiments, are methods for knock-out of a gene in a cell or a subject. The methods described herein can comprise: (a) introducing into the cell or the subject: a guide RNA (gRNA), a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein, a repair template comprising an insert, or any combination thereof, and (b) incorporating by homology directed repair (HDR) the insert at a cut site introduced by the Cas protein in the gene. In some cases, the incorporating results in the knock-out of the gene.

**[0087]** Introducing a gRNA, a Cas protein, a nucleic acid molecule encoding the Cas protein, a repair template, or any combination thereof into a cell or a subject can comprise electroporation, nucleofection, microinjection, lipofection, sonication, impalefection, or use of a vector.

Introducing a gRNA, a Cas protein, a nucleic acid molecule encoding the Cas protein, a repair template, or any combination thereof into a cell can comprise electroporation. Introducing a gRNA, a Cas protein, a nucleic acid molecule encoding the Cas protein, a repair template, or any combination thereof into a cell comprises the use of a heat shock applied to the cell.

**[0088]** The term "vector" can refer to a composition for transferring, delivering or introducing a nucleic acid (or nucleic acids) into a cell or a subject. Non-limiting examples of general classes of vectors, include but are not limited to, a viral vector, a plasmid vector, a phage vector, a phagemid vector, a cosmid vector, a fosmid vector, a bacteriophage, and an artificial chromosome. The vector can be a viral vector. The viral vector can be a lentivirus, a retrovirus, an adenovirus, an adeno-associated virus (AAV), or a baculovirus. The viral vector can be a vector capable of replication or a non-replicating vector. The vector can be administered to the cell or a subject via injection. The injection can be intrathecal, intramuscular, intracranial, intraperitoneal, subretinal, subcutaneous, intravitreal, or intravenous. The injection can be a stereotactic injection. The vector can comprise a nucleic acid sequence encoding: a guide RNA, a Cas protein, or any combination thereof. The vector can comprise a nucleic acid sequence of the repair template. In some cases, one vector is administered to the cell or the subject. In other cases, more than one vector is administered to the cell or the subject. Two, three, four, five, six, seven, eight, nine, ten, or more than ten vectors can be administered to the cell or the subject. The vector can further comprise at least one promoter. The vector can comprise an expression cassette comprising at least one sequence to be expressed and at least one promoter operably linked to the at least one sequence to be expressed. The at least one sequence to be expressed can

be a gRNA, a Cas protein, or a combination thereof. The vector can comprise the nucleic acid sequence of the repair template.

**[0089]** For example, a single vector can be used to deliver a nucleic acid encoding a gRNA and a Cas protein, and further comprising a nucleic acid sequence of a repair template. In another example, a first vector can be used to deliver a nucleic acid encoding a gRNA, a second vector can be used to deliver a nucleic acid encoding a Cas protein, and a DNA sequence comprising the repair template can be used to deliver the repair template.

**[0090]** After gene editing by a gene editing tool (e.g., CRISPR/Cas9, TALEN, or ZFN, etc.), a single cell or a population of cells can have one or more indels. A software can be implemented to use sequencing to predict genomic edits of most likely genetic outcomes. The sequencing can be Sanger sequencing or next generation sequencing (NGS). The sequencing can be next generation sequencing. For instance, an algorithm can use a target sequence of the gene to align a first Sanger sequencing trace from a cell (or a first population of cells) without an exposure to the gene editing tool and a second Sanger sequencing trace from the same type of cell but with the exposure to the gene editing tool. The algorithm can generate a library of plausible genomic edits at the target sequence of the first Sanger sequencing trace, where the plausible genomic edits can be from non-homologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), homology directed repair (HDR), single target sequence editing, and/or multiplex editing (e.g., editing multiple target sequences). Subsequently, the algorithm can perform a mathematical analysis between the library and the second Sanger sequencing trace to identify a subset of the library of genomic edits that are likely to be a result of the gene editing tool. The mathematical analysis can comprise regression analysis. Examples of the regression analysis can include Bayesian methods, percentage regression, least absolute deviations, nonparametric regression, distance metric learning, Lasso regression analysis, and non-negative least squares (NNLS) regression analysis. In some cases, the algorithm can use a donor sequence (i.e., repair template) provided from a user to determine an efficiency of the “knock-in” experiment.

**[0091]** The methods of the present disclosure can be used in a variety of cells. Cells can be any prokaryotic or eukaryotic living cells, cell lines derived from these organisms for *in vitro* cultures, primary cells from animal or plant origin. Eukaryotic cells can refer to a fungal, plant, algal or animal cell or a cell line. The eukaryotic cell can be derived from the organisms listed below and established for *in vitro* culture. The fungus can be of the genus *Aspergillus*, *Penicillium*, *Acremonium*, *Trichoderma*, *Chrysosporium*, *Mortierella*, *Kluyveromyces* or *Pichia*; *More preferably, the fungus is of the species Aspergillus niger, Aspergillus nidulans, Aspergillus oryzae, Aspergillus terreus, Penicillium chrysogenum, Penicillium citrinum, Acremonium*

*Chrysogenum*, *Trichoderma reesei*, *Mortierella alpine*, *Chrysosporium lucknowense*, *Kluyveromyceslactis*, *Pichia pastoris* or *Pichia ciferrii*. The plant can be of the genus *Arabidospis*, *Nicotiana*, *Solanum*, *lactuca*, *Brassica*, *Oryza*, *Asparagus*, *Pisum*, *Medicago*, *Zea*, *Hordeum*, *Secale*, *Triticum*, *Capsicum*, *Cucumis*, *Cucurbita*, *Citrullis*, *Citrus*, *Sorghum*. The plant can be of the species *Arabidospis thaliana*, *Nicotiana tabaccum*, *Solanum lycopersicum*, *Solanum tuberosum*, *Solanum melongena*, *Solanum esculentum*, *Lactuca saliva*, *Brassica napus*, *Brassica oleracea*, *Brassica rapa*, *Oryza glaberrima*, *Oryza sativa*, *Asparagus officinalis*, *Pisumsativum*, *Medicago sativa*, *zea mays*, *Hordeum vulgare*, *Secale cereal*, *Triticuma estivum*, *Triticum durum*, *Capsicum sativus*, *Cucurbitapepo*, *Citrullus lanatus*, *Cucumis melo*, *Citrus aurantifolia*, *Citrus maxima*, *Citrus medica*, and *Citrus reticulata*. The animal cell can be of the genus *Homo*, *Rattus*, *Mus*, *Cricetulus*, *Pan*, *Sus*, *Bos*, *Danio*, *Canis*, *Felis*, *Equus*, *Salmo*, *Oncorhynchus*, *Gallus*, *Meleagris*, *Drosophila*, *Caenorhabditis*. The animal cell can be of the species *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Cricetulus griseus*, *Pan paniscus*, *Sus scrofa*, *Bos taurus*, *Canis lupus*, *Cricetulus griseus*, *Danio rerio*, *Felis catus*, *Equus caballus*, *Rattus norvegicus*, *Salmo salar*, *Oncorhynchus mykiss*, *Gallus gallus*, *Meleagris gallopavo*, *Drosophila melanogaster*, and *Caenorhabditis elegans*. In some instances, the cell is a human cell. In some embodiments, the methods of the present disclosure are used in an organism comprising the cells described herein.

**[0092]** Examples cell lines include, but are not limited to, Chinese hamster ovary (CHO) cells (e.g., CHO-K1); HEK293 cells; Caco2 cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; DG44 cells; K-562 cells, U-937 cells; MC5 cells; MCF-7 cells; Saos-2 cells; P3 cells; Sf9 cells; IMR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HL-60 cells; HT-1080 cells; HCT-116 cells; Hu-h7 cells; Huvec cells; and Molt 4 cells. Examples of other cells applicable to the scope of the present disclosure can include stem cells, embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs). A cell lines and/or cells can be modified by the methods of the present disclosure to provide cell line models comprising a knockout of a gene.

**[0093]** Incorporating the insert, by HDR at the cut site introduced by the nuclease can result in disruption of the binding site of the gRNA. In some cases, disruption of the binding site is a result of the repair template breaking up the binding site upon incorporation into the gene, so that the binding site is no longer present. In some cases, incorporating the insert results in the inability of the gRNA to bind to the binding site.

**[0094]** In one example, the MAST3 gene can be the expressible sequence targeted for incorporation of a repair template comprising an insert, and the method comprises introducing a gRNA designed to target a sequence in the MAST3 gene, which in combination with a Cas

protein will introduce a cut-site into the MAST3 gene, (**FIG. 8A**) and introducing the repair template at the cut site (**FIG. 8B**) in order to incorporate the insert into the sequence of the MAST3 gene so that expression of this gene is knocked. Further, incorporation of the repair template comprising the insert can disruption the site where the gRNA binds, so that gRNA will no longer recognize this sequence and no additional cutting will occur (**FIG. 8C**).

**[0095]** In some cases, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the population of cells. In some cases, introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in about 10%, about 15%, about 20%, about 25%, about 30%, about 35%, about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, or about 95% of the population of cells.

**[0096]** In some cases, the use of a repair template 100 bases in length in the methods described herein produces a higher proportion of knockout of a gene relative to the use of a repair template greater than 100 bases in length in a population of cells. In some cases, the use of a repair template 74 bases in length in the methods described herein produces a higher proportion of knockout of a gene relative to the use of a repair template greater than 74 bases in length in a population of cells. In some cases, the use of a repair template comprising an insert of 14 bases in length in the methods described herein produces a higher proportion of knockout of a gene relative to the use of a repair template comprising an insert of greater than 14 bases in length in a population of cells.

**[0097]** The methods and compositions of the present disclosure can be used to generate a knock-out of a gene in a human. The human can suffer from a condition. The condition can be caused by a gain-of-function mutation in the gene. The condition can be selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy. The gain-of-function mutation can produce a dominant negative allele. In some instances, described herein are methods of treating the condition by knock-out of a dominant negative allele using the compositions and methods described herein. The dominant negative allele can be one or both alleles of the gene.

**[0098]** In some embodiments, the compositions of the present disclosure are suitable for parenteral administration. Parenteral administration can comprise injection of the compositions into an intended recipient. The injection can be intrathecal, intramuscular, intracranial, intraperitoneal, subretinal, subcutaneous, intravitreal, or intravenous. Injectable compositions can comprise sterile aqueous and non-aqueous injection solutions. In some embodiments, aqueous and non-aqueous sterile suspensions include suspending agents and thickening agents. In some embodiments, compositions disclosed herein are presented in unit/dose or multi-dose containers, for example sealed ampoules and vials, and are stored in a freeze-dried (lyophilized) condition requiring only the addition of the sterile liquid carrier, for example, saline or water for injection on immediately prior to use.

**[0099]** The guide RNA (gRNA), a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein, a repair template comprising an insert, or any combination thereof can be in form of a pharmaceutical formulation. In some embodiments, a pharmaceutical formulation comprises an excipient. Excipients can include, but are not limited to, solvents, dispersion media, diluents, or other liquid vehicles, dispersion or suspension aids, surface active agents, isotonic agents, thickening or emulsifying agents, preservatives, solid binders, and lubricants.

**[0100]** In some cases, the method further comprises administering a therapeutic agent. The therapeutic agent can be administered before, in conjunction with, or after the compositions described herein. The therapeutic agent can be any therapeutic agent with the ability to treat the disease or condition or relieve the symptoms of the disease or condition. For example, the therapeutic agent can be an anticonvulsant, a nerve pain medication, an analgesic, an anesthetic, a narcotic, an antiarrhythmic agent (e.g. quinidine), a selective serotonin reuptake inhibitor (e.g. fluoxetine), a dopamine promoter (e.g. Carbidopa-Levodopa), an antidepressant, a cognition-enhancing medication, an anti-tremor medication, a catechol-O-methyl transferase (COMT) inhibitor, monoamine oxidase inhibitor (MAO) inhibitors, anticholinergic agents, or stem cells.

**[0101]** In some embodiments, the compositions disclosed herein are administered before, during, or after the occurrence of the disease or condition. In some embodiment, the timing of administering the composition varies. In some embodiments, the pharmaceutical compositions are used as a prophylactic and are administered prior to onset of the symptoms of the disease or condition. For example, the composition can be administered to an individual upon a diagnostic sequencing test showing that the individual has a gain-of-function mutation. In some embodiments, pharmaceutical compositions are administered to a subject during or as soon as

possible after the onset of the symptoms of a disease or condition associated with the gain-of-function mutation.

[0102] Disclosed herein, in certain embodiments, are engineered cells comprising a gene comprising an insert described herein. The insert in an engineered cell can comprise a nucleic acid sequence of from 12 to 23, from 13 to 22, from 14 to 21, from 14 to 20, from 14 to 19, from 14 to 18, from 14 to 17, from 14 to 16, or from 14 to 15 bases in length.

### Computer Systems

[0103] The present disclosure provides computer systems that are programmed to implement methods of the disclosure or portions of methods of the disclosure. Computer systems of the present disclosure can be used to design one or more guide RNAs for hybridizing the gene, design one or more repair templates for incorporating an insert into the gene, or a combination thereof. Information from any of the computer systems can provide a report to a remote computer. The computer systems described herein can be used to implement a software to predict genomic edits of most likely genetic outcomes, as previously described herein.

[0104] FIG. 9 shows a computer system 901 that is programmed or otherwise configured to communicate with and regulate various aspects of a computer system of the present disclosure. The computer system 901 can regulate various aspects of the present disclosure, such as, for example, designing one or more guide RNAs for hybridizing the gene, calculating an off-target value by enumerating a number of mismatches to potential guide RNA hybridizing sites in the genome of interest, or designing one or more repair templates for incorporating an insert into the gene. The computer system 901 can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

[0105] The computer system 901 includes a central processing unit (CPU, also “processor” and “computer processor” herein) 905, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system 901 also includes memory or memory location 910 (e.g., random-access memory, read-only memory, flash memory), electronic storage unit 915 (e.g., hard disk), communication interface 920 (e.g., network adapter) for communicating with one or more other systems, and peripheral devices 925, such as cache, other memory, data storage and/or electronic display adapters. The memory 910, storage unit 915, interface 920 and peripheral devices 925 are in communication with the CPU 905 through a communication bus (solid lines), such as a motherboard. The storage unit 915 can be a data storage unit (or data repository) for storing data. The computer system 901 can be operatively

coupled to a computer network (“network”) 930 with the aid of the communication interface 920. The network 930 can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network 930 in some cases is a telecommunication and/or data network. The network 930 can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network 930, in some cases with the aid of the computer system 901, can implement a peer-to-peer network, which can enable devices coupled to the computer system 901 to behave as a client or a server.

**[0106]** The CPU 905 can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions can be stored in a memory location, such as the memory 910. The instructions can be directed to the CPU 905, which can subsequently program or otherwise configure the CPU 905 to implement methods of the present disclosure. Examples of operations performed by the CPU 905 can include fetch, decode, execute, and writeback.

**[0107]** The CPU 905 can be part of a circuit, such as an integrated circuit. One or more other components of the system 901 can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

**[0108]** The storage unit 915 can store files, such as drivers, libraries and saved programs. The storage unit 915 can store user data, e.g., user preferences and user programs. The computer system 901 in some cases can include one or more additional data storage units that are external to the computer system 901, such as located on a remote server that is in communication with the computer system 901 through an intranet or the Internet.

**[0109]** The computer system 901 can communicate with one or more remote computer systems through the network 930. For instance, the computer system 901 can communicate with a remote computer system of a user. Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC’s (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system 901 via the network 930.

**[0110]** Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system 901, such as, for example, on the memory 910 or electronic storage unit 915. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor 905. In some cases, the code can be retrieved from the storage unit 915 and stored on the memory 910 for ready access by the processor 905. In some situations, the electronic storage unit 915 can be precluded, and machine-executable instructions are stored on



memory 910.

**[0111]** The code can be pre-compiled and configured for use with a machine having a processor adapted to execute the code or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

**[0112]** Aspects of the systems and methods provided herein, such as the computer system 901, can be embodied in programming. Various aspects of the technology can be thought of as “products” or “articles of manufacture” typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. “Storage” type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which can provide non-transitory storage at any time for the software programming. All or portions of the software can at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, can enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that can bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also can be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible “storage” media, terms such as computer or machine “readable medium” refer to any medium that participates in providing instructions to a processor for execution.

**[0113]** Hence, a machine readable medium, such as computer-executable code, can take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as can be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media can take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio

frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer can read programming code and/or data. Many of these forms of computer readable media can be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[0114] The computer system 901 can include or be in communication with an electronic display 935 that comprises a user interface (UI) 940 for providing the use, for example, the ability to select a species of interest and gene of interest from the species of interest. Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

[0115] Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit 905. The algorithm can, for example, design one or more gRNAs for hybridizing a gene, activate and initiate synthesis of at least one of the one or more gRNAs, design one or more repair templates for incorporating an insert into the gene, activate and initiate synthesis of at least one of the one or more repair templates, or any combination thereof.

## EXAMPLES

### Example 1. Synthesis of modified sgRNA, donor DNA, and transfection optimization

[0116] Modified sgRNAs targeting the Exon 1 of the human T cell receptor alpha (TRAC) gene were synthesized (TABLE 2; FIG. 1).

**TABLE 2. Modified sgRNAs targeting exon 1 of TRAC.**

<b>sgRNA Modified Name</b>	<b>sgRNA sequence</b>	<b>SEQ ID NO.</b>
TRAC sgRNA1	GUCAGGGUUCUGGAUAUCUG	19
TRAC sgRNA2	GCUGGUACACGGCAGGGUCA	20
TRAC sgRNA3	CUCUCAGCUGGUACACGGCA	21
TRAC sgRNA4	UCUCUCAGCUGGUACACGGC	22
TRAC ex1 sgRNA	UCAGGGUUCUGGAUAUCUG	23

[0117] A single stranded oligonucleotide repair template containing a sequence with all three stop codon reading frames, flanked by their respective regions of homology, was designed and

synthesized to serve as a template for incorporation into exon 1 of TRAC (exon 1, SEQ ID NO. 24) via homology directed repair.

[0118] Different transfection parameters for optimization of homology directed repair (HDR) in primary human T cells (CD3+) were tested. The most optimized transfection parameters produced >85% editing efficiency with TRAC sgRNA3. TRAC sgRNA3 was therefore selected as the preferred gRNA. The editing efficiency of incorporation of the repair template was determined using the primers described in **TABLE 3** for generating Sanger sequencing data.

**TABLE 3. Genotyping primers.**

Primer	Sequence	SEQ ID NO.
TRAC geno F2	TCAGGTTTCCTTGAGTGGCAGG	25
TRAC geno R2	TAAGGCCGAGACCACCAATCAG	26

### **Example 2. Transfection and detection of knock-in (KI) and indel formation**

[0119] Following the synthesis of the modified sgRNA and single strand donor DNA (repair template), primary human T cells were transfected using the optimized protocol derived in Example 1.

[0120] A software to predict genomic edits of most likely genomic outcomes was used to analyze the Sanger sequencing results for the knock-out (KO) pool. The software inferred that the pool contained 90% multi-allelic indels (genotype contributions <5% were not considered significant). The software's KO-Score predicted that 83% of the pool contained indels that results in a frameshift. Knock-in of the repair template comprising the insert, represented by the +14 in **FIG. 3**, was detected in 11.6% of the population (**FIG. 2**).

[0121] Fluorescence activated cell sorting (FACS) analysis was subsequently used to detect knockout of TCRA. A fluorescein isothiocyanate (FITC) conjugated anti-CD3 antibody was used to identify populations of CD3+ PAN T-cells. 95% of the wild-type T-cells were found to be CD3+ using FACS (**FIGS. 5A-5B**). FACS analysis on the TCRA KO T-cell pool revealed only 17.6% of the pool is positive for CD3 suggesting ~83% TCRA knock-out (**FIGS. 6A-6B**).

### **Example 3. Quality control, shipment of live cell pools, and cryogenic cell pools**

[0122] The edited cell populations were frozen down into 4 vials at  $5 \times 10^6$  cells/vial. Test samples were thawed on the subsequent day, genotype reconfirmed (using the software to predict genomic edits of most likely genomic outcomes) and QC tested.

**Example 4. Determination of optimum transfection conditions for nucleofection of a repair template comprising an insert into the MAST3 gene in HEK293 cells**

[0123] HEK293 cells (20,000 cells) were electroporated with precomplexed ribonucleoprotein (RNP) composed of 20 pmoles of Cas9 + 180 pmoles MAST3 sgRNA. Varying amounts of the repair template comprising an insert (the donor) at different concentrations (0, 10, 30, and 60 pmol) were co-transfected along with the RNPs to identify the optimal concentration for the highest KI efficiency (**FIG. 7A**). Transfected cells were recovered in normal growth medium, transferred to 96-well tissue culture dishes, and allowed to recover/edit for 48 hours. At 48 hours post transfection, genomic DNA was extracted from the cells. PCR was performed using the genomic DNA as template and primers flanking the site of insertion. After a standard PCR clean-up to remove nucleotides and primers, the amplified DNA was submitted for Sanger sequencing. The resulting chromatograms were used for analysis using a software to predict genomic edits of most likely genomic outcomes in order to determine the percentage of sequences containing the repair template comprising the insert (**FIG. 7B**). Restriction enzyme cleavage analysis by a fragment analyzer is presented in **FIG. 7C**.

[0124] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein can be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

## CLAIMS

## WHAT IS CLAIMED IS:

1. A method for introducing at least three stop codons in at least three different reading frames of a gene in a cell, comprising:
  - (a) introducing into the cell:
    - (i) a guide RNA (gRNA) capable of hybridizing to the gene;
    - (ii) a clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein; and
    - (iii) a repair template comprising:
      - (a) a first homology region comprising homology to a first portion of the gene;
      - (b) an insert comprising:
        - (i) a first stop codon in a first reading frame;
        - (ii) a second stop codon in a second reading frame; and
        - (iii) a third stop codon in a third reading frame,wherein the first reading frame, the second reading frame, and the third reading frame are different; and
      - (c) a second homology region comprising homology to a second portion of the gene;wherein the repair template comprises a nucleotide sequence of less than 100 bases in length; and
  - (b) incorporating by homology directed repair the insert at a cut site introduced by the Cas protein in the gene, wherein the incorporating results in knock-out of the gene in the cell.
2. The method of claim 1, wherein the incorporating is in an exon of the gene.
3. The method of claim 1 or claim 2, wherein the first stop codon, the second stop codon, and the third stop codon are identical.
4. The method of any one of claims 1-3, wherein the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA.
5. The method of any one of claims 1-4, wherein the insert comprises a nucleotide sequence of a length that is not an integer of 3.
6. The method of any one of claims 1-5, wherein the insert comprises a nucleotide sequence of a length of no more than 20 bases.

7. The method of any one of claims 1-6, wherein the insert comprises a nucleotide sequence of a length of 14 bases.
8. The method of any one of claims 1- 7, wherein the insert has a GC content from 40% to 50%.
9. The method of any one of claims 1-7, wherein the insert comprises a nucleic acid sequence selected from any one of SEQ ID NO: 1- 18.
10. The method of any one of claims 1-7, wherein the insert comprises SEQ ID NO:1.
11. The method of any one of claims 1-8, wherein the insert comprises SEQ ID NO: 2.
12. The method of any one of claims 1-10, wherein the insert comprises an identifier sequence.
13. The method of claim 12, wherein the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof.
14. The method of claim 13, wherein the restriction endonuclease recognition sequence is recognized by a restriction endonuclease.
15. The method of claim 14, wherein the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspAII, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmlI, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI.
16. The method of claim 14, wherein the restriction endonuclease is BcII.
17. The method of any one of claims 1-16, wherein the first homology region comprises a nucleotide sequence of a length of 30 bases or less.
18. The method of any one of claims 1-17, wherein the second homology region comprises a nucleotide sequence of a length of 30 bases or less.
19. The method of any one of claims 1-18, wherein the repair template comprises a nucleotide sequence of a length of 74 bases.
20. The method of any one of claims 1-19, wherein the Cas protein is Cas9.
21. The method of any one of claims 1-20, wherein the gRNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA).

22. The method of any one of claims 1-21, wherein the introducing comprises electroporation.
23. The method of any one of claims 1-22, wherein the cell is from a cell line.
24. The method of claim 23, wherein the cell line is selected from the group consisting of: Chinese hamster ovary (CHO) cells; HEK293 cells; Caco2 cells; U2-OS cells; NIH 3T3 cells; NSO cells; SP2 cells; DG44 cells; K-562 cells, U-937 cells; MC5 cells; MCF-7 cells; Saos-2 cells; P3 cells; Sf9 cells; IMR90 cells; Jurkat cells; HepG2 cells; HeLa cells; HL-60 cells; HT-1080 cells; HCT-116 cells; Hu-h7 cells; Huvec cells; and Molt 4 cells.
25. The method of any one of claims 1-24, wherein the cell is a mammalian cell.
26. The method of claim 25, wherein the mammalian cell is a human cell.
27. The method of claim 26, wherein the human cell is a T cell.
28. The method of claim 26, wherein the gene is a T cell receptor alpha (TRAC).
29. The method of any one of claims 1-28, wherein the gene causes a condition in a human.
30. The method of claim 29, wherein the condition is caused by a gain-of-function mutation in the gene.
31. The method of claim 30, wherein the condition is selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy.
32. The method of any one of claims 1-31, wherein introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 10% of the population of cells.
33. The method of any one of claims 1-32, wherein introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 40% of the population of cells.
34. The method of any one of claims 1-33, wherein introducing the gRNA, Cas protein, and repair template into a population of cells results in an incorporating of the insert at the cut site in at least 50% of the population of cells.
35. The method of any one of claims 1-34, wherein introducing the nuclease and the repair template into a population of cells results in an incorporating of the insert at the cut site in about 10% - 60% of the population of cells.

36. A composition for knock-out of a gene comprising a repair template comprising:
- (i) a first homology region comprising homology to a first portion of the gene;
  - (ii) an insert comprising:
    - (a) a first stop codon in a first reading frame;
    - (b) a second stop codon in a second reading frame; and
    - (c) a third stop codon in a third reading frame,wherein the first reading frame, the second reading frame, and the third reading frame are different; and
  - (iii) a second homology region comprising homology to a second portion of the gene;
- wherein the repair template comprises a nucleotide sequence length of less than 100 bases.
37. The composition of claim 36, wherein the first stop codon, the second stop codon, and the third stop codon are identical.
38. The composition of claim 36 or claim 37, wherein the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA.
39. The composition of any one of claims 36-38, wherein the insert comprises a nucleotide sequence of a length that is not an integer of 3.
40. The composition of any one of claims 36-39, wherein the insert comprises a nucleotide sequence of a length of no more than 20 bases.
41. The composition of any one of claims 36-40, wherein the insert comprises a nucleotide sequence of a length of 14 bases.
42. The composition of any one of claim 36-41, wherein the insert has a GC content from 40% to 50%.
43. The method of any one of claims 36-41, wherein the insert comprises a nucleic acid sequence selected from any one of SEQ ID NO: 1-18.
44. The composition of any one of claims 36-41, wherein the insert comprises SEQ ID NO: 1.
45. The composition of any one of claims 36-42, wherein the insert comprises SEQ ID NO: 2.
46. The composition of any one of claims 36-45 wherein the insert comprises an identifier sequence.



47. The composition of claim 46, wherein the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof.
48. The composition of claim 46, wherein the restriction endonuclease recognition sequence is recognized by a restriction endonuclease.
49. The composition of claim 48, wherein the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI.
50. The composition of claim 48, wherein the restriction endonuclease is BcII.
51. The composition of any one of claims 36-50, wherein the first homology region comprises a nucleotide sequence of a length of 30 bases or less.
52. The composition of any one of claims 36-51, wherein the second homology region comprises a nucleotide sequence of a length of 30 bases or less.
53. The composition of any one of claims 36-52, wherein the repair template comprises a nucleotide sequence of a length of 74 bases.
54. The composition of any one of claims 34-53, wherein the composition is in the form of a powder or a solution.
55. The composition of claim 54, wherein the powder is a lyophilized powder.
56. A kit comprising the composition of any one of claims 36-55 and instructions.
57. The kit of claim 56, further comprising a guide RNA capable of hybridizing to a gene.
58. The kit of claim 57, wherein the guide RNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA).
59. The kit of any one of claims 56-58, further comprising a clustered, regularly interspersed, short palindromic repeats (CRISPR) associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein.
60. The kit of claim 59, wherein the Cas protein is Cas9.

61. A method of treating a subject comprising a gain-of-function mutation in a gene, the method comprising administering to the subject a composition of any one of claims 36-55.
62. The method of claim 61, further comprising administering to the subject a guide RNA capable of hybridizing to the gene.
63. The method of claim 62, wherein the guide RNA comprises a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA).
64. The method of any one of claims 61-63, further comprising administering to the subject a clustered, regularly interspersed, short palindromic repeats (CRISPR) associated protein (Cas protein) or a nucleic acid molecule encoding the Cas protein.
65. The method of claim 64, wherein the Cas protein is Cas9.
66. The method of any one of claims 61-65, wherein the gain-of-function mutation results in a condition.
67. The method of claim 66, wherein the condition is selected from the group consisting of: hereditary motor and sensory neuropathy type IIC, postsynaptic slow-channel congenital myasthenic syndrome, PRPS1 Huntington's disease, Parkinson's disease, tubular aggregate myopathy, achondroplasia, Lubs X-linked mental retardation syndrome, spinocerebellar ataxia, spinal and bulbar muscular atrophy, Freiderich ataxia, myotonic dystrophy, and oculopharyngeal muscular dystrophy.
68. The method of any one of claims 61-67, wherein the subject is a mammal.
69. The method of claim 68, wherein the mammal is a human.
70. An engineered cell comprising a gene engineered to comprise an insert in a , wherein the insert comprises a first stop codon in a first reading frame, a second stop codon in a second reading frame, and a third stop codon in a third reading frame, wherein the insert is a non-naturally occurring sequence, and wherein the insert comprises a nucleic acid sequence of 14 bases in length.
71. The engineered cell of claim 70, wherein the insert is in an exon of the gene.
72. The engineered cell of claim 70 or claim 71, wherein the first stop codon, the second stop codon, and the third stop codon are identical.
73. The engineered cell of any one of claims 70-72, wherein the first stop codon, the second stop codon, the third stop codon, or any combination thereof are TGA.
74. The engineered cell of any one of claim 70-73, wherein the insert has a GC content from 40% to 50%.

75. The engineered cell of any one of claims 70-73, wherein the insert comprises SEQ ID NO: 1.
76. The engineered cell of any one of claims 70-74, wherein the insert comprises SEQ ID NO: 2.
77. The engineered cell of any one of claims 70-76 wherein the insert comprises an identifier sequence.
78. The engineered cell of claim 77, wherein the identifier sequence is selected from the group consisting of: a restriction endonuclease recognition sequence, an oligonucleotide probe binding sequence, a sequencing primer annealing sequence, a barcode, and a combination thereof.
79. The engineered cell of claim 78, wherein the restriction endonuclease recognition sequence is recognized by a restriction endonuclease.
80. The engineered cell of claim 79, wherein the restriction endonuclease is selected from the group consisting of: AatII, Acc65I, AccI, AcII, AfeI, AfIII, AgeI, ApaI, ApaLI, ApoI, AscI, AseI, AsiSI, AvrII, BamHI, BcII, BgIII, Bme1580I, BmtI, BsaHI, BsiEI, BsiWI, BspEI, BspHI, BsrGI, BssHII, BstBI, BstZ17I, BtgI, ClaI, DraI, EaeI, EagI, EcoRI, EcoRV, FseI, FspI, HaeII, HincII, HindIII, HpaI, KasI, KpnI, MfeI, MluI, MscI, MspA1I, NaeI, NarI, NcoI, NdeI, NgoMIV, NheI, NotI, NruI, NsiI, NspI, PacI, PciI, PmeI, PmII, PsiI, PspOMI, PstI, PvuI, PvuII, SacI, SacII, SaII, SbfI, SfcI, SfoI, SgrAI, SmaI, SmII, SnaBI, SpeI, SphI, SspI, StuI, SwaI, XbaI, XhoI, and XmaI.
81. The engineered cell of claim 79, wherein the restriction endonuclease is BcII.
82. The engineered cell of any one of claims 70-81, wherein the engineered cell is a mammalian cell.
83. The engineered cell of claim 82, wherein the mammalian cell is a human cell.
84. A kit comprising the engineered cell of any one of claims 70-82 and instructions.

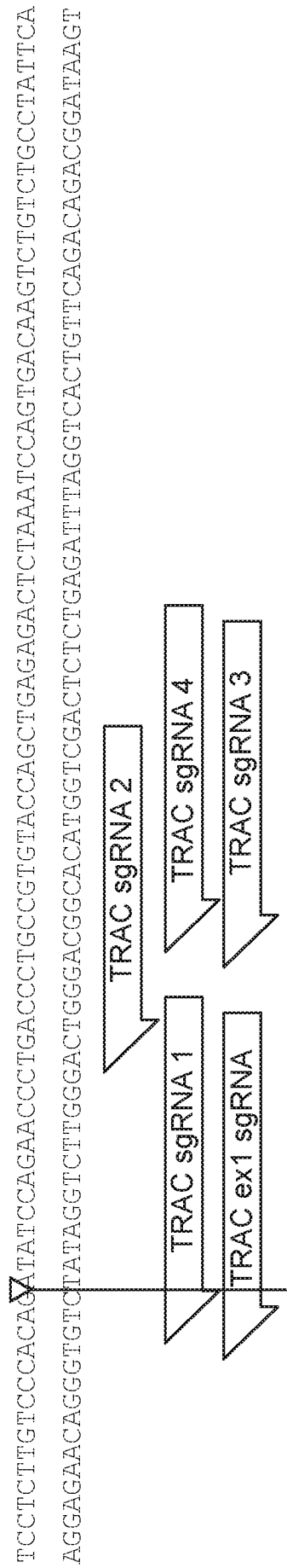


FIG. 1

STATUS <sup>?</sup> Succeeded | GUIDE TARGET <sup>?</sup> CTCACGCTGGTACACGGCA | PAM SEQUENCE <sup>?</sup> GGG | Intel percentage score 90 | R<sup>2</sup> <sup>?</sup> 0.94 | KO SCORE <sup>?</sup> 83

POWERED BY >SYNTHIGO ICE

RELATIVE CONTRIBUTION OF EACH SEQUENCE (NORMALIZED)

INDEL CONTRIBUTION	SEQUENCE
+1	38.8% CAGATAATCCAGAAACCCCTGACCCCTGC
+14	11.6% CAGATAATCCAGAAACCCCTGACCCCTGC
-1	9.0% CAGATAATCCAGAAACCCCTGACCCCTGC
-11	5.3% CAGATAATCCAGAAACCCCTGACCCCTGC
0	4.5% CAGATAATCCAGAAACCCCTGACCCCTGC
-2	2.7% CAGATAATCCAGAAACCCCTGACCCCTGC
-5	2.7% CAGATAATCCAGAAACCCCTGACCCCTGC
-10	2.1% CAGATAATCCAGAAACCCCTGACCCCTGC
-14	1.9% CAGATAATCCAGAAACCCCTGACCCCTGC
-12	1.8% CAGATAATCCAGAAACCCCTGACCCCTGC
-12	1.4% CAGATAATCCAGAAACCCCTGACCCCTGC
-24	1.4% CAGATAATCCAGAAACCCCTGACCCCTGC
-11	1.4% CAGATAATCCAGAAACCCCTGACCCCTGC
-18	1.4% CAGATAATCCAGAAACCCCTGACCCCTGC
+2	1.1% CAGATAATCCAGAAACCCCTGACCCCTGC
-24	1.1% CAGATAATCCAGAAACCCCTGACCCCTGC
-11	1.1% CAGATAATCCAGAAACCCCTGACCCCTGC
-2	1.0% CAGATAATCCAGAAACCCCTGACCCCTGC
-3	0.8% CAGATAATCCAGAAACCCCTGACCCCTGC
+3	0.7% CAGATAATCCAGAAACCCCTGACCCCTGC
-20	0.6% CAGATAATCCAGAAACCCCTGACCCCTGC
-18	0.5% CAGATAATCCAGAAACCCCTGACCCCTGC
-17	0.5% CAGATAATCCAGAAACCCCTGACCCCTGC
-20	0.4% CAGATAATCCAGAAACCCCTGACCCCTGC
-18	0.2% CAGATAATCCAGAAACCCCTGACCCCTGC
-7	0.1% CAGATAATCCAGAAACCCCTGACCCCTGC
-18	0.1% CAGATAATCCAGAAACCCCTGACCCCTGC

FIG. 2

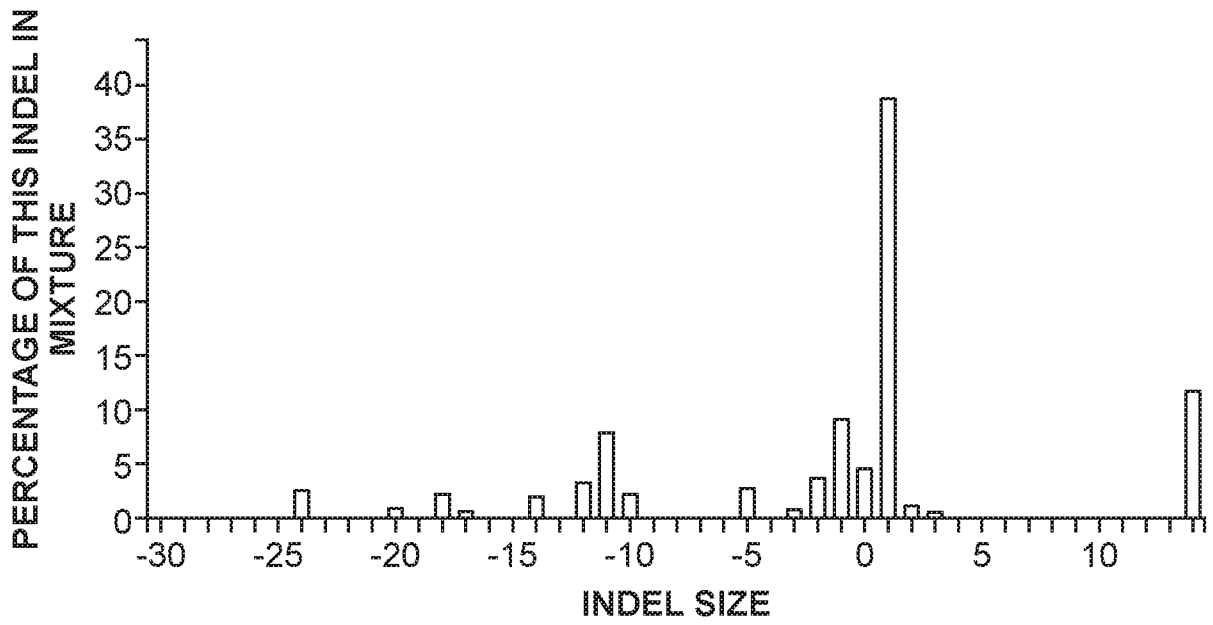


FIG. 3

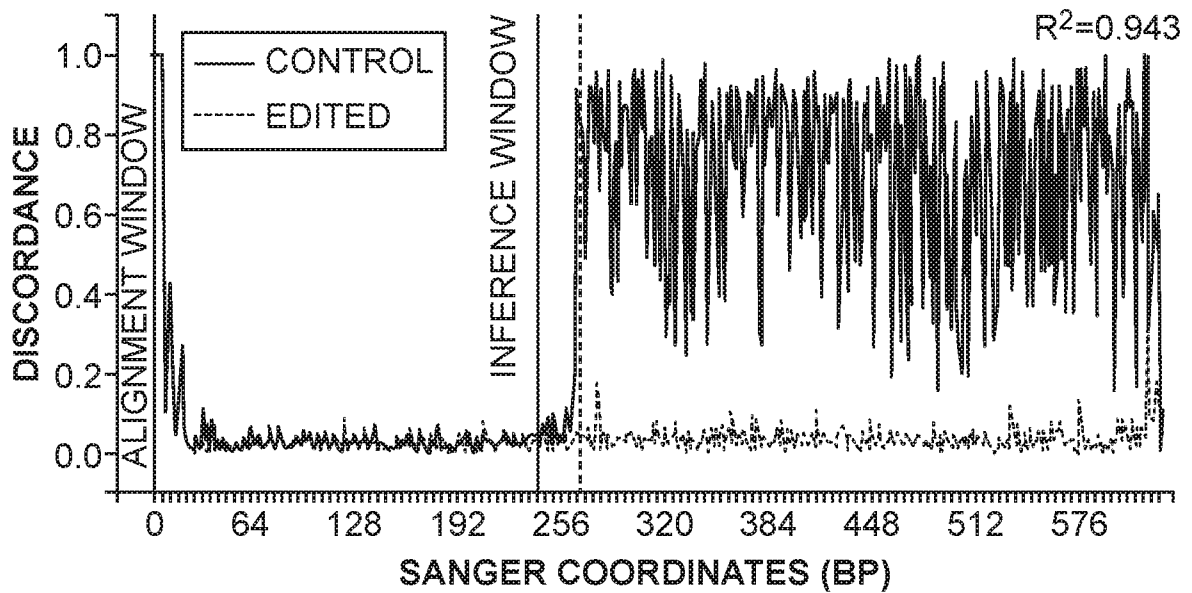


FIG. 4

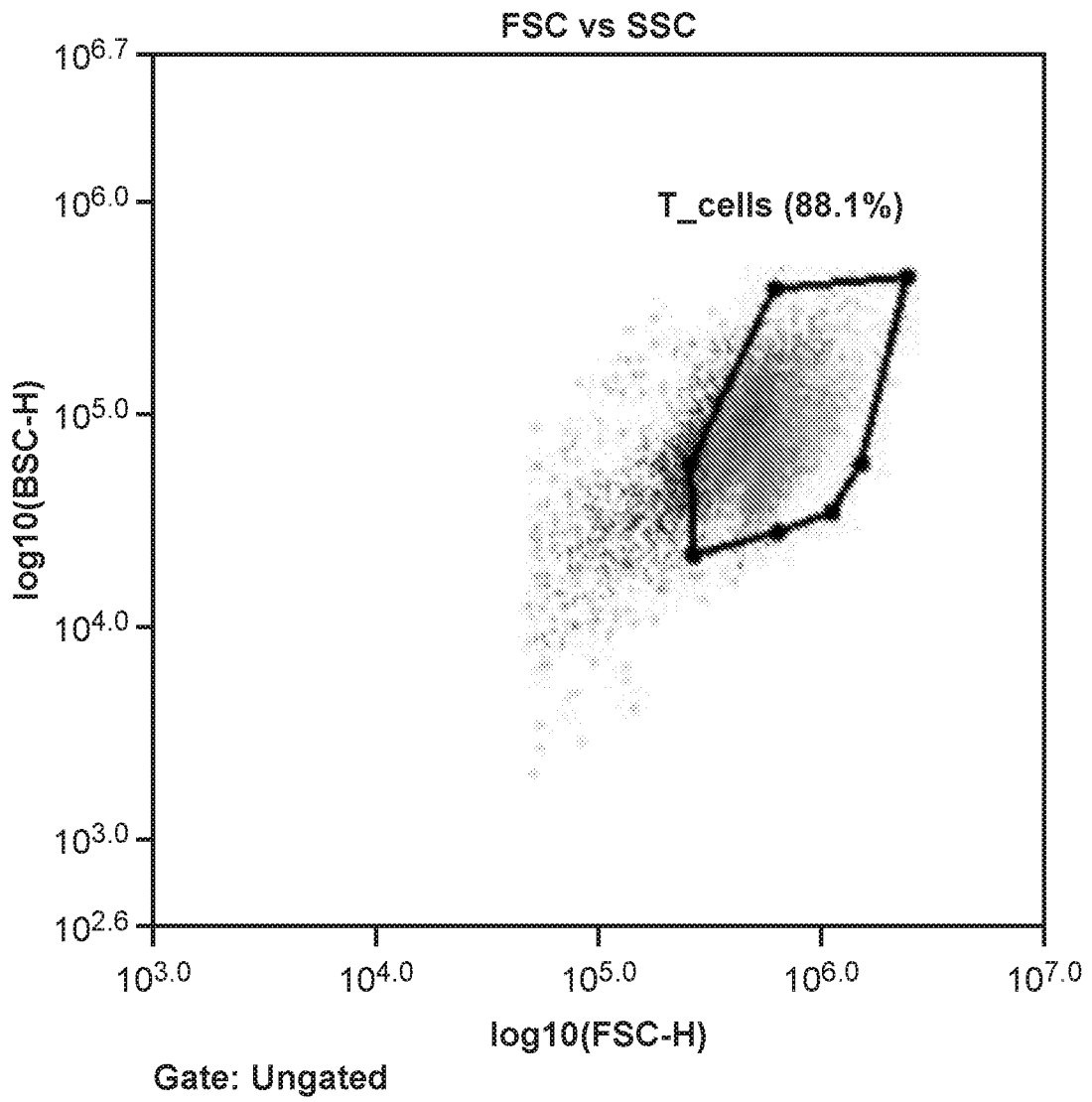


FIG. 5A

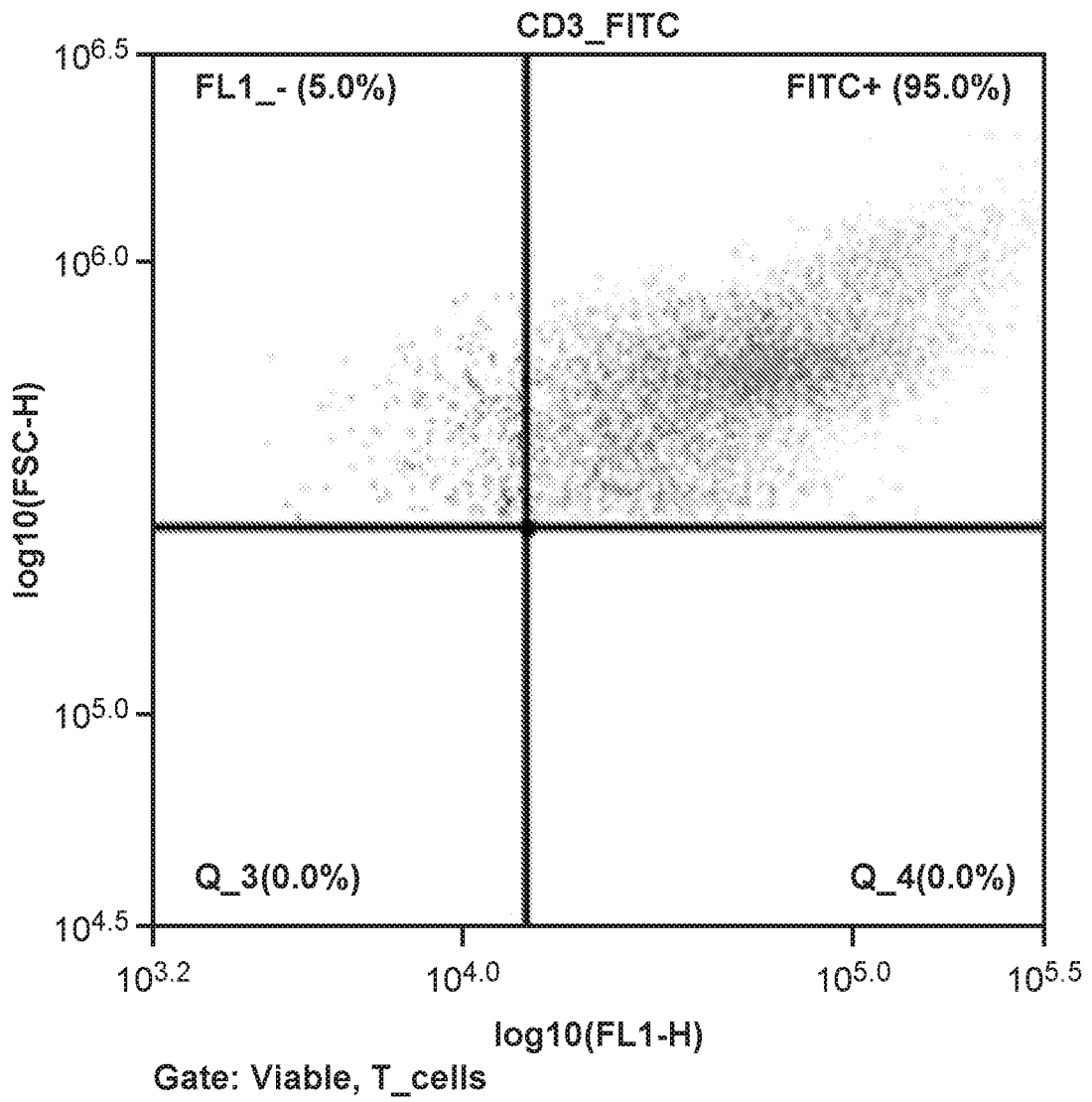


FIG. 5B



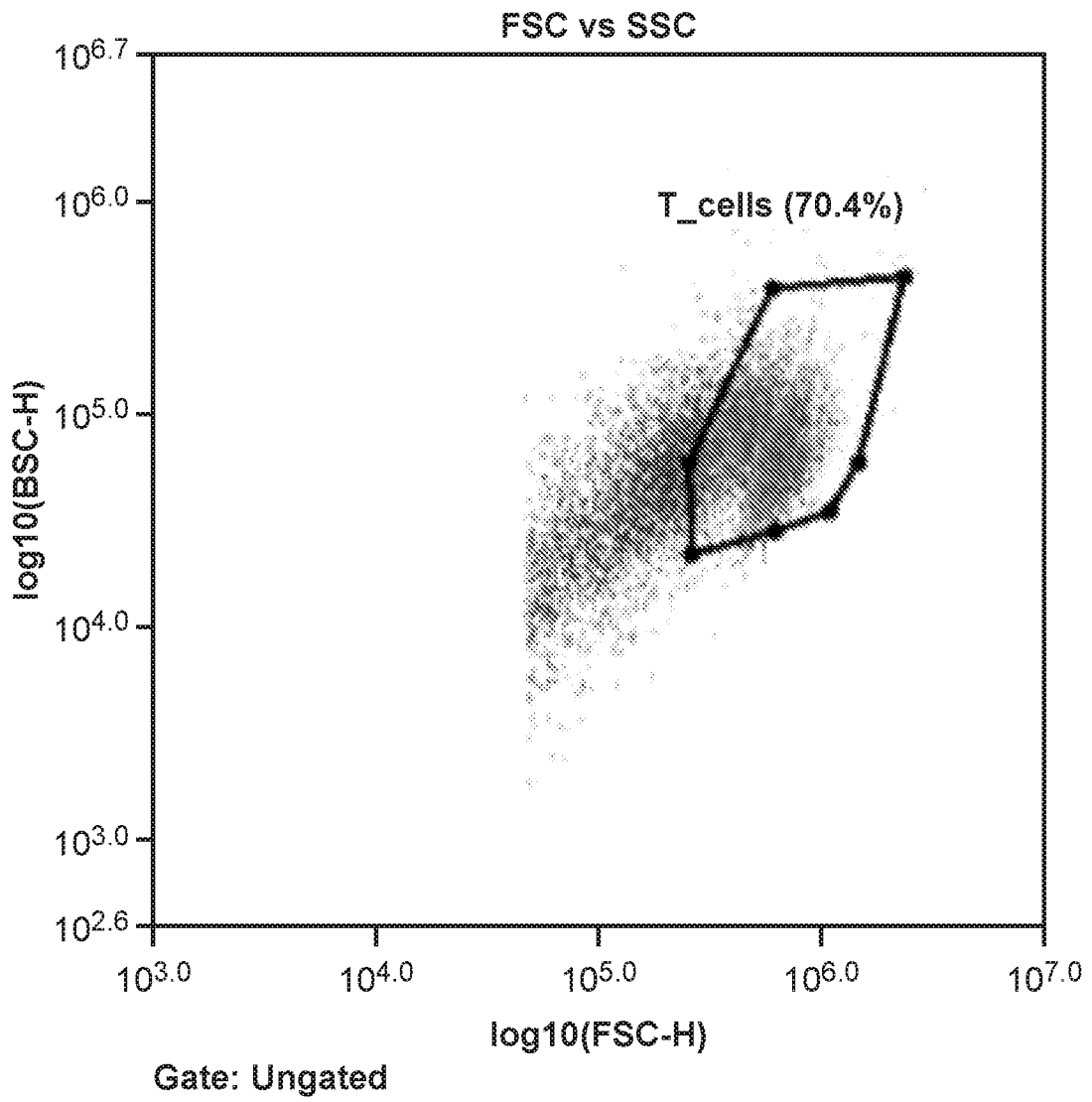


FIG. 6A

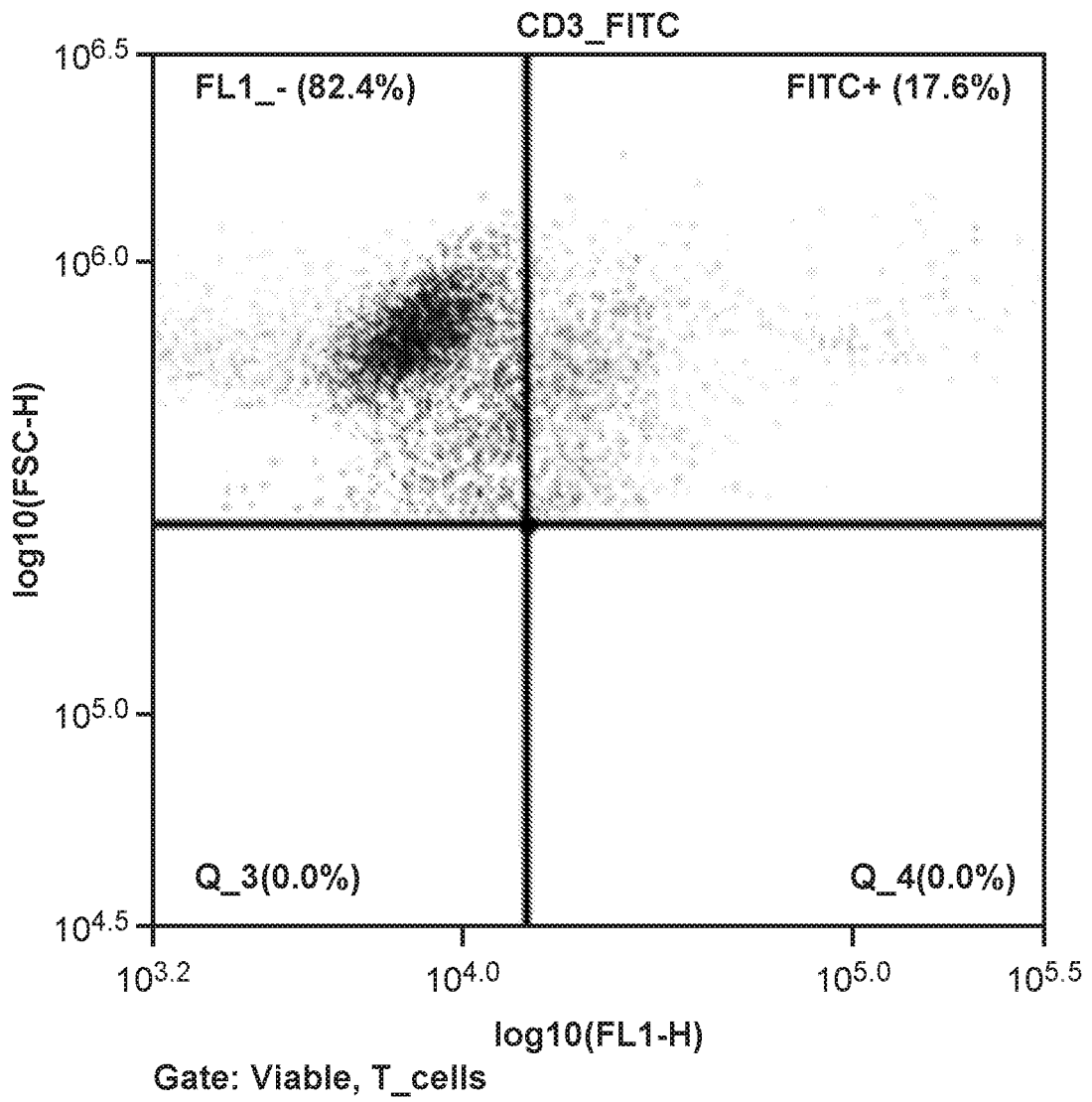


FIG. 6B

Experiment	Nucleofection Experimental Setup				Donor Amount (pmol)	Experimental Conditions: gRNA	gRNA Amount (pmol)	NEB Cass9 Amount (pmol)	Delivery Method
	NF #	Expt Conditions: Donor	Donor Amount (pmol)	Experimental Conditions: gRNA					
Delivery Metho	1	MAST3 ssODN_T	30	MAST3 -18123265 gRNA	180	20	RNP -> D		
	2	MAST3 ssODN_NT	30	MAST3 -18123265 gRNA	180	20	RNP -> D		
	3	RNP only_no donor	0	MAST3 -18123265 gRNA	180	20	RNP -> D		
	4	Neg Cl_no gRNA (ssODN)	30	None	0	20	RNP -> D		
	5	Mock	0	Mock	0	0	RNP -> D		
Donor Orientat	6	MAST3 ssODN_T	10	MAST3 -18123265 gRNA	180	20	Co-deliver		
	7	MAST3 ssODN_T	30	MAST3 -18123265 gRNA	180	20	Co-deliver		
	8	MAST3 ssODN_T	60	MAST3 -18123265 gRNA	180	20	Co-deliver		
	9	MAST3 ssODN_NT	10	MAST3 -18123265 gRNA	180	20	Co-deliver		
	10 [1]	MAST3 ssODN_NT	30	MAST3 -18123265 gRNA	180	20	Co-deliver		
Controls	11 [2]	MAST3 ssODN_NT	60	MAST3 -18123265 gRNA	180	20	Co-deliver		
	12	RNP only_no donor	0	MAST3 -18123265 gRNA	180	20	Co-deliver		
	13	Neg Cl_no gRNA (ssODN)	30	None	0	20	Co-deliver		
	14	Mock	0	Mock	0	0	Co-deliver		

FIG. 7A

Sanger Sequencing Analysis by ICE										
Total = ICE-D (signal from all edited), NHEJ = ICE (signal explained by potential +/- 10bp indels), KI = (ICE-D score) - (ICE score)										
Experiment	Total Efficiency Rep1 (%)	Total Efficiency Rep2 (%)	Total Efficiency Avg (%)	Estimated NHEJ Rep1 (%)	Estimated NHEJ Rep2 (%)	Estimated NHEJ Avg (%)	Estimated KI Rep1 (%)	Estimated KI Rep2 (%)	Estimated KI Avg (%)	Estimated KI Avg (%)
Delivery Method	86.10	87.06	86.58	75.42	78.14	76.78	10.68	8.92	9.80	9.80
	86.27	86.77	86.52	65.69	77.40	71.55	20.58	9.38	14.98	14.98
Controls	83.02	84.05	83.53	85.76	87.28	86.52	NA	NA	NA	NA
	1.82	1.20	1.51	0.89	0.00	0.45	0.93	1.20	1.06	1.06
	-	-	-	0.00	0.00	0.00	0.00	0.00	-	-
Donor Orientation	86.38	87.05	86.71	86.19	86.90	86.55	0.19	0.15	0.17	0.17
	87.90	88.83	88.36	71.29	67.51	69.40	16.60	21.32	18.96	18.96
	89.05	89.85	89.45	50.47	48.67	49.57	38.58	41.17	39.87	39.87
	82.31	83.10	82.70	77.97	79.52	78.74	4.34	3.57	3.96	3.96
	Discarded	87.22	87.22	Discarded	48.13	48.13	Discarded	39.09	39.09	39.09
	Discarded	86.94	86.94	Discarded	33.37	33.37	Discarded	53.57	53.57	53.57
Controls	84.50	85.96	85.23	86.75	87.24	86.99	NA	NA	NA	NA
	1.83	1.12	1.48	0.41	0.00	0.21	1.42	1.12	1.27	1.27
	-	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.00

FIG. 7B

**Restriction Enzyme Cleavage Analysis by  
Fragment Analyzer  
HDR = %Cleavage**

Experiment	%Cleavage Rep1 (%)	%Cleavage Rep2 (%)	%Cleavage Avg (%)
Delivery Metho	19.71	15.93	17.82
	23.73	13.67	18.70
Controls	NaN	NaN	NaN
	NaN	NaN	NaN
	NaN	NaN	NaN
Donor Orientat	NaN	NaN	NaN
	33.91	24.73	29.32
	34.31	37.67	35.99
	NaN	7.39	7.39
	Discarded	33.12	33.12
	Discarded	46.14	46.14
Controls	NaN	NaN	NaN
	NaN	NaN	NaN
	NaN	NaN	NaN

**FIG. 7C**

**MAST3 cut and KI repair template (vertical line designates guide cut-site)**

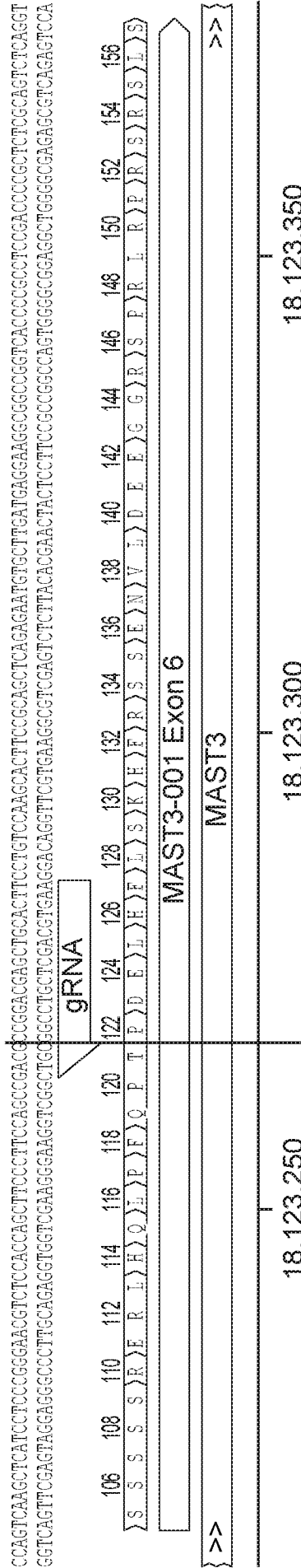


FIG. 8A

The repair template will locate the sequence based on flanking homology arms and will insert the core sequence between the cut-site

GTCAAAGCTCATTCCTCCCGGGAAAGCTCTCCACACAGCTTCCCTTCAGAGCCGACGGtgattgatcaaatgaCCGGACGAGCTGCACCTTCCGTCCCAAGCACTTCCGGCAGCTCAGAGAAATGTGCTTGA TCAGGGAAGGGGGCCGGTCAACCCGGCCTCCGACC  
CAGTTCAGTAGGAGGGCCCTTGCAGAGGTGGTCGAAGGGAGGTCGGGTCGactaaactagttactGGCTGCTCGAGCTGAAGGACAGGTTCCGTGAGGGCTCGACTCTTTACACGACTACTCCTTCCGCCGGGCAATGGGGGGAGGCTGG

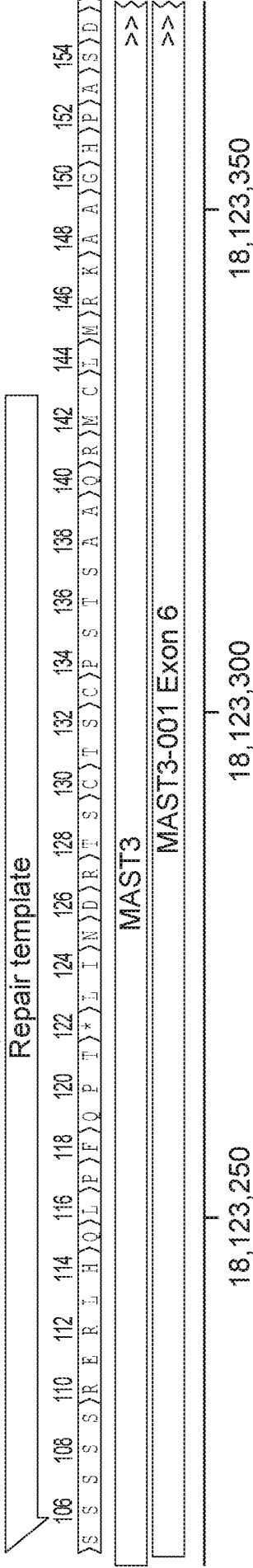
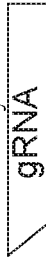
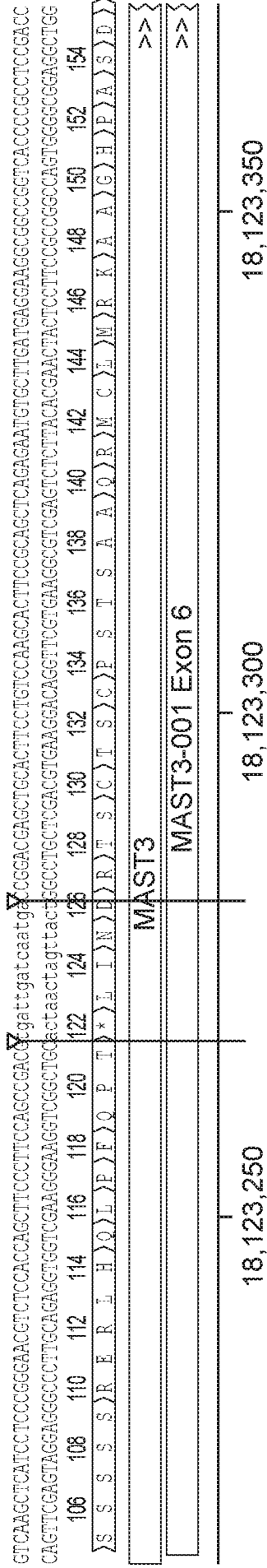


FIG. 8B

After cut and successful KI of the repair template, MAST3, exon 6, will look like the following:



The guide will no longer recognize the cut-site because the core sequence has been inserted and there is a stop codon in the first reading frame.

FIG. 8C



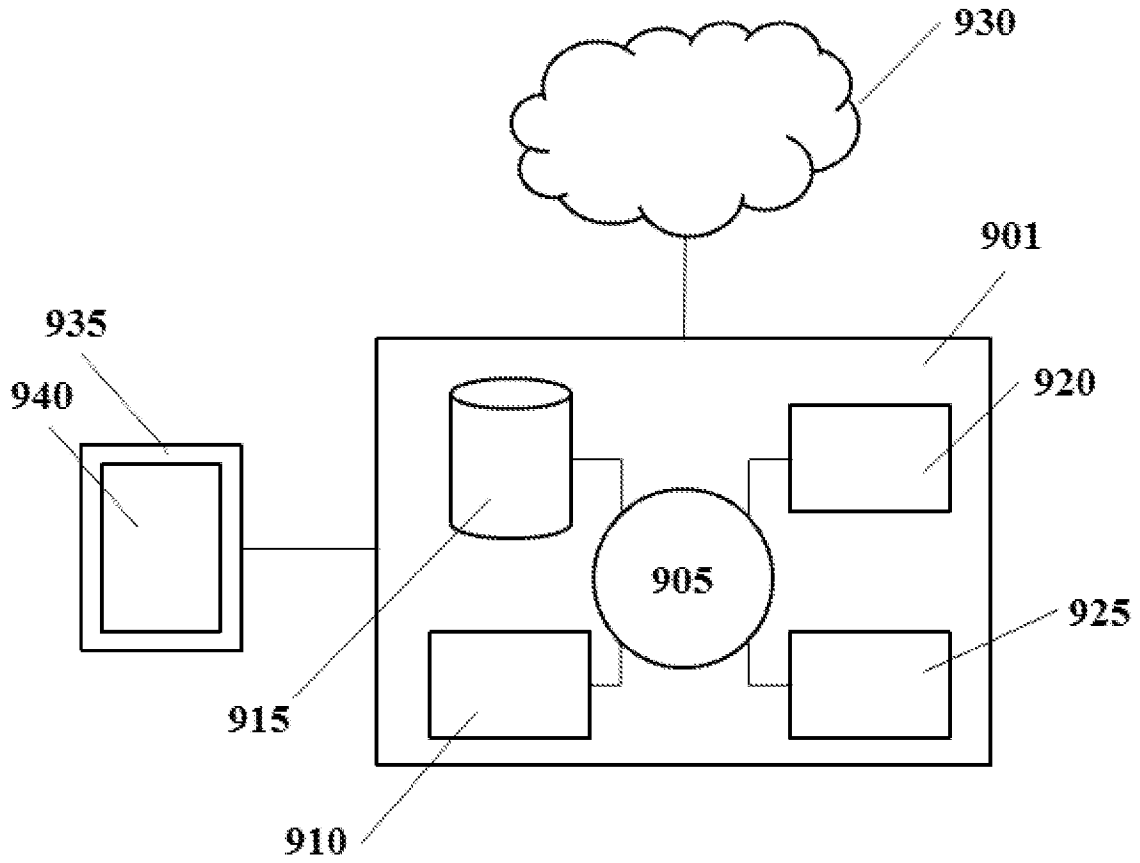


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2019/059557

A. CLASSIFICATION OF SUBJECT MATTER  
IPC(8) - A01H 1/00; C12N 5/00; C12N 5/10; C12N 15/00; C12N 15/85 (2020.01)  
CPC - C12N 9/00; C12N 9/22; C12N 15/00; C12N 15/10; C12N 15/113; C12N 15/66; C12N 15/79; C12N 2310/00; C12N 2310/122; C12N 2310/531 (2020.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)  
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
USPC - 435/6.1; 514/43; 514/44; 536/22.1 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2017/0198302 A1 (THE CHINESE UNIVERSITY OF HONG KONG) 13 July 2017	1, 2, 36, 70, 71
--	(13.07.2017) entire document	-----
Y		3, 37, 38, 72
Y	JP 4243680 B2 (NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE & TECHNOLOGY) 25 March 2009 (25.03.2009) entire document	3, 37, 38, 72
P, X	WO 2019/204668 A1 (CASEBIA THERAPEUTICS LIMITED LIABILITY PARTNERSHIP) 24 October 2019 (24.10.2019) entire document	1-3, 36-38, 70-72
A	WO 2014/093595 A1 (THE BROAD INSTITUTE, INC. et al) 19 June 2014 (19.06.2014) entire document	1-3, 36-38, 70-72
A	WO 2018/067826 A1 (CELLULAR DYNAMICS INTERNATIONAL, INC.) 12 April 2018 (12.04.2018) entire document	1-3, 36-38, 70-72
A	US 2012/0142051 A1 (OGAWA et al) 07 June 2012 (07.06.2012) entire document	1-3, 36-38, 70-72

Further documents are listed in the continuation of Box C.  See patent family annex.

\* Special categories of cited documents:  
 "A" document defining the general state of the art which is not considered to be of particular relevance  
 "E" earlier application or patent but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed  
 "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
 "&" document member of the same patent family

Date of the actual completion of the international search  
22 January 2020

Date of mailing of the international search report  
**20 FEB 2020**

Name and mailing address of the ISA/US  
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, VA 22313-1450  
Facsimile No. 571-273-8300

Authorized officer  
Blaine R. Copenheaver  
PCT Helpdesk: 571-272-4300  
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2019/059557

**Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2.  Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3.  Claims Nos.: 4-35, 39-69, 73-84  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
  - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
  - No protest accompanied the payment of additional search fees.