

(12) UK Patent

(19) GB

(11) 2616539

(13) B

(45) Date of B Publication

20.12.2023

(54) Title of the Invention: A non-toxic cas9 enzyme and application thereof

(51) INT CL: C12N 15/90 (2006.01) C12N 15/62 (2006.01)

(21) Application No: 2307873.6

(22) Date of Filing: 06.01.2020

Date Lodged: 25.05.2023

(30) Priority Data:

(31) 62789347	(32) 07.01.2019	(33) US
(31) 62823477	(32) 25.03.2019	(33) US
(31) 62824164	(32) 26.03.2019	(33) US
(31) 62855612	(32) 31.05.2019	(33) US

(62) Divided from Application No  
2110538.2 under section 15(9) of the Patents Act 1977

(43) Date of A Publication 13.09.2023

(72) Inventor(s):  
Christopher Hackley

(73) Proprietor(s):  
Crisp-HR Therapeutics, Inc.  
(Incorporated in USA - California)  
3108 Patrick Henry Drive, Santa Clara 95054,  
California, United States of America

(74) Agent and/or Address for Service:  
Avidity IP  
Broers Building, Hauser Forum, 21 J J Thomson Ave,  
CAMBRIDGE, Cambridgeshire, CB3 0FA,  
United Kingdom

(56) Documents Cited:

WO 2016/057961 A1 WO 2016/007604 A1  
US 20150071898 A1  
Genesis, 2017, 55(8), 10.1002/dvg.23044, Clements et  
al., "RICE CRISPR: Rapidly increased cut ends by an  
exonuclease Cas9 fusion in zebrafish"

(58) Field of Search:

As for published application 2616539 A viz:  
INT CL C07K, C12N  
Other: WPI, EPODOC, INTERNET, BIOSIS, MEDLINE  
updated as appropriate

Additional Fields  
Other: None

GB  
2616539  
B

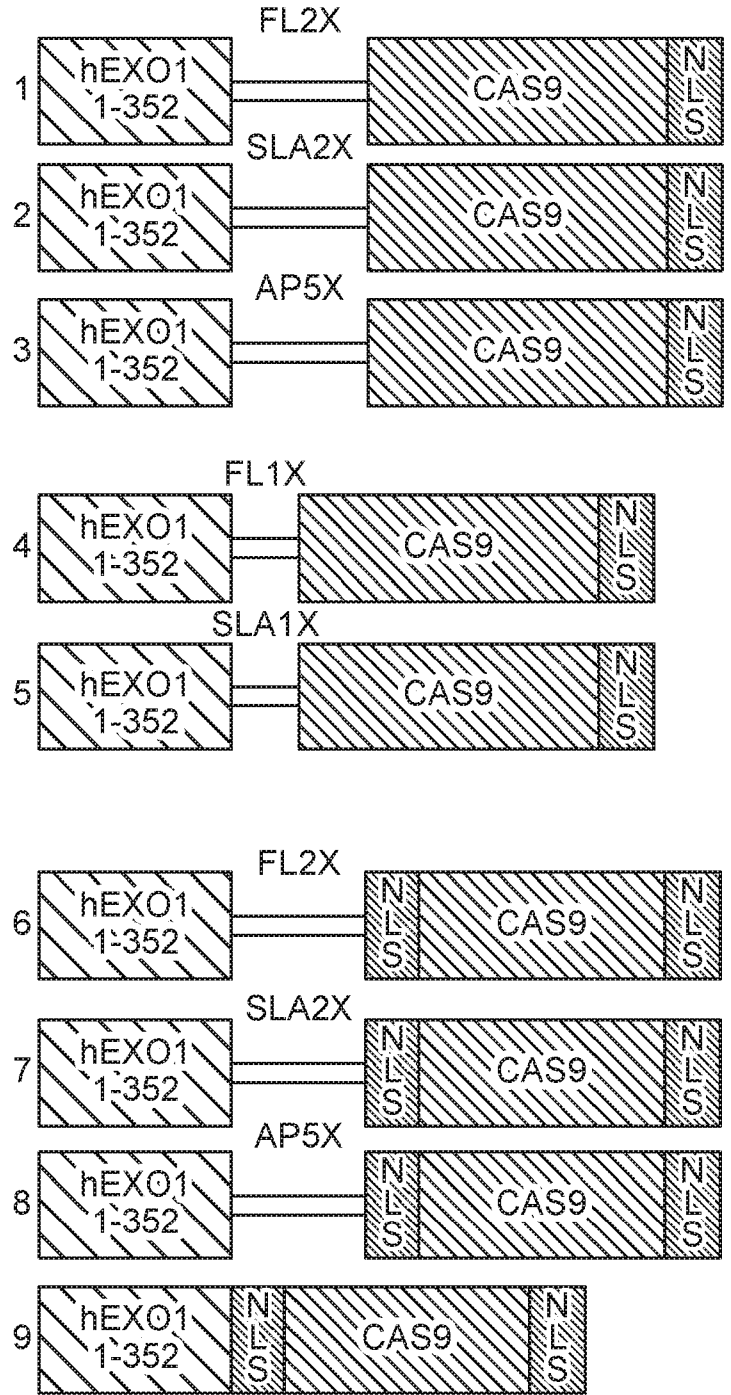


FIG. 1

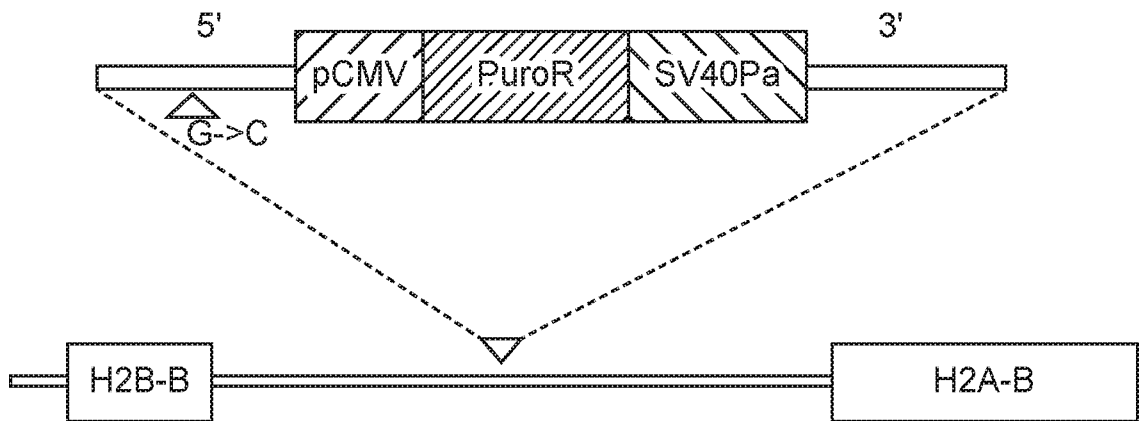


FIG. 2

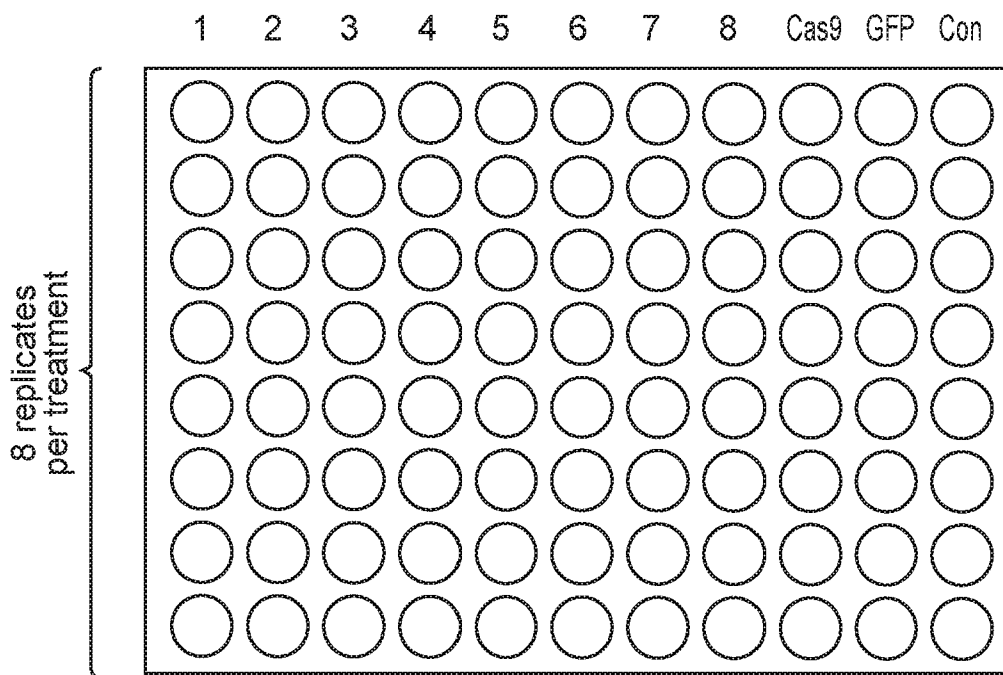


FIG. 3

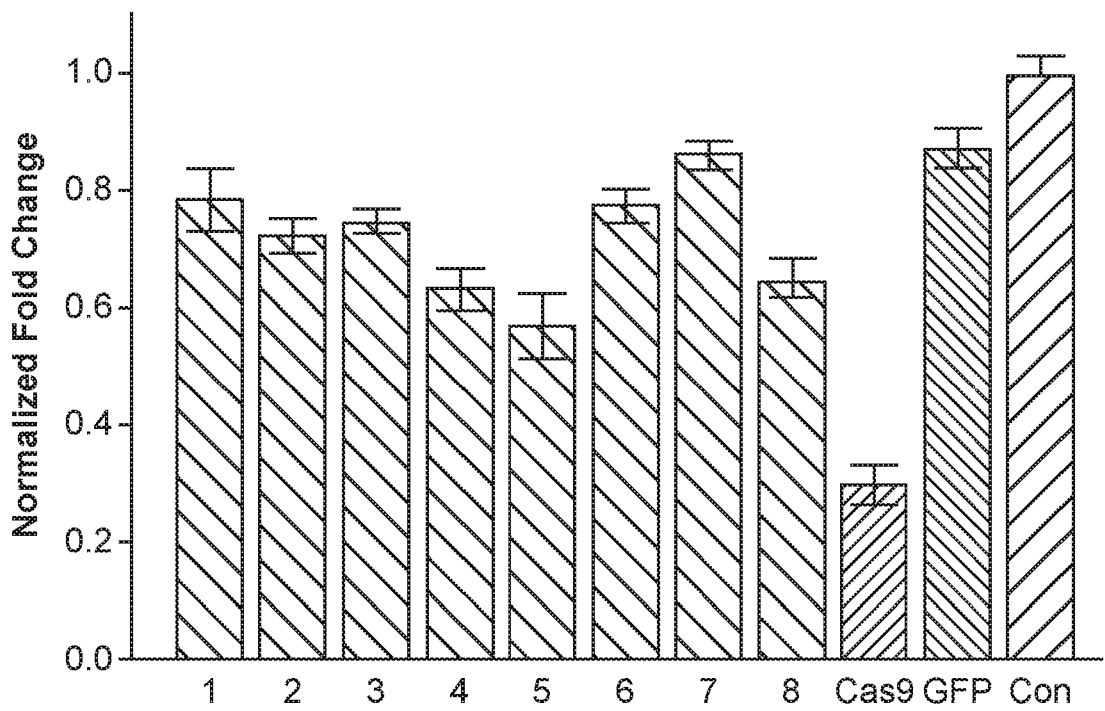


FIG. 4

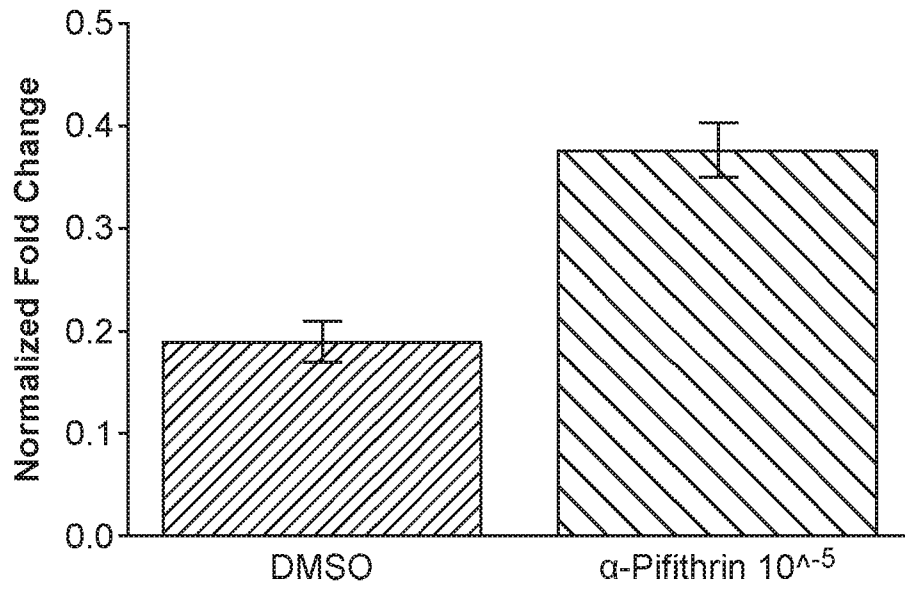


FIG. 5

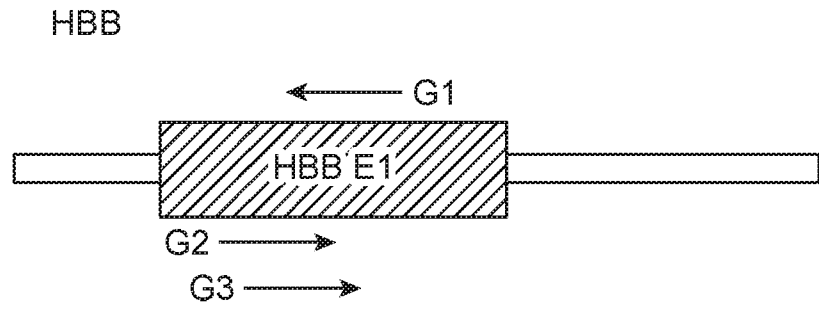


FIG. 6A

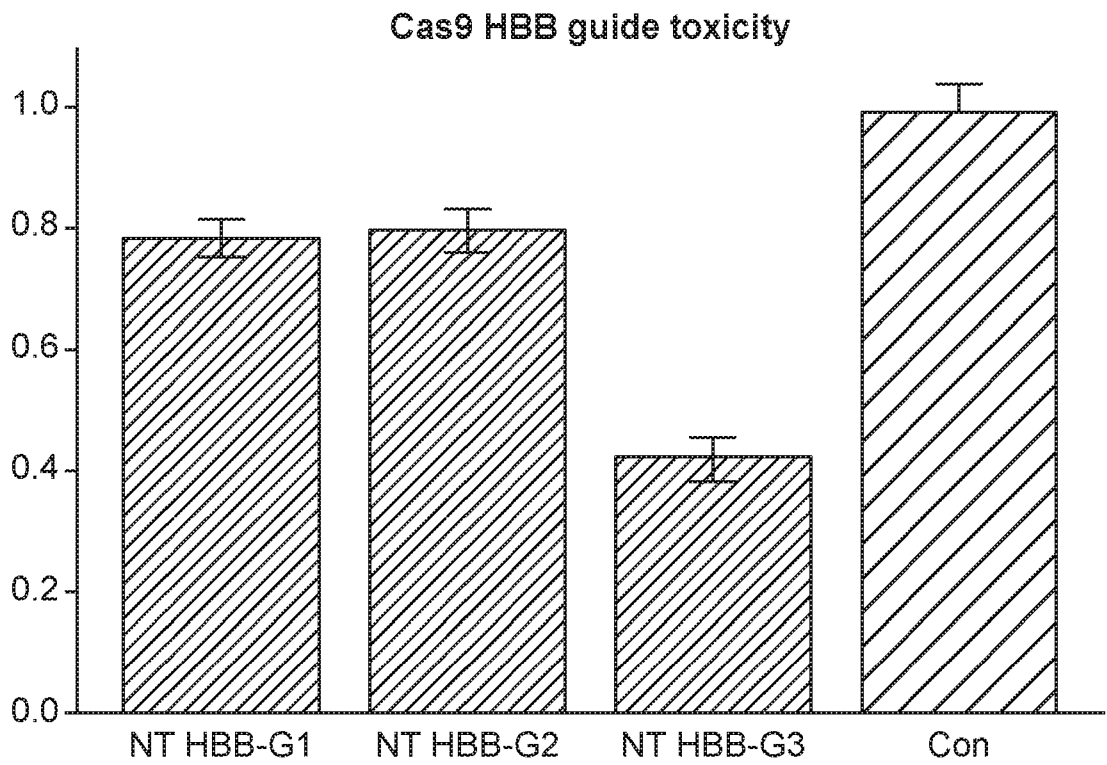
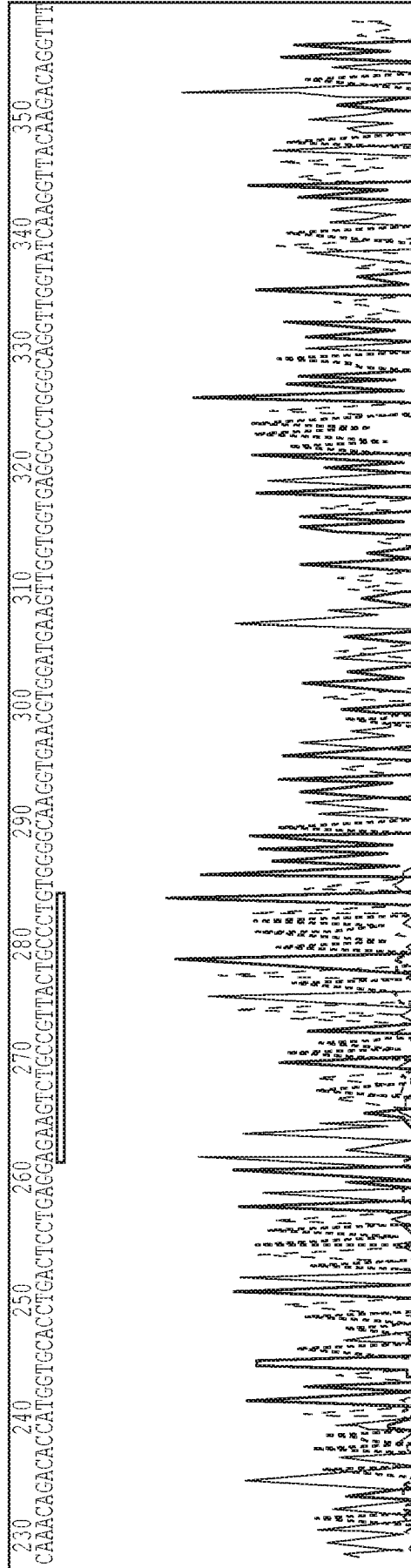


FIG. 6B

**Cas9 HBB-G3 reverse sequence trace**



**FIG. 6C**



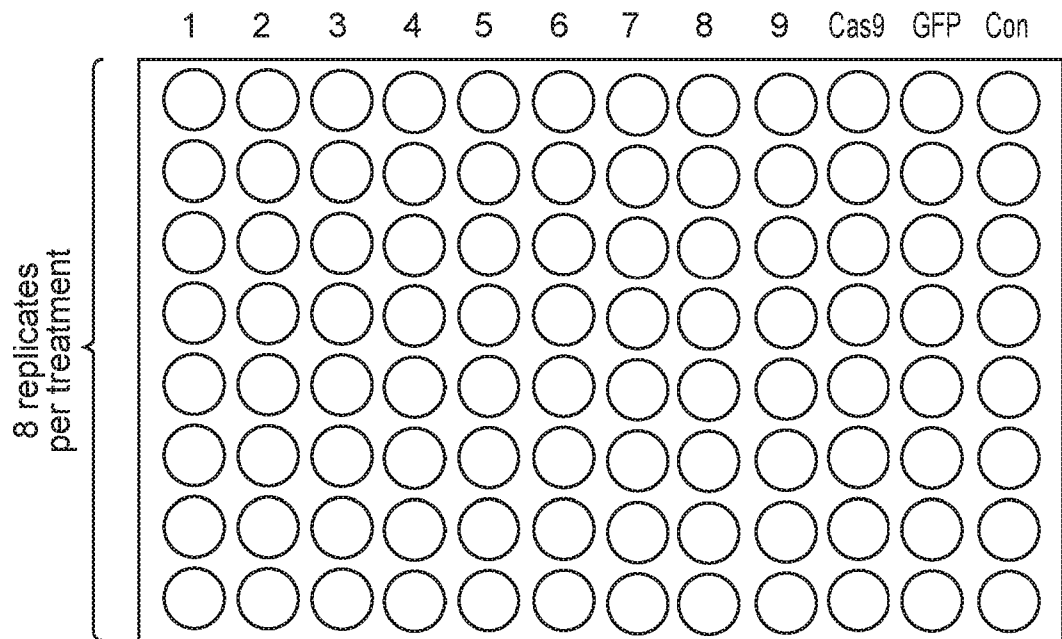


FIG. 7

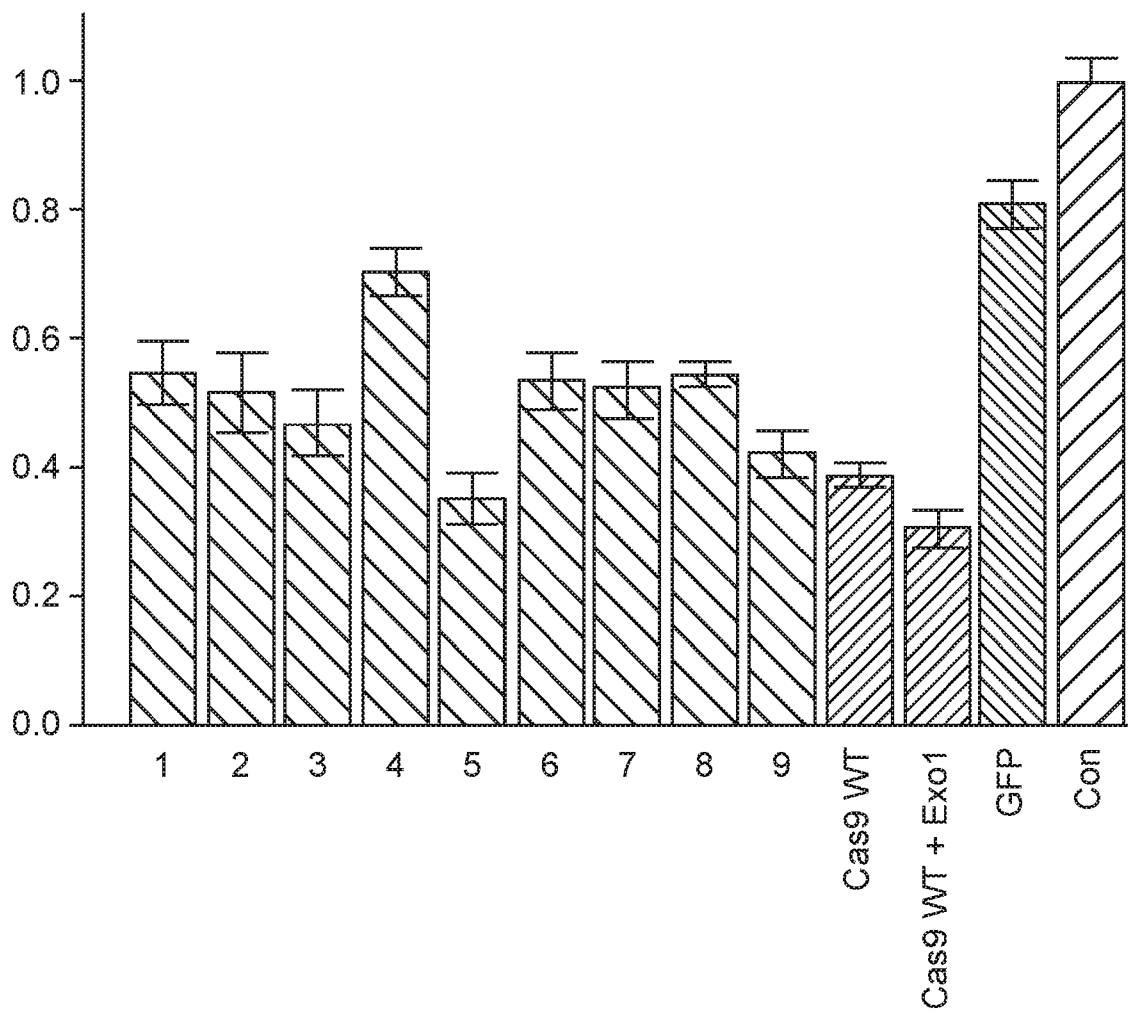


FIG. 8

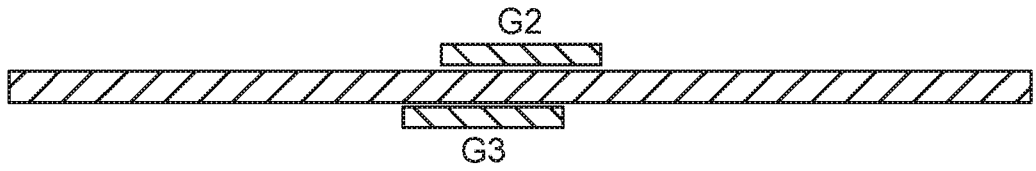


FIG. 9A

Int-G2 Toxicity

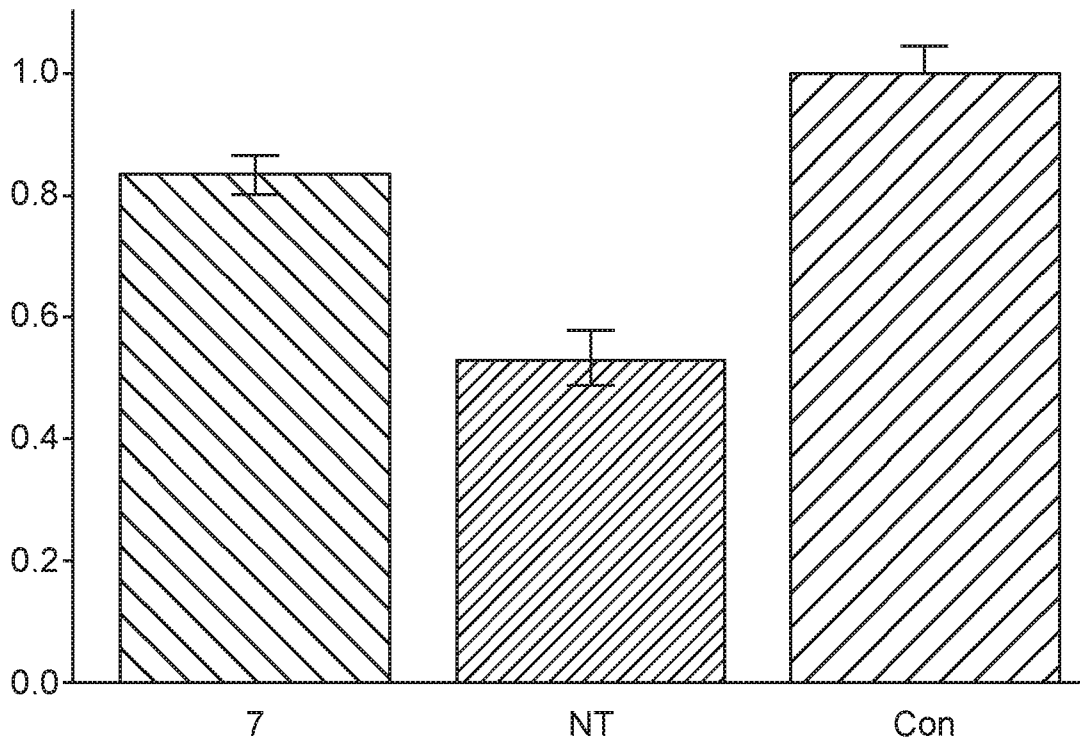


FIG. 9B

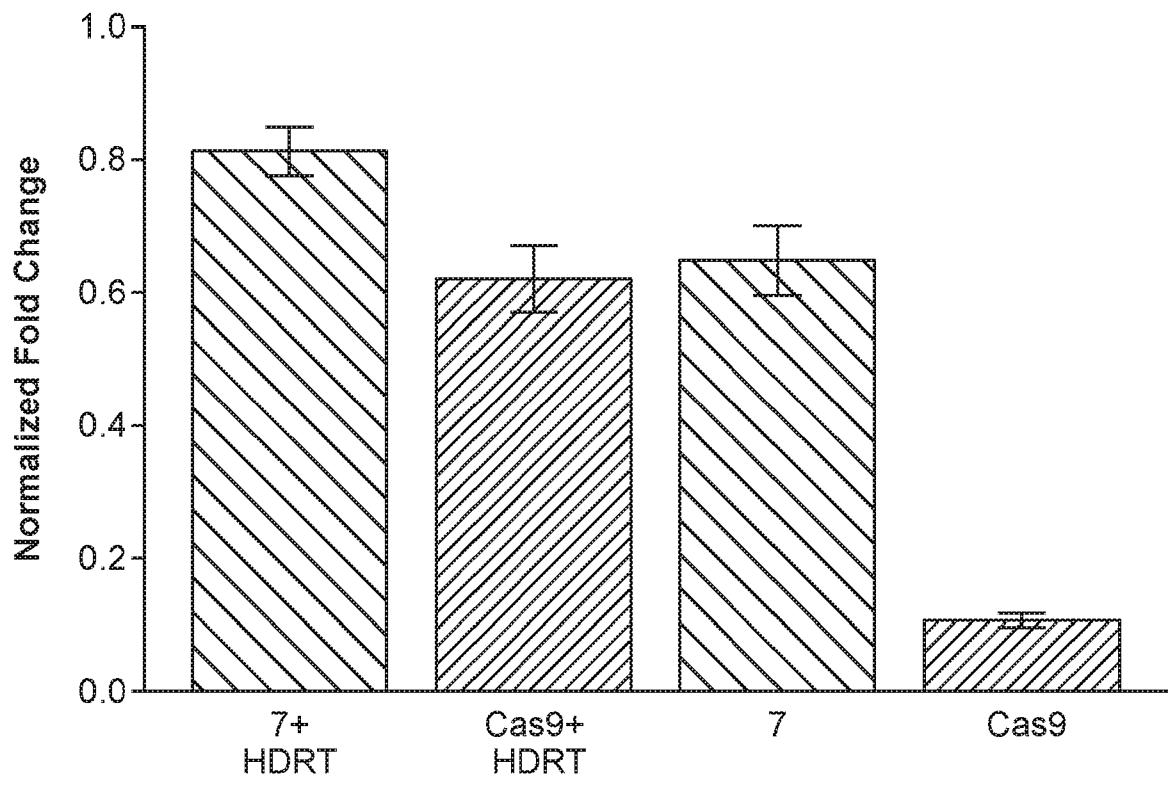


FIG. 10

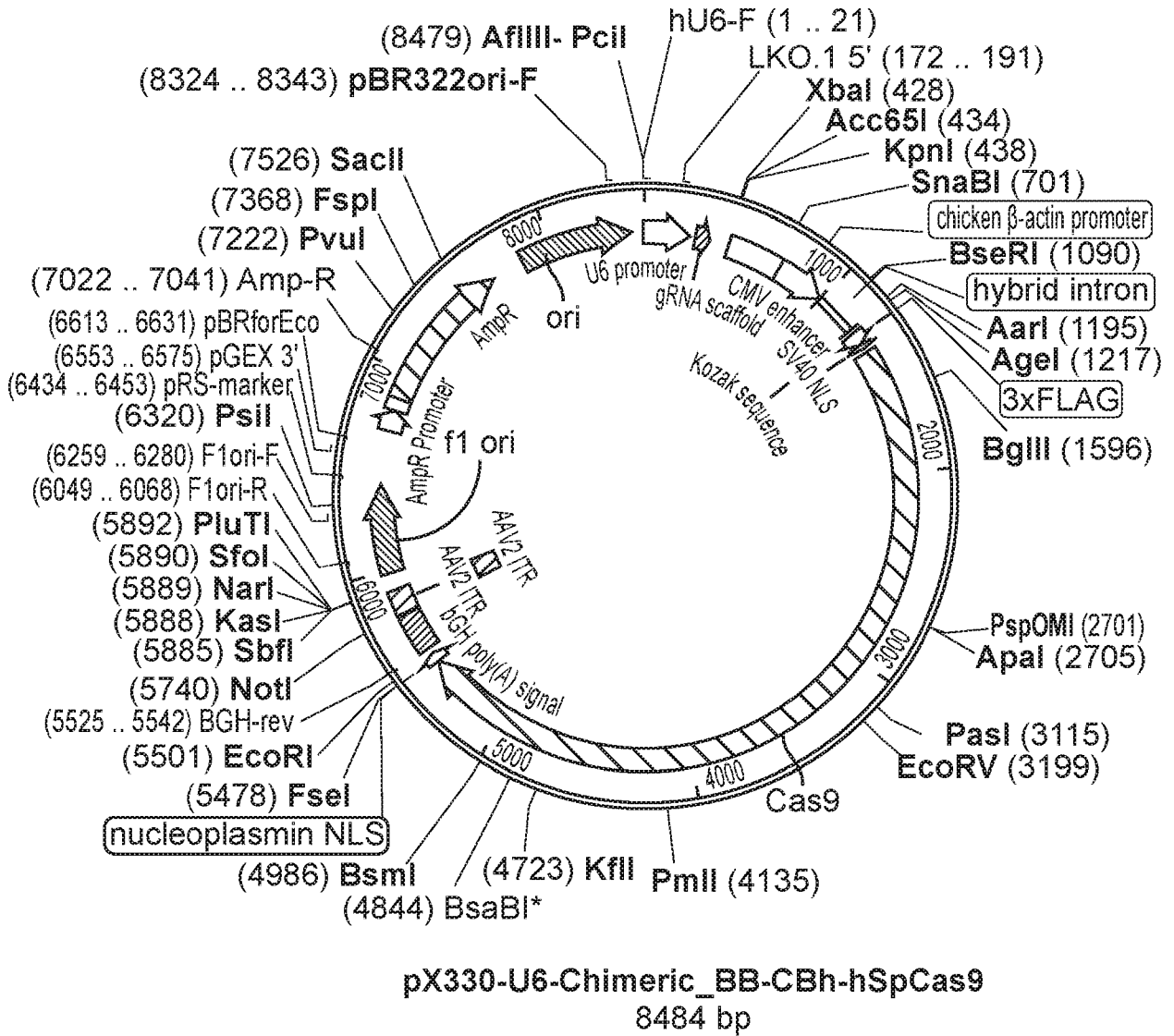


FIG. 11A

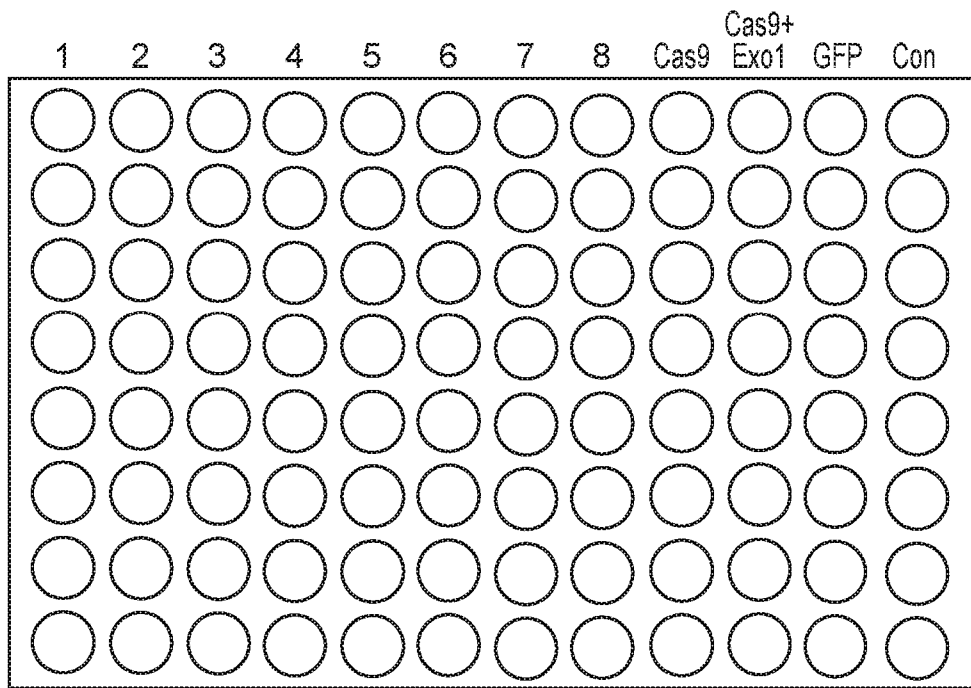


FIG. 11B

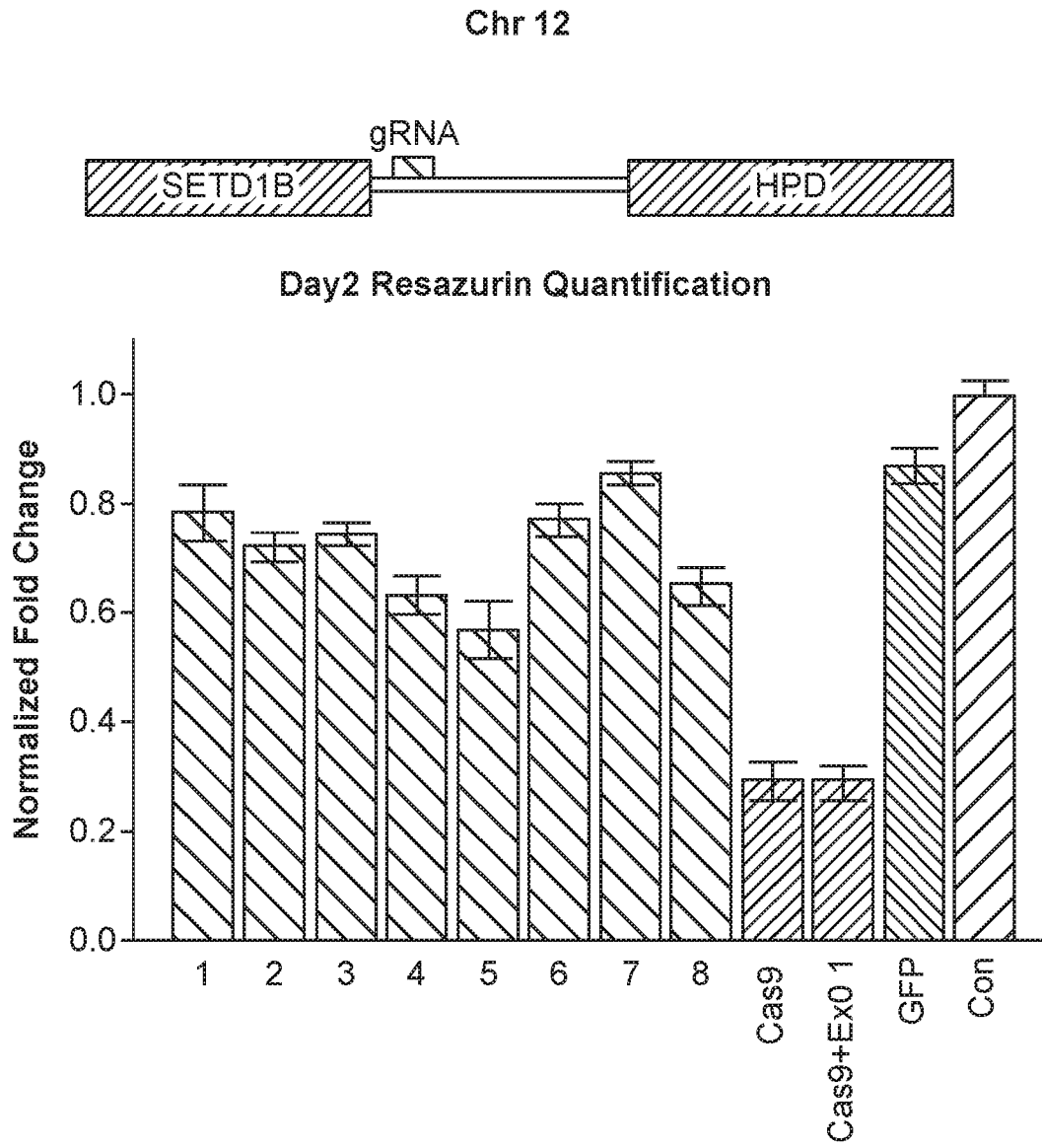


FIG. 11C

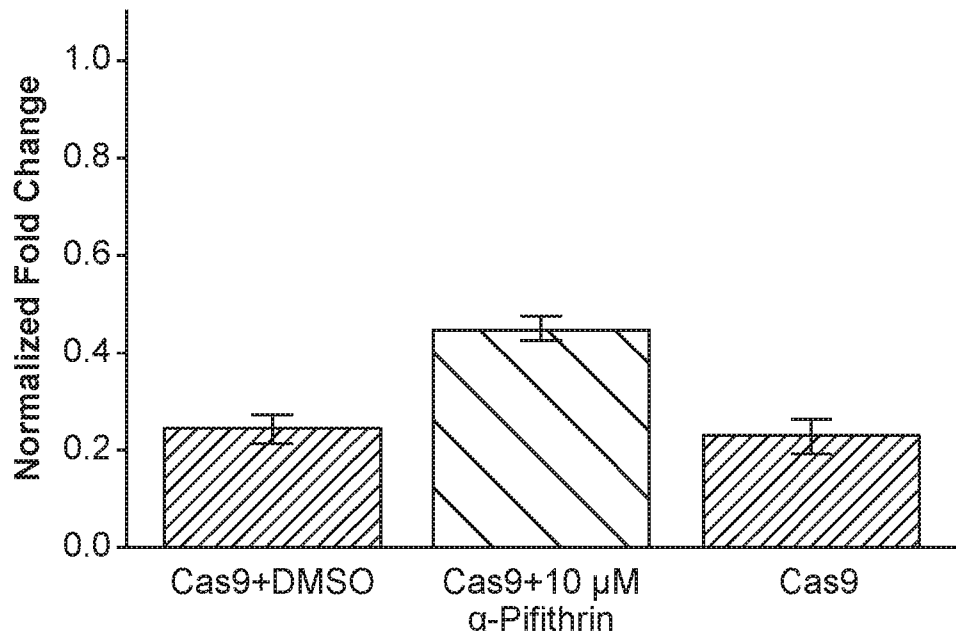


FIG. 11D



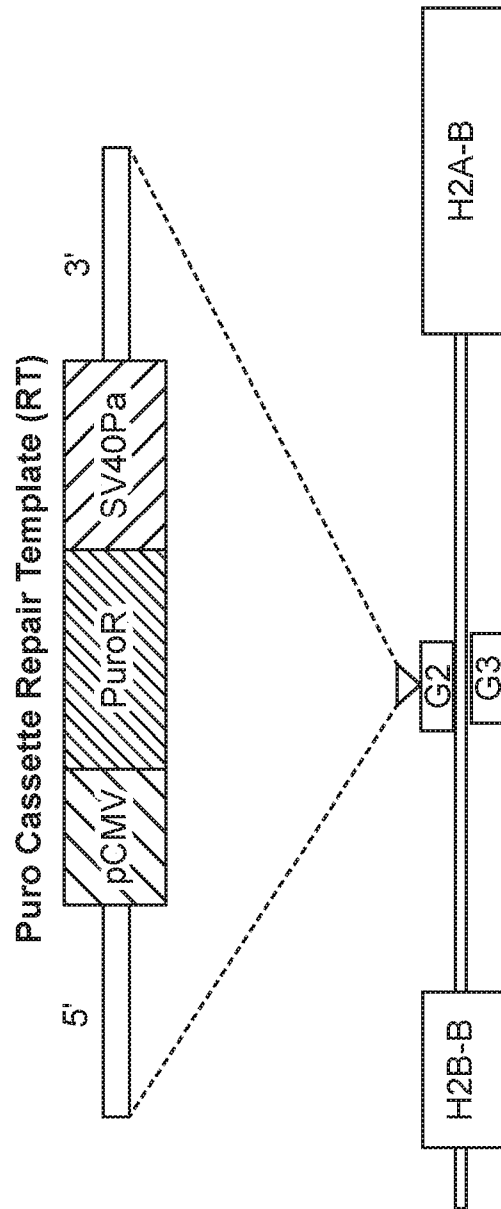
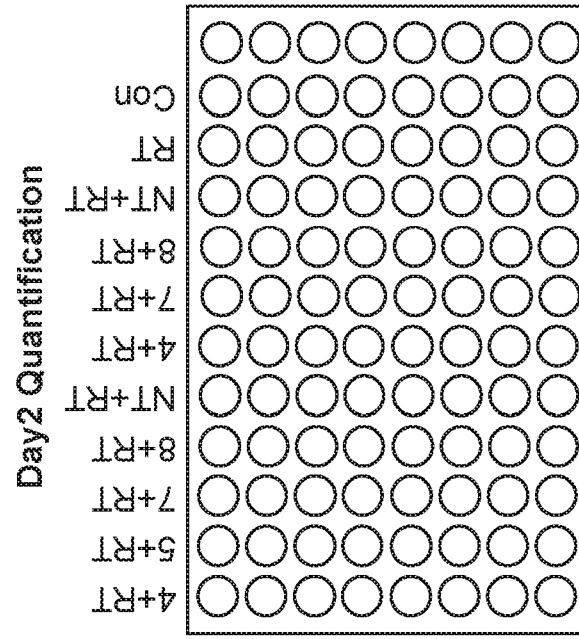
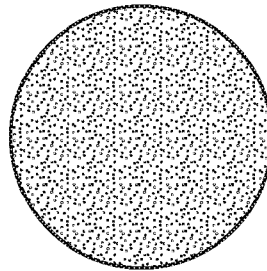


FIG. 12A



A549



Cal-Phos



FIG. 12B

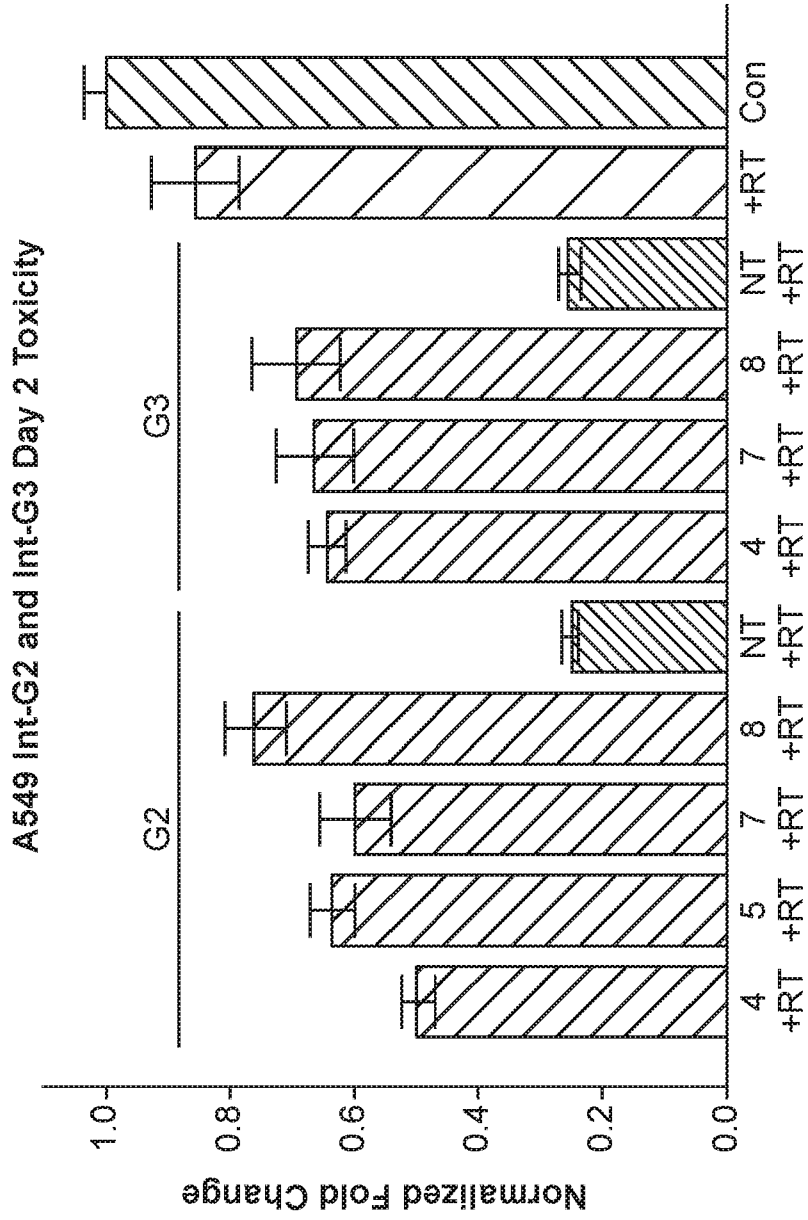
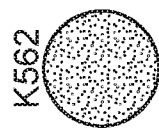
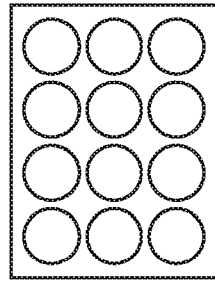


FIG. 12C

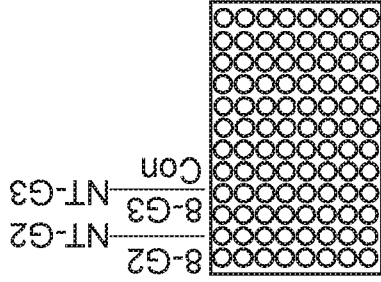


K562

8+RT NT  
+RT  
Electroporation

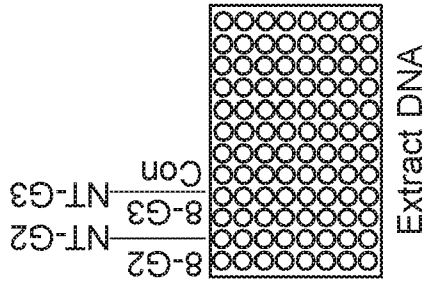


3 Days  
0.5 mg/mL  
Puromycin



Day 3 Puro Quantification

10 Days  
0.5 mg/mL  
Puromycin



Day 10 Puro Quantification

FIG. 12D

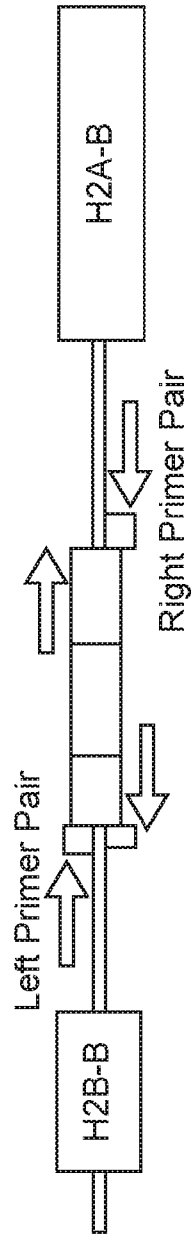


FIG. 12E

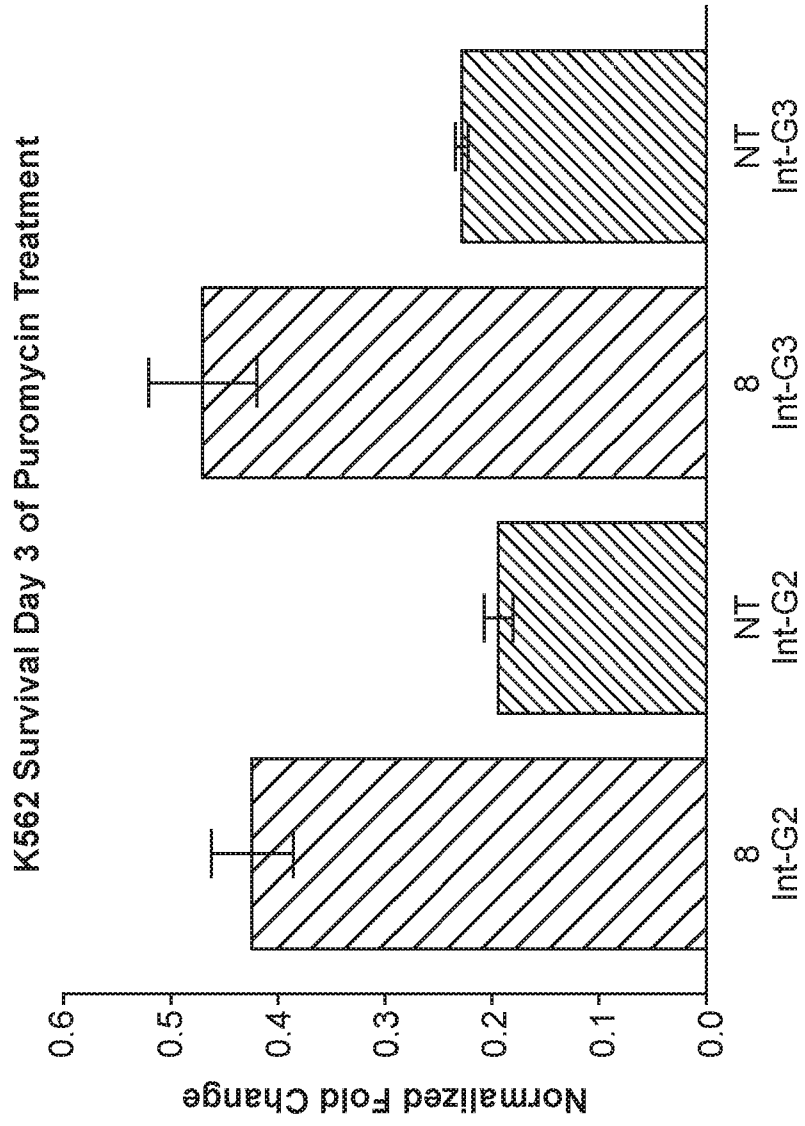


FIG. 12F

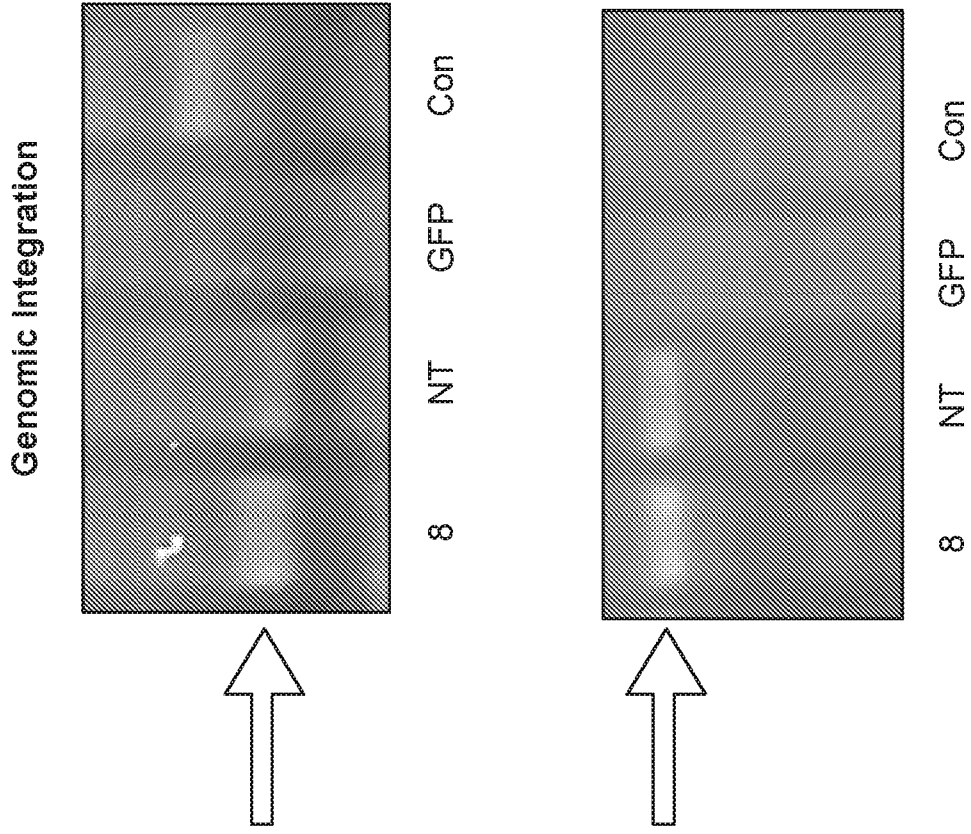


FIG. 12G

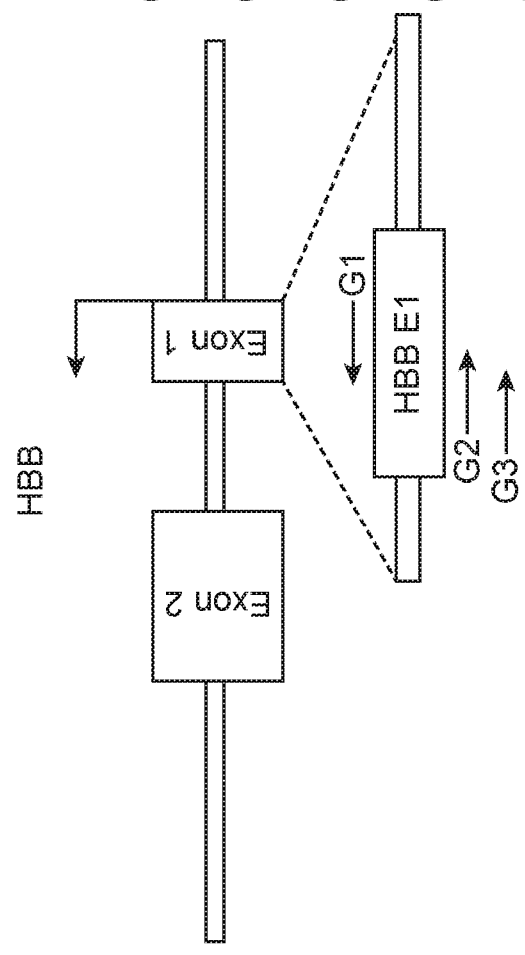
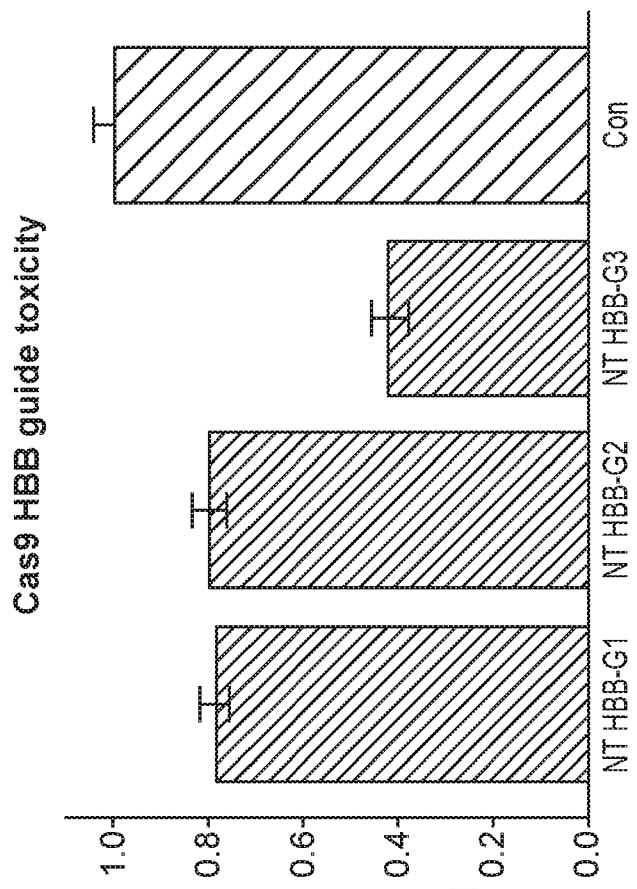
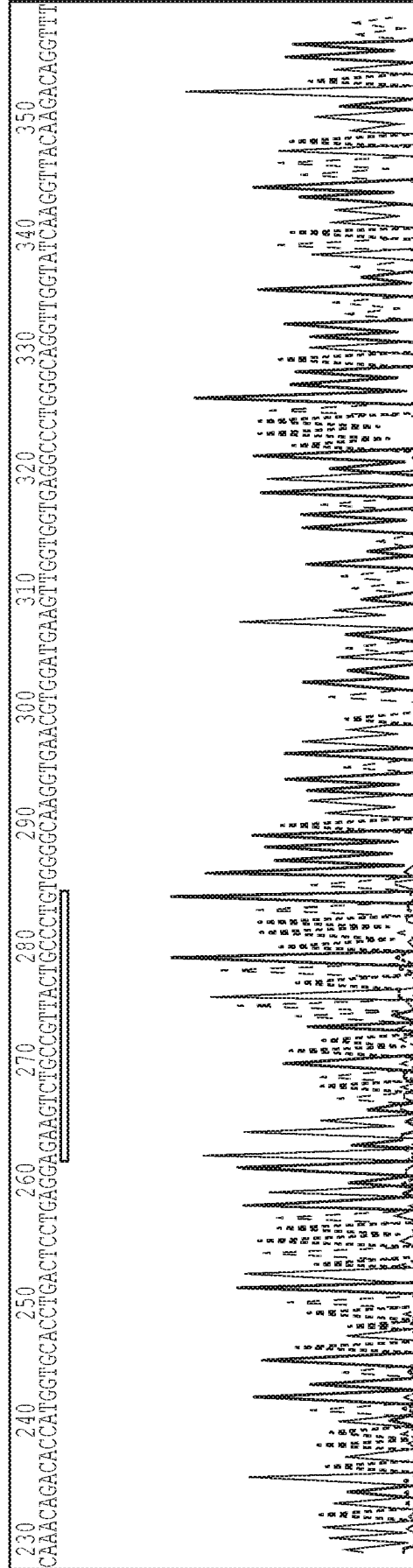


FIG. 13A



**Cas9 HBB-G3 reverse sequence trace**



**FIG. 13B**

```

ctgactcctgaggagaagtcctgcccctgtggggcaaggctgaacgctggatgaagttggtggtga... WT
ctgactcctgtggagaattctgcaagttaactgcaactgtgaggcaaggctgaacgctggaagaggttgggtggtga... SSRT G3
SS EcoRI MM MM
Allele

```

FIG. 13C

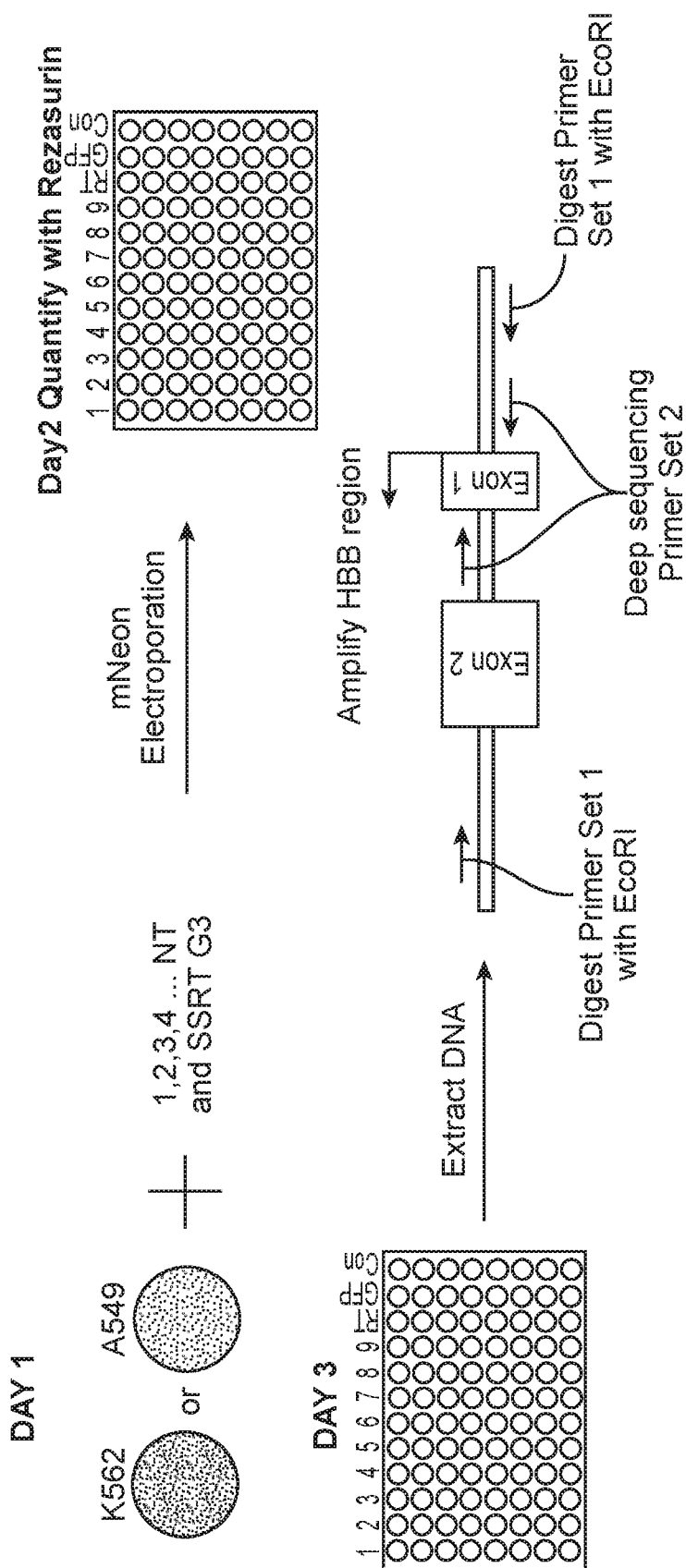
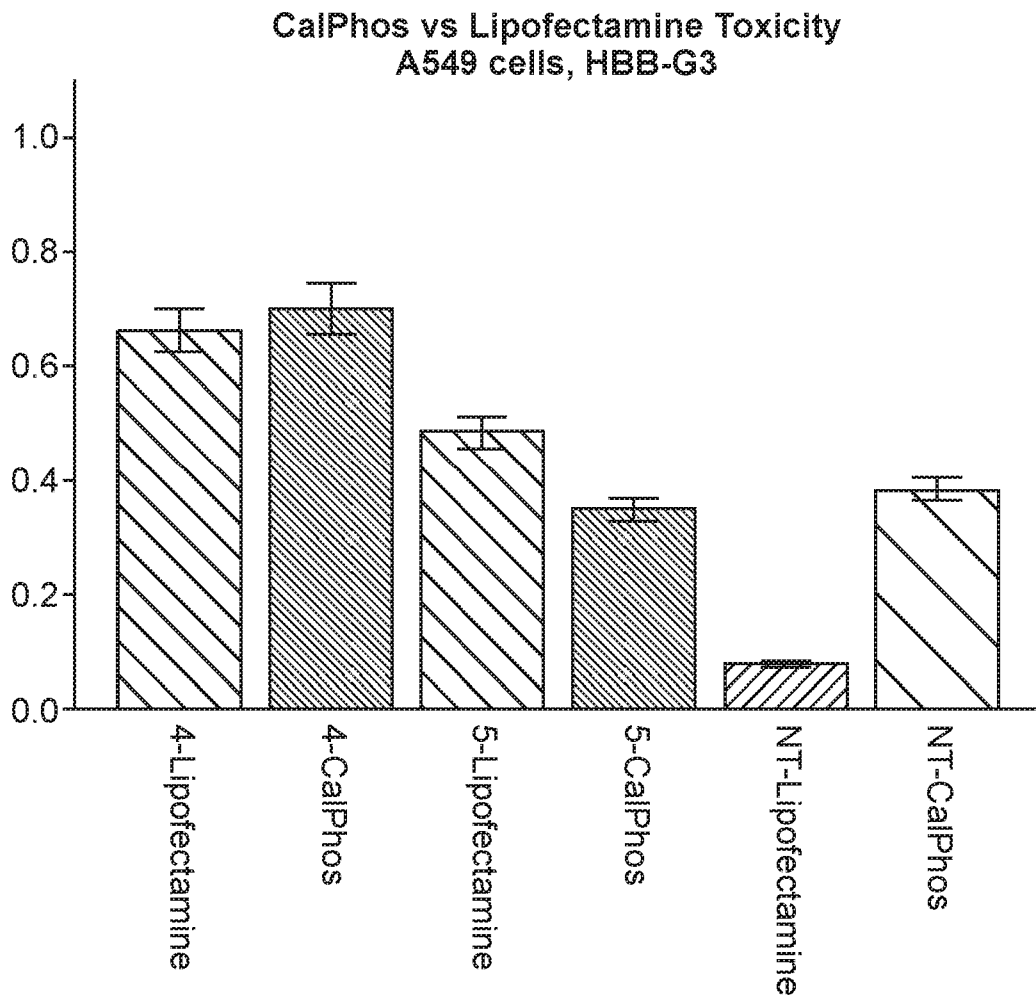


FIG. 13D



**FIG. 14**

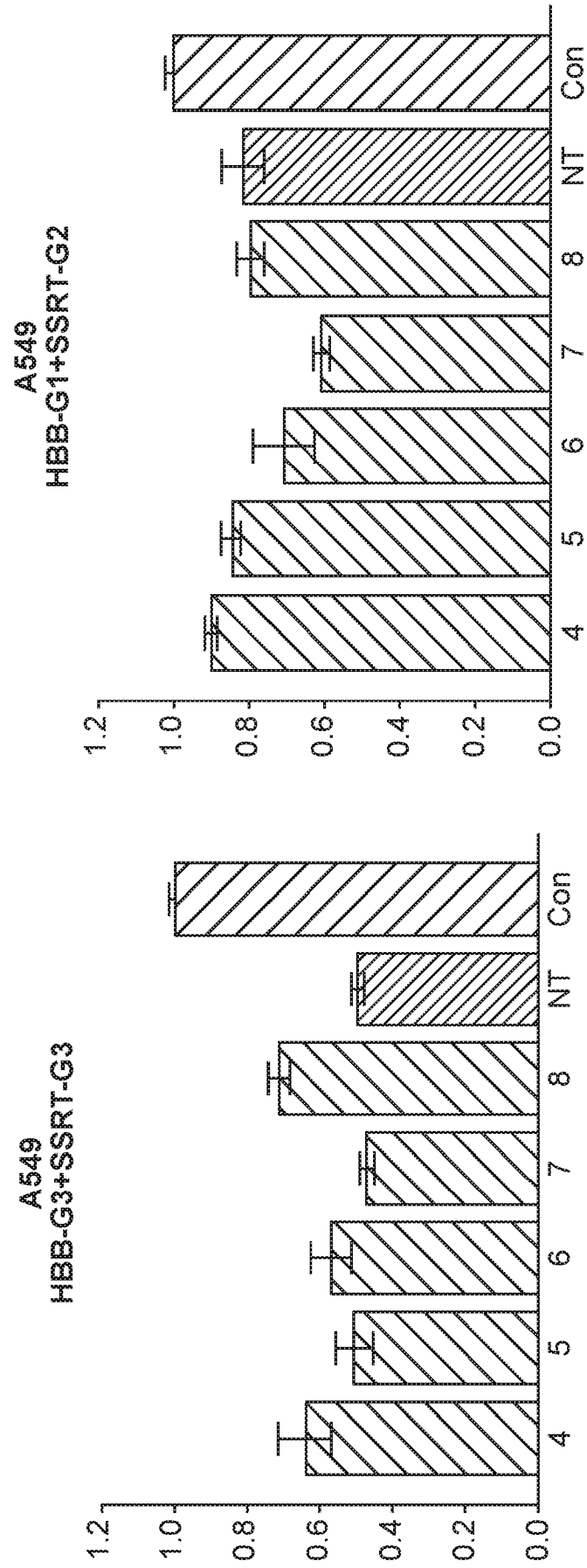


FIG. 15

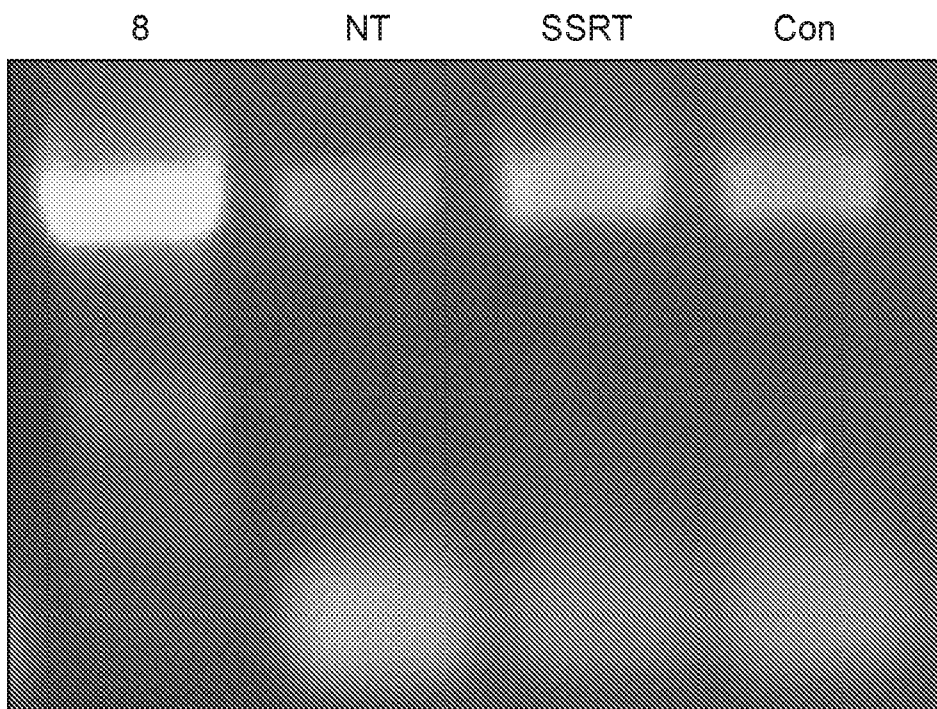


FIG. 16A

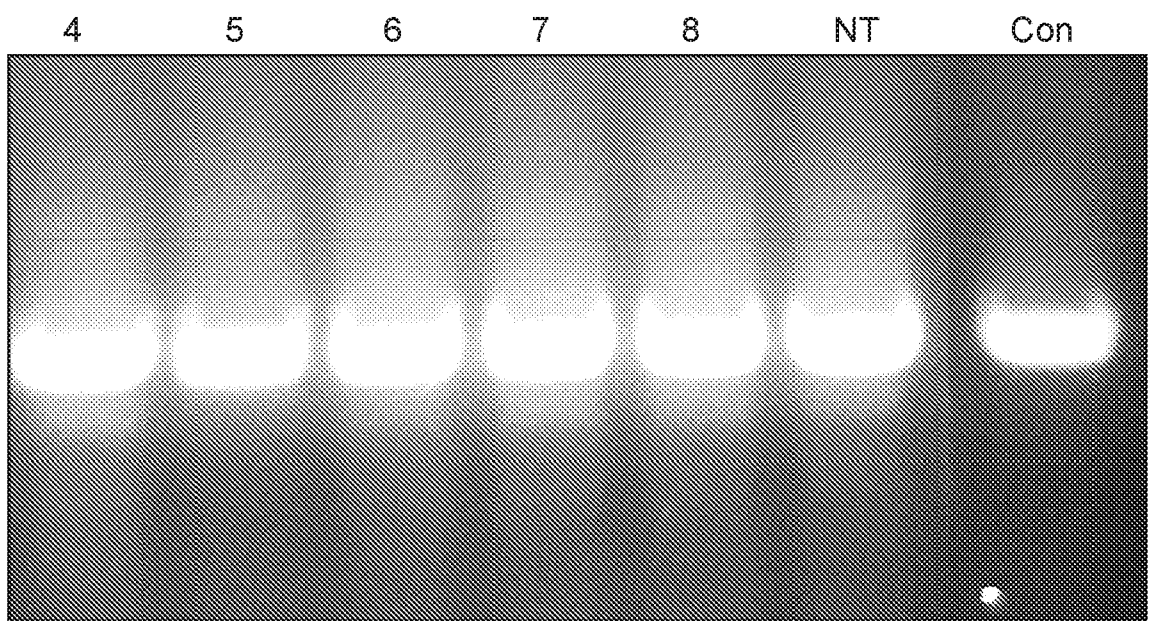


FIG. 16B

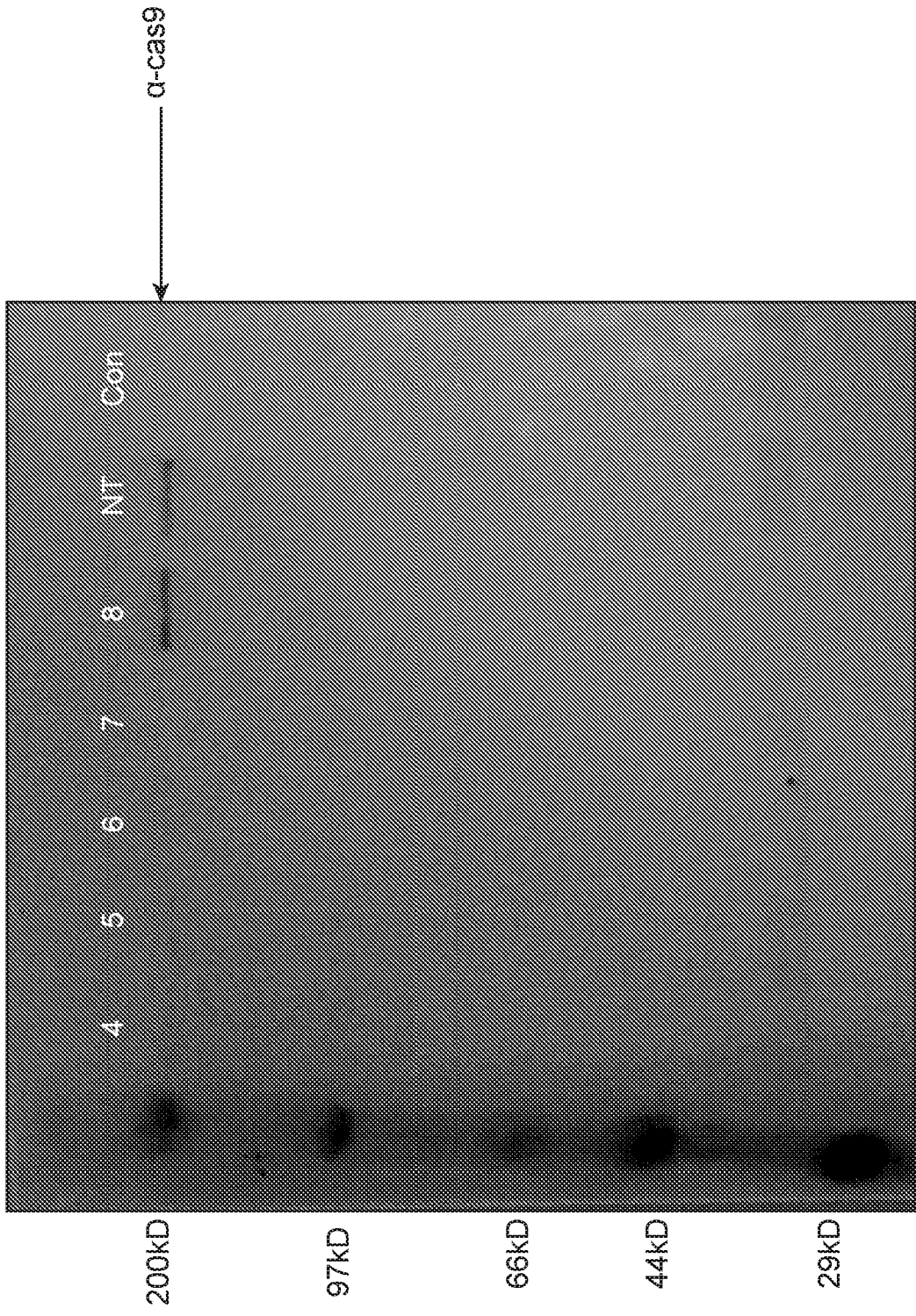


FIG. 16C



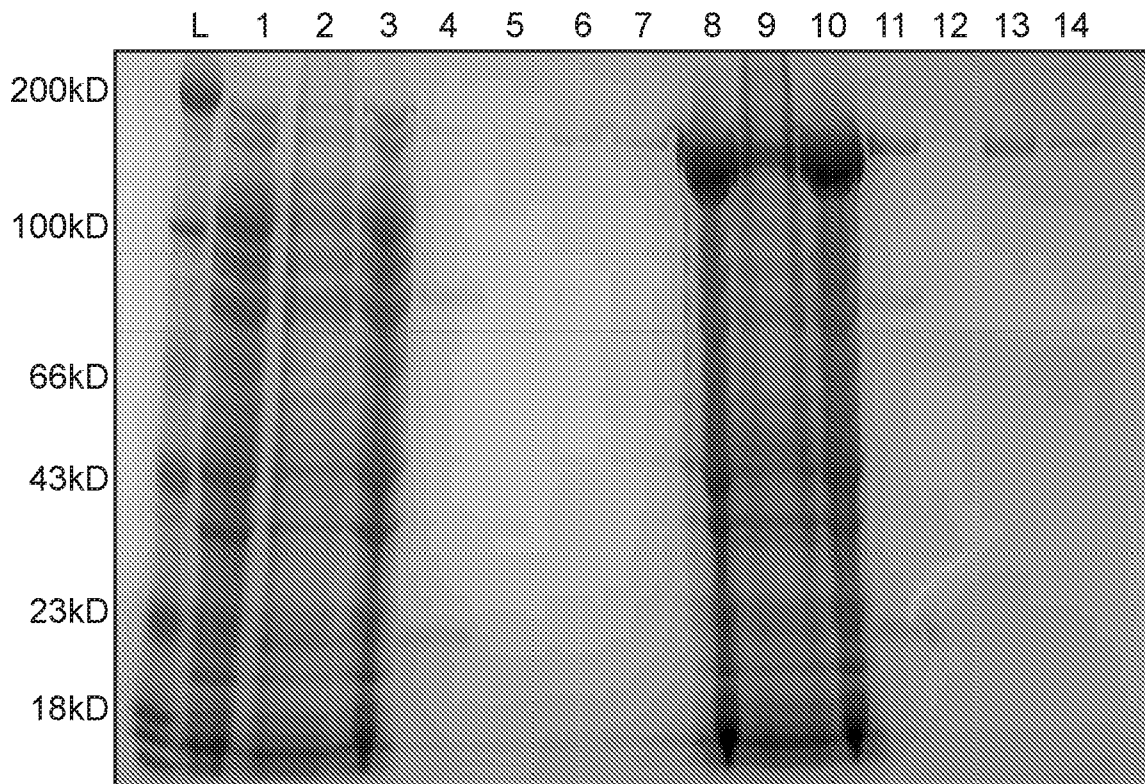


FIG. 16D

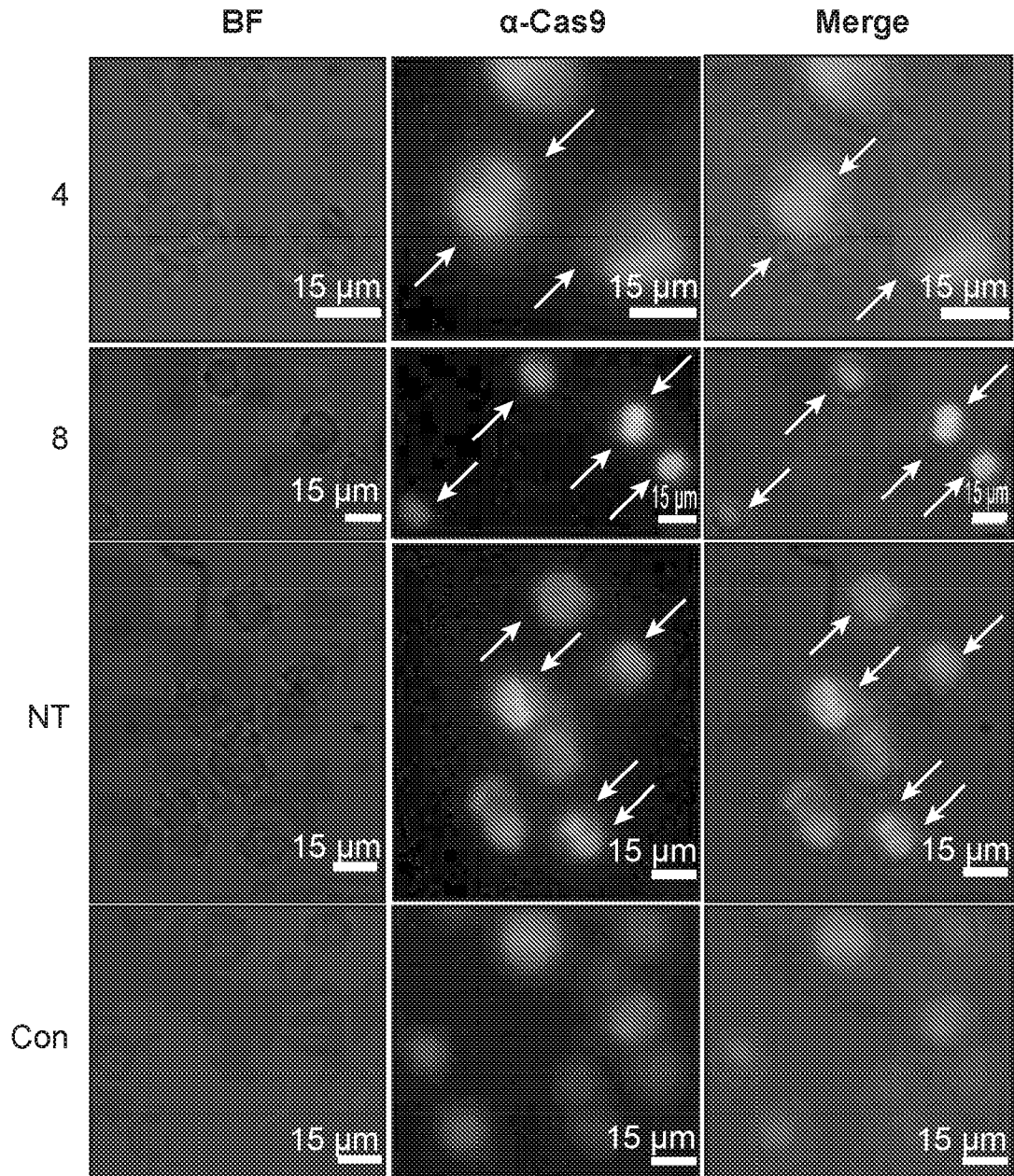


FIG. 16E

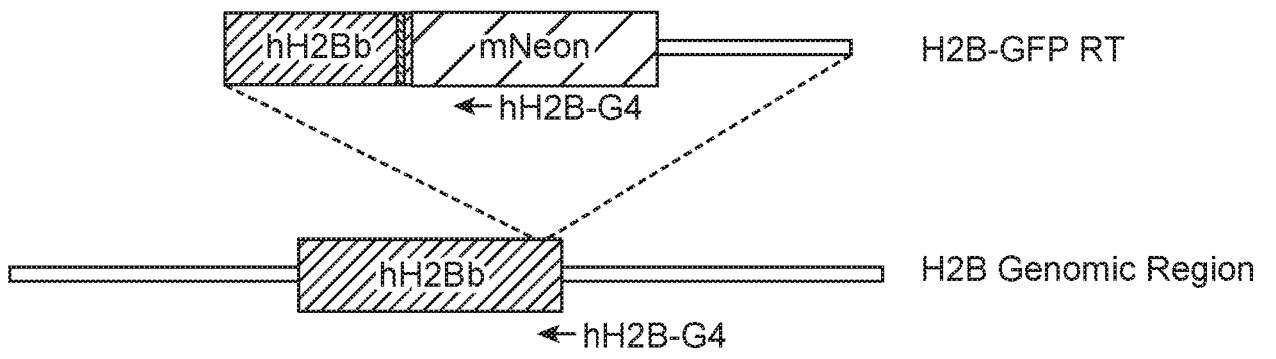


FIG. 17A

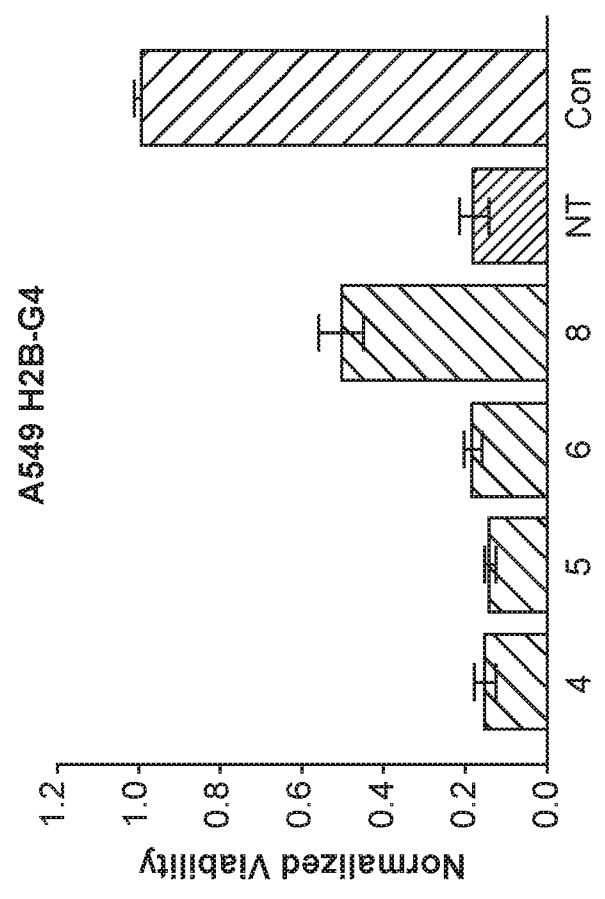
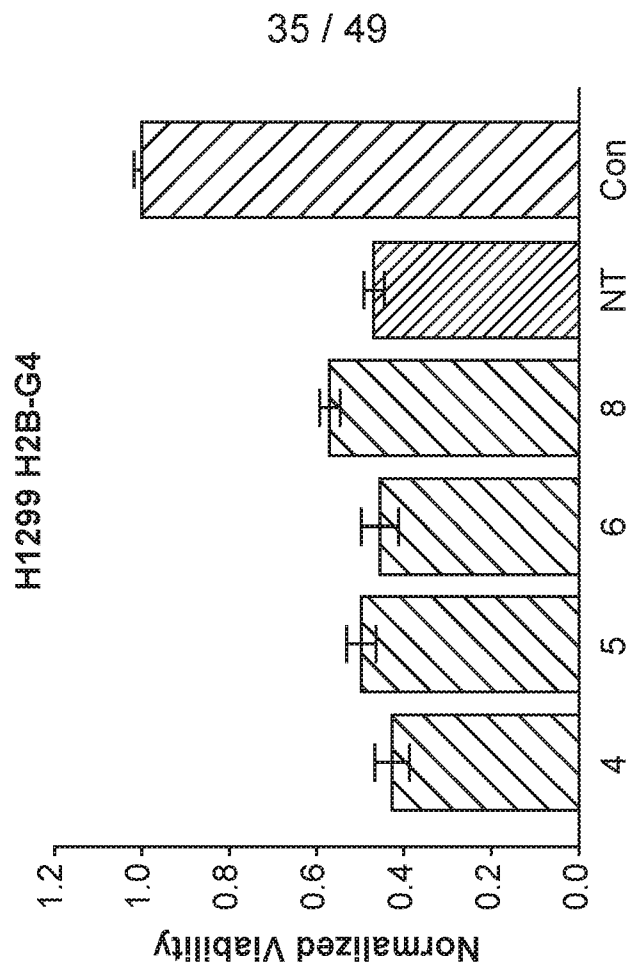


FIG. 17B

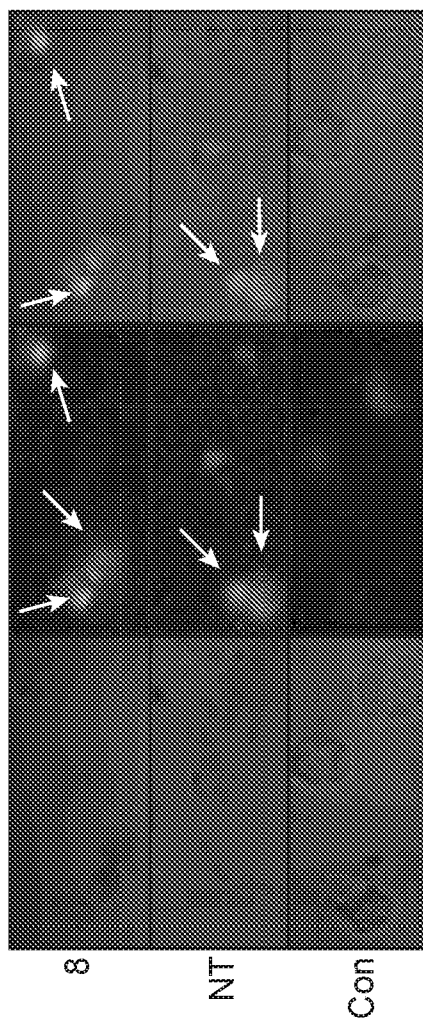
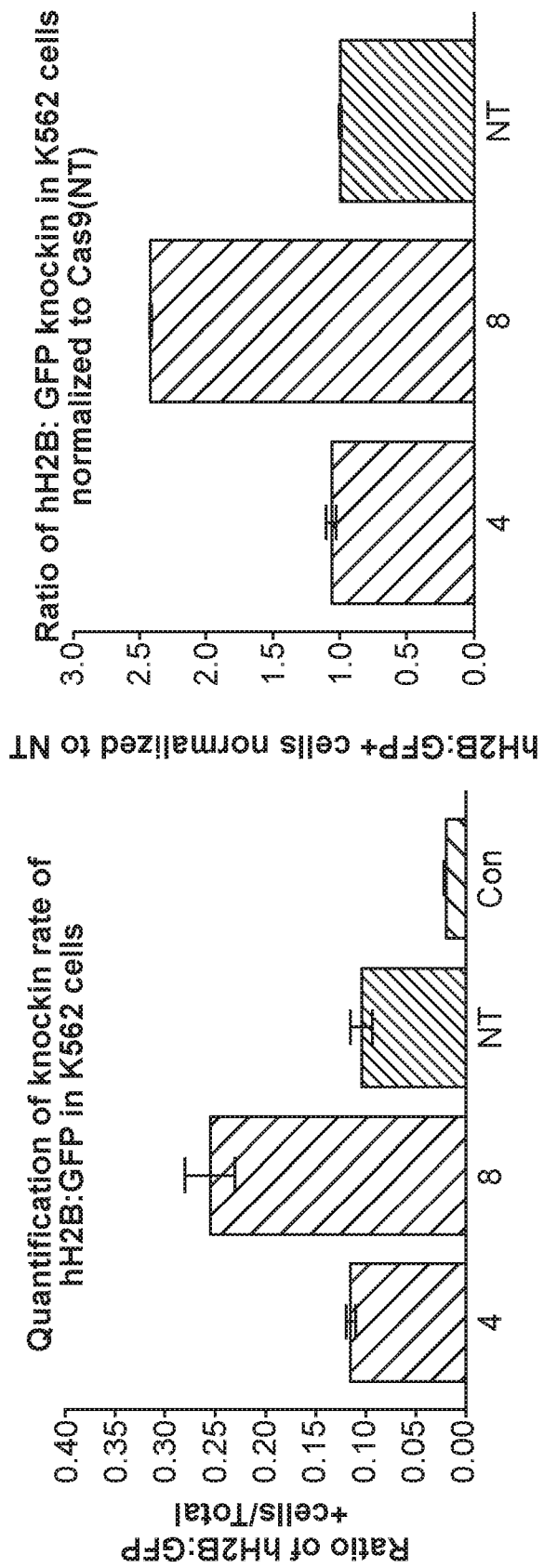


FIG. 17C

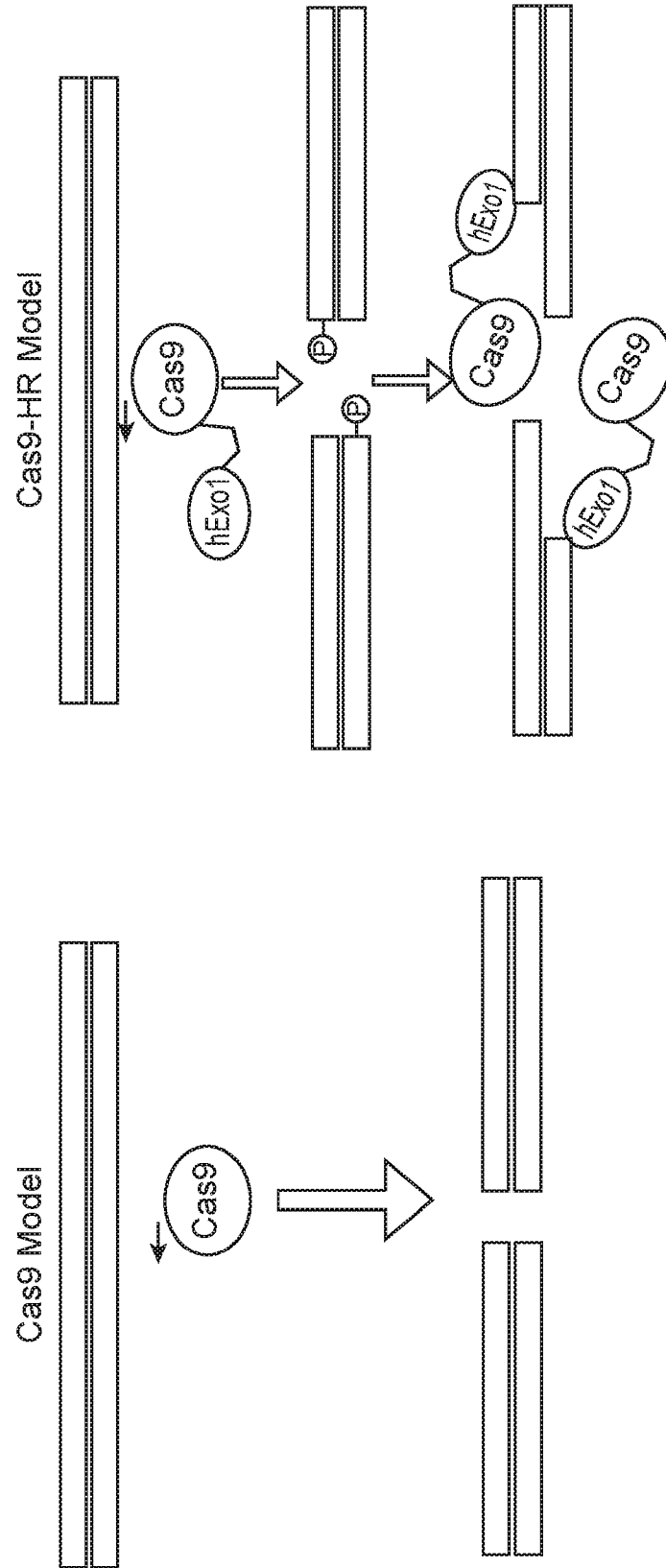


FIG. 18A

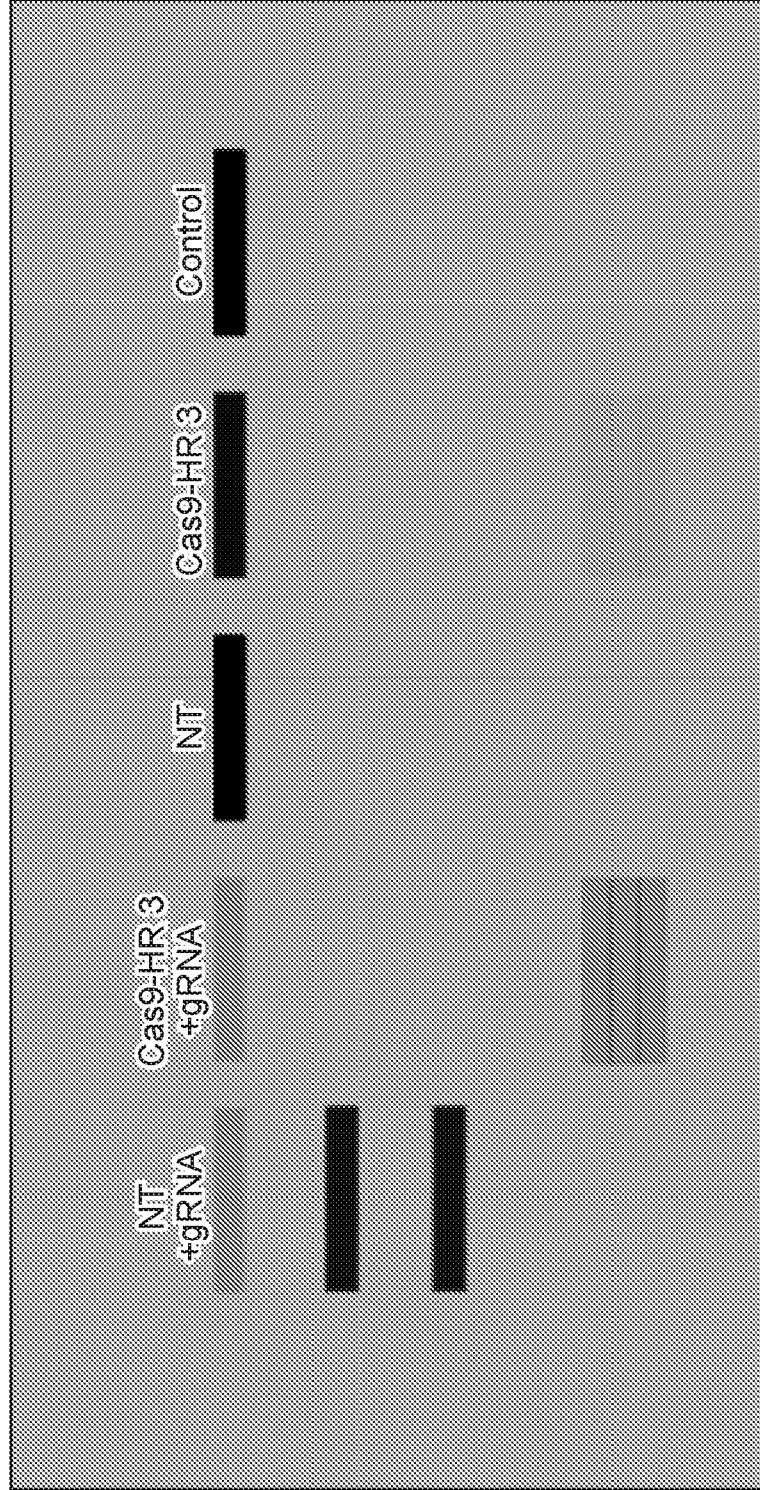


FIG. 18B

1 2 3 4 5

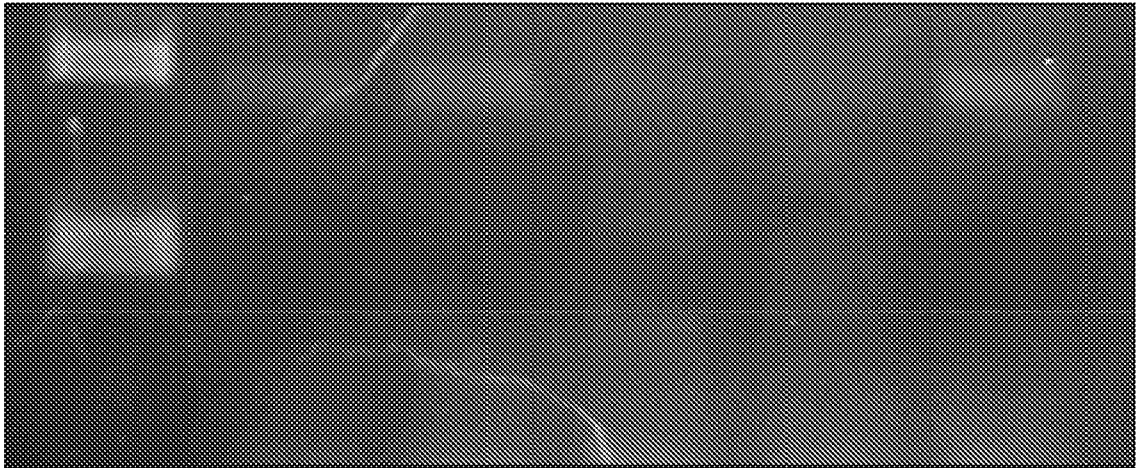


FIG. 18C



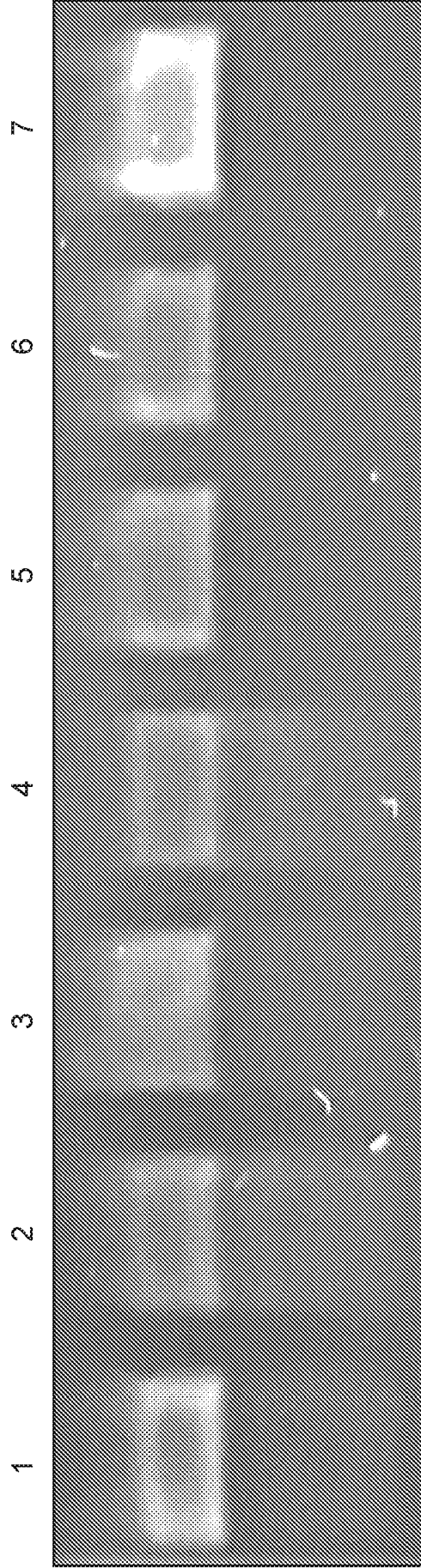


FIG. 18D



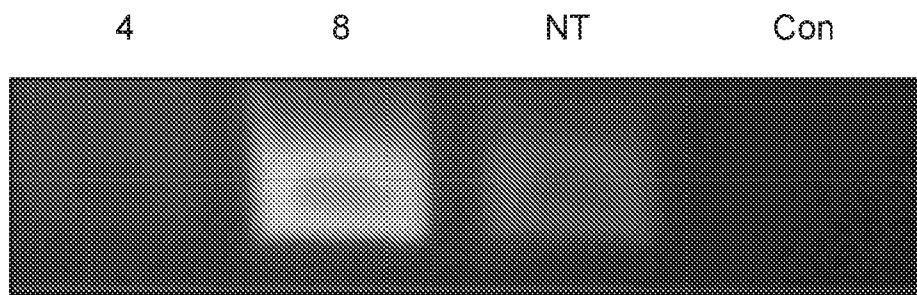


FIG. 19B

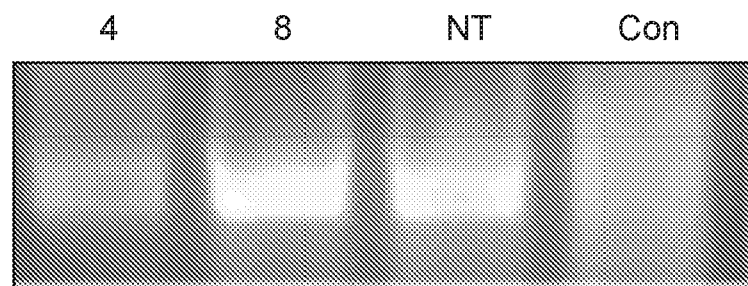
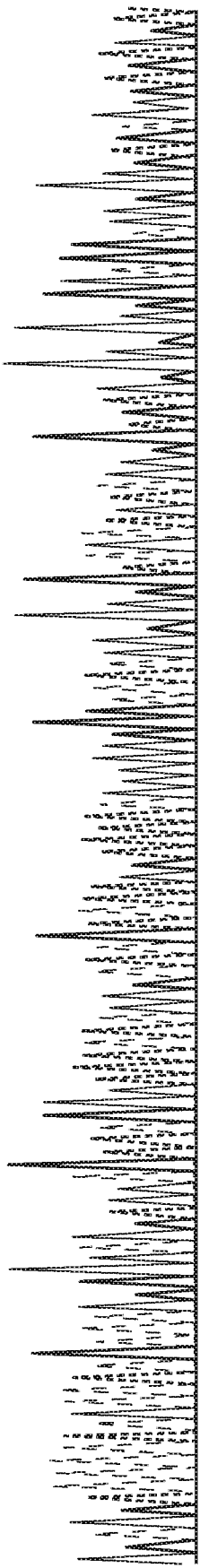


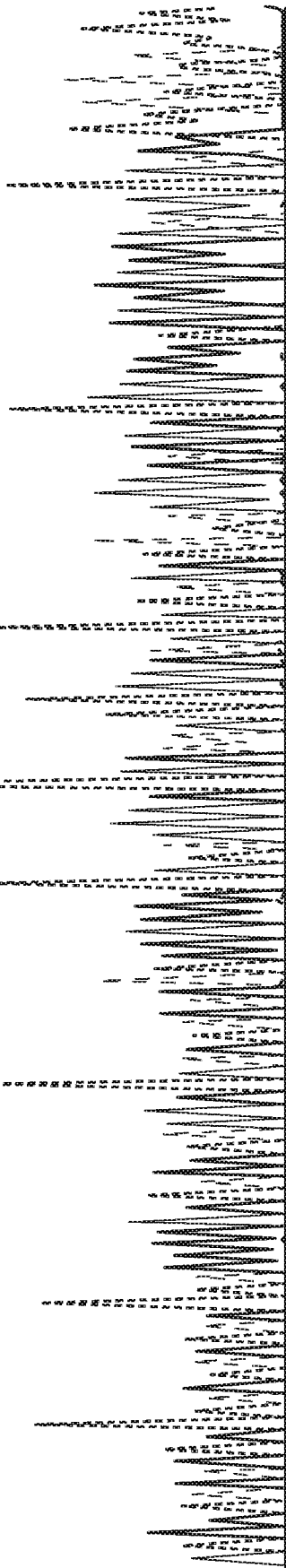
FIG. 19C

TAGTAGCTTTTCTATTCTCTGTTTAGGAATAGCAATGCCCTGAACCTCTAACTCTGCTCCAGCCCTAAAAGGGTTCTAAGAGCCTATCTACTAGCCGCCAGAAAGGATGGTAAAGAGCGCTAAGCCGACGCC



□ □ □ □

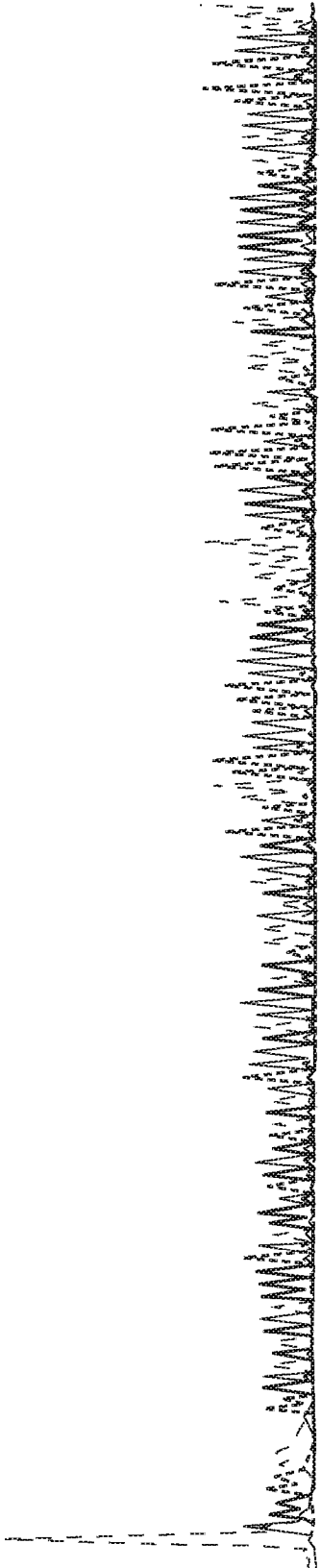
GACGGCTGTGGCCCTGCTGCTGGGAGCTAGCATGCTGTCTGGAGGGCAGTAAGCAGTACCAAGTACTAGCTTAAAGTACGCAAGGGCGGAGGATACATGGCCCTCTCTCCCA



□ □

FIG. 19D

NTTTGGCCIT TNCGGATGTGATGGGATGGACCGCTGTACAAGTATAGTGCCTTATGTAGGCACTCCAAAGCCCAAGGGCTCTTTTCAGAGCCCACTACTTTGTCACAAAGGAGGCTATAGCCACAATTT



- 
- 
- 
- 

360 370 380 390 400 410 420 430 440 450 460 470 480  
TGTAGTCCAGCTACTCGGGAGCTGAGCGGAGAAATGGCGTCAACCGTGGAGGGGGAGGCTTGCAGTGAAGCCGAGATCGGGCCACTGCACCTCCAGCTGGGTGACAGAGCCGAGAGACTCCN TAAA NNTA N

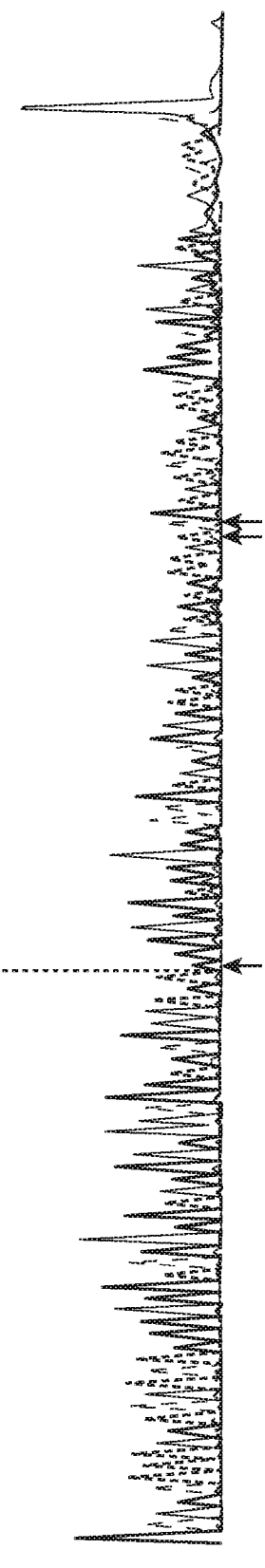


FIG. 19E

4-5/1-683 ..... STGCTTTATCCAAATGAAAAAGATAAATAATCCATATTTCCATAAACCACCACCCCTCAGTGAACCCTGTTTTCTTTGTCCAAATCAGAGCTGAGGAA 100  
 8-5/1-708 ..... CAATAGAGGCTATCCGTAATTTGGTCAATTAATTTGGCTGCTTCCAAATGAAAAAGATAAATAATCCATATTTCCATAAACCACCACCCCTCAGTGAACCCTGTTTTCTTTGTCCAAATCAGAGCTGAGGAA 135  
 NT-5/1-717 ..... CTGCAATAGGAAGCTATCCGTAATTTGGTCAATTAATTTGGCTGCTTCCAAATGAAAAAGATAAATAATCCATATTTCCATAAACCACCACCCCTCAGTGAACCCTGTTTTCTTTGTCCAAATCAGAGCTGAGGAA 138  
 ref\_seq/1-742 lcaaacTgcaaggctgcaaataggaagctatccttatTggtcaattatggtttccgtgctttatccaatagaaaaagataacaataaattccatatttgcataaaacccccctcagtgaaaaccgTggtttcttttgcctcaatcagaagtTgagaa 150  
 4-5/1-683 101 TCTTAACCCGTCATTTGAAATCTCAGGACTATAAATACATGGCCCTGCAACTGTTCTCTACTACTCTGTTAGTGGAGACTGTTAGAGCTTTCTTTAATTCTGTTTAGGAAATGCAATGGCTGAACCCCTTAAGTCTGCTCCAGCCCTATAA 250  
 8-5/1-708 136 TCTTAACCCGTCATTTGAAATCTCAGGACTATAAATACATGGCCCTGCAACTGTTCTCTACTACTCTGTTAGTGGAGACTGTTAGAGCTTTCTTTAATTCTGTTTAGGAAATGCAATGGCTGAACCCCTTAAGTCTGCTCCAGCCCTATAA 285  
 NT-5/1-717 139 TCTTAACCCGTCATTTGAAATCTCAGGACTATAAATACATGGCCCTGCAACTGTTCTCTACTACTCTGTTAGTGGAGACTGTTAGAGCTTTCTTTAATTCTGTTTAGGAAATGCAATGGCTGAACCCCTTAAGTCTGCTCCAGCCCTATAA 288  
 ref\_seq/1-742 151 tctttaaaccgttcatttgaatctcaggactataaatacatgggctctgaaactgttctgtaactatctctgtagtggagagtgttagtagtcttctattctgttttggaaatgcaatgctgaaccctttaaagtctgctccagccccctaaa 300  
 4-5/1-683 251 AAGGGTTCTAAGAGGCTATACCTAAGCCCGAGAGAGCGAATGGTAAAGACGTTAAGCCACGCCAAGGAGACTAATCTATCTATGTTCTACAAGGTTCTGAAGXCCGGTCCACCCCGACACCCGGGCTCTCATCCAGGCCATGGGGATC 400  
 8-5/1-708 286 AAGGGTTCTAAGAGGCTATACCTAAGCCCGAGAGAGCGAATGGTAAAGACGTTAAGCCACGCCAAGGAGACTAATCTATCTATGTTCTACAAGGTTCTGAAGXCCGGTCCACCCCGACACCCGGGCTCTCATCCAGGCCATGGGGATC 435  
 NT-5/1-717 289 AAGGGTTCTAAGAGGCTATACCTAAGCCCGAGAGAGCGAATGGTAAAGACGTTAAGCCACGCCAAGGAGACTAATCTATCTATGTTCTACAAGGTTCTGAAGXCCGGTCCACCCCGACACCCGGGCTCTCATCCAGGCCATGGGGATC 438  
 ref\_seq/1-742 301 aagggttctaaagaagctatcactaaagccgcagaagaaggatggttaagaagcgtaaagccagcccaaggagactaatctatctatggtacaaggttctgaagxcnggtccacccccgacccccggcatctcaatccaaaggccatggggatc 450  
 4-5/1-683 401 ATGAATTCCTTGGTCAAGCAGATCTCCAGGGCATGGGGCCGAGGCTTCTGGCTGGCTCACTACAATAAAGGCTCGACCATCAGCTCCAGGGAGATTTCAGAGGCTGTGGGCTGTGCTGCTGGGAGCTGGCTAAGCATGCTGTG 550  
 8-5/1-708 436 ATGAATTCCTTGGTCAAGCAGATCTCCAGGGCATGGGGCCGAGGCTTCTGGCTGGCTCACTACAATAAAGGCTCGACCATCAGCTCCAGGGAGATTTCAGAGGCTGTGGGCTGTGCTGCTGGGAGCTGGCTAAGCATGCTGTG 585  
 NT-5/1-717 439 ATGAATTCCTTGGTCAAGCAGATCTCCAGGGCATGGGGCCGAGGCTTCTGGCTGGCTCACTACAATAAAGGCTCGACCATCAGCTCCAGGGAGATTTCAGAGGCTGTGGGCTGTGCTGCTGGGAGCTGGCTAAGCATGCTGTG 588  
 ref\_seq/1-742 451 atgaattccttctgtcaagacatcttccagagcactcggggggcaggcttctcgccctggctcactacaataaagcctcgaaccatcaccctccagggagattcagacggctgctggcctggcctgggagcctggtcctaaagcctgctggtg 600  
 4-5/1-683 551 TCGGAGGGCATAAAGCAGTTACCAAGTACACTAGCTTAAAGTGTGCGAAGGGCGANAGGATACATGGCCCTCTCCGAGCGCATCAGTTACACATTTTGGTCCATCAACCGGTTGACTTTGACA ..... 683  
 8-5/1-708 586 TCGGAGGGCATAAAGCAGTTACCAAGTACACTAGCTTAAAGTGTGCGAAGGGCGATACACTGGCCCTCTCCGAGCGCATCAGTTACACATTTTGGTCCATCAACCGGTTG ..... 708  
 NT-5/1-717 589 TCGGAGGGCATAAAGCAGTTACCAAGTACACTAGCTTAAAGTGTGCGAAGGGCGANAGGATACATGGCCCTCTCCGAGCGCATCAGTTACACATTTTGGTCCATCAACCGGTTGACTTT ..... 717  
 ref\_seq/1-742 601 tccgagggcactaaagcagtlacccaagtlaccactagctctaaaagtggccaagggcgagaggatcaatgggctctctcccaggactcagagttacaactcttggctccatcaacgggtgTggaactttgacatTgTgggTc 742

FIG. 19F

```

4_5/1-394 .....AGTCTTATGTAGCAGCTCCAAACCCAAAGGCTCTTTTCAGAGCCACCTACTTTTCACAGGAGAGCTATAACCAAAATTTCTTAAGTGTGCTGCTG 100
8_5/1-481 .....CCGATGTGATGGGCATGGACGAGCTGTACAGTAAAGTGGCTCCAAACCCAAAGGCTCTTTTCAGAGCCACCTACTTTTCACAGGAGAGCTATAACCAAAATTTCTTAAGTGTGCTGCTG 135
NT_3/1-417 .....CCGATGTGATGGGCATGGACGAGCTGTACAGTAAAGTGGCTCCAAACCCAAAGGCTCTTTTCAGAGCCACCTACTTTTCACAGGAGAGCTATAACCAAAATTTCTTAAGTGTGCTGCTG 135
ref_seq/1-500 1 aggcctttaccgatgtgatgggcatggaagagctgtacaagttaagttagcaacttccaaaccccaaaaggctcttttcagagaccacctacttttcacaagggagagctataaccacaatttcttaagggtggtgctgctg 150

4_5/1-394 101 CTATTCCTTTCAGTTCTAGAGGATCAACTGGATGTTAGCGACACAGTTTTAGAGTCAAGGTTAACTTGGACGCGGGCTGGCTCTTTCGCTTTTAAATCCCGGCAATTTGGGAGCGCGAGGCGGCGATCACGAG 244
8_5/1-481 136 CTATTCCTTTCAGTTCTAGAGGATCAACTGGATGTTAGCGACACAGTTTTAGAGTCAAGGTTAACTTGGACGCGGGCTGGCTCTTTCGCTTTTAAATCCCGGCAATTTGGGAGCGCGAGGCGGCGATCACGAG 279
NT_3/1-417 136 CTATTCCTTTCAGTTCTAGAGGATCAACTGGATGTTAGCGACACAGTTTTAGAGTCAAGGTTAACTTGGACGCGGGCTGGCTCTTTCGCTTTTAAATCCCGGCAATTTGGGAGCGCGAGGCGGCGATCACGAG 279
ref_seq/1-500 145 ctattctcttccagttctcagaggatcaactgggaatgttagcgaagacaagtcttagagagccaagtcttaacttggacgggcccgtgcccgtgacctcttgccttlaatccccggcaatttggggagcccgggaggatcacgag 288

4_5/1-394 245 GTCAGGAGATGAGACCATCCCTGCTTACACCGATGAACCCCTCTCTACTAAATAATACAAATAATATTAGCTGGGCTGATGGTGGGCGCTGTANTCCAGCTACTCGGGAGGCTGAGGCGAGGAGATGGGCTGAACGCATGA 388
8_5/1-481 280 GTCAGGAGATGAGACCATCCCTGCTTACACCGATGAACCCCTCTCTACTAAATAATACAAATAATATTAGCTGGGCTGATGGTGGGCGCTGTACTCCAGCTACTCGGGAGGCTGAGGCGAGGAGATGGGCTGAACGCATGA 423
NT_3/1-417 280 GTCAGGAGATGAGACCATCCCTGCTTACACCGATGAACCCCTCTCTACTAAATAATACAAATAATATTAGCTGGGCTGATGGTGGGCGCTGTACTCCAGCTACTCGGGAGGCTGAGGCGAGGAGATGGGCTGAAC ..... 417
ref_seq/1-500 289 gtcaggagatggagaccatccctgcttlaaacacgatgaaccccctctctactaaaaatacaaaaataattagctgggcccgtgatgggtgggacctagctcccagctactcgggaggtgagggcagagagaatggcgtgaaccgcagga 432

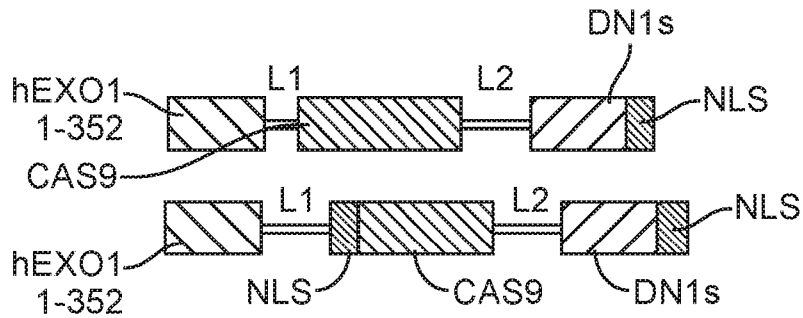
4_5/1-394 389 GGGCGEA .....
8_5/1-481 424 GCGGGAGCTTGCAGTGGCGGAGATCGGGCACTGCACTCCAGGCTGGGTGACAGAGC .....
NT_3/1-417 .....
ref_seq/1-500 433 ggcggagccttcagtgagcccagatgcgccactgcactccagcctgggtcacagagagagactccgt

```

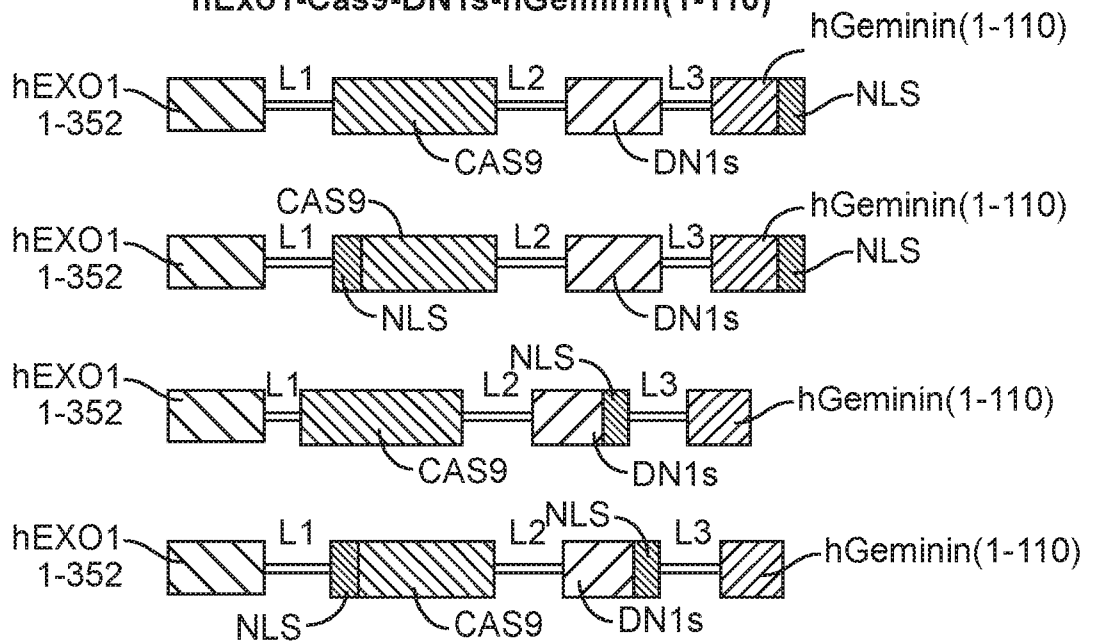
FIG. 19G



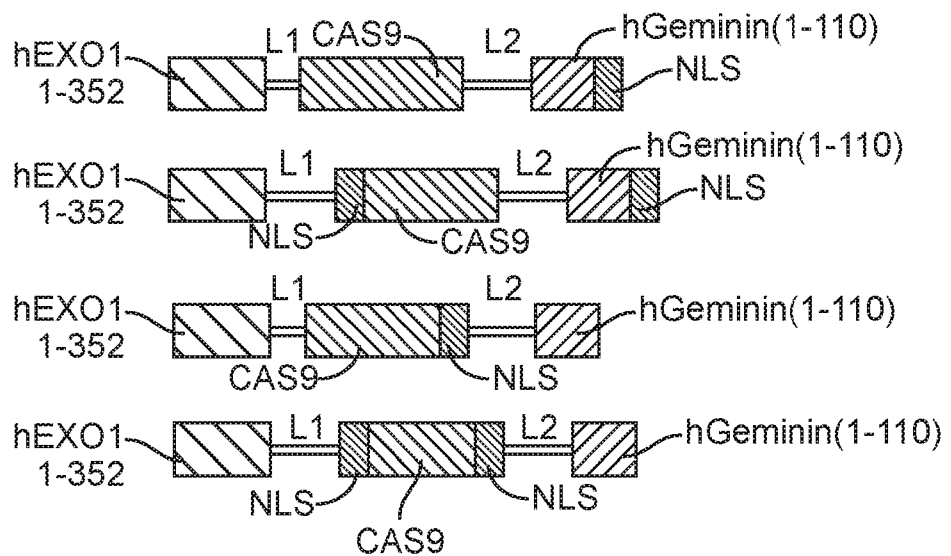
**hExo1-Cas9-DN1s**



**hExo1-Cas9-DN1s-hGeminin(1-110)**



**hExo1-Cas9-hGeminin(1-110)**



**FIG. 20**

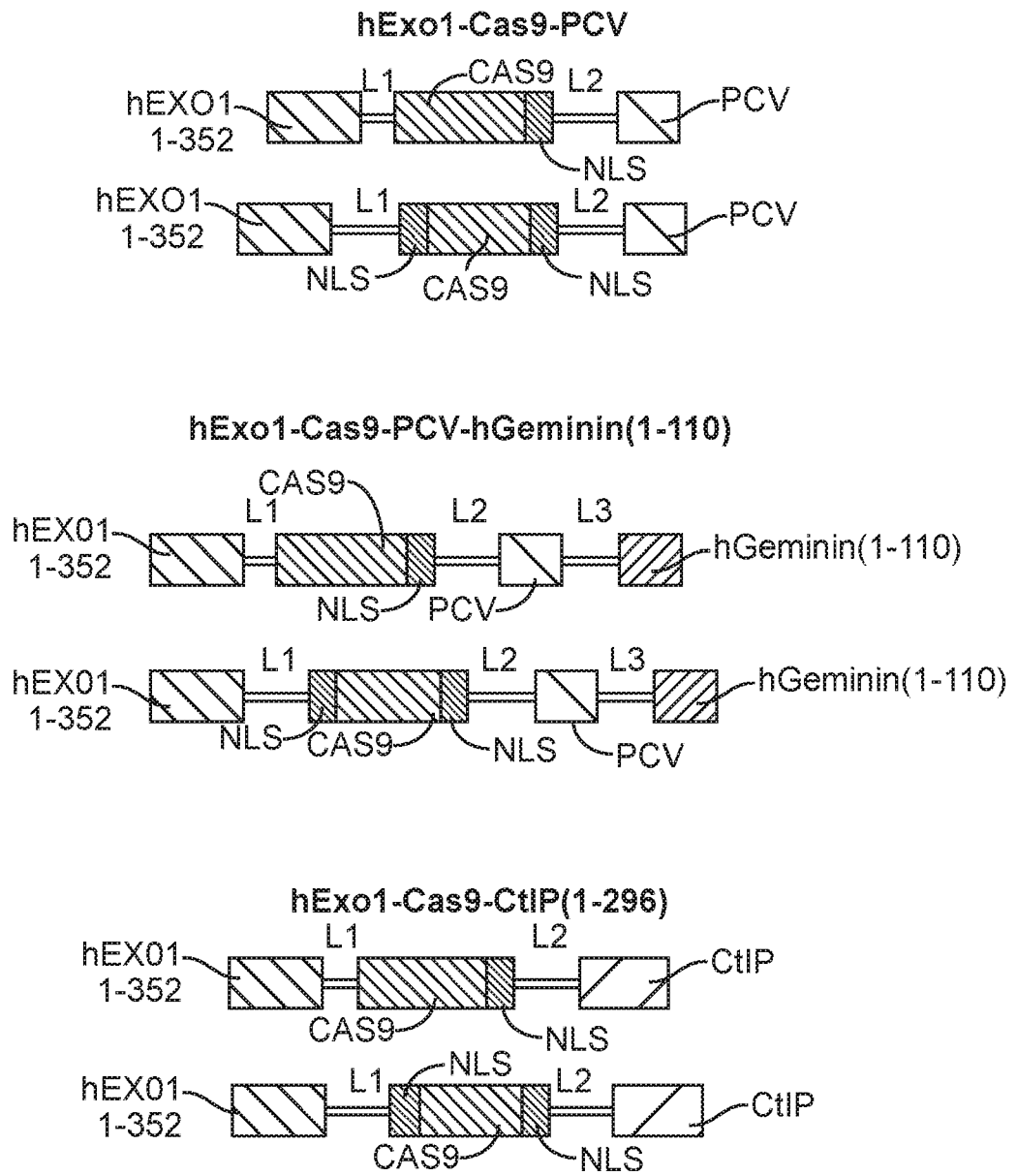


FIG. 20 (Cont.)

# **A NON-TOXIC CAS9 ENZYME AND APPLICATION THEREOF**

## **CROSS-REFERENCE**

**[0001]** This application claims priority to U.S. provisional application 62/789,347, filed on January 7, 2019; U.S. provisional application 62/823,477, filed on March 25, 2019; U.S. provisional application 62/824,164, filed on March 26, 2019, and U.S. provisional application 62/855,612, filed on May 31, 2019, the entirety of which are hereby incorporated by reference herein.

## **SEQUENCE LISTING**

**[0002]** The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on January 3, 2020, is named 55190-701\_601\_SL.txt and is 236,317 bytes in size.

## **BACKGROUND**

**[0003]** Targeted editing of nucleic acids is a highly promising approach for studying genetic functions and for treating and ameliorating symptoms of genetic disorders and diseases. Most notable target-specific genetic modification methods involve engineering and using of zinc finger nucleases (ZFNs), transcription activator like effector nucleases (TALENs), and RNA-guided DNA endonuclease Cas. Frequency of introducing mutations such as deletions and insertions at the targeted nucleic acids through the non-homologous end joining (NHEJ) repair mechanism limits the applications of genetic targeting and editing in the development of therapeutics.

## **SUMMARY**

**[0004]** The disclosure is summarized here in part in the claims disclosed herein. Disclosed herein is a method comprising introducing a first vector into a plurality of cells wherein said first vector encodes a fusion protein complex comprising a Cas9 nuclease fused to an exonuclease; wherein a viability of said plurality of cells comprising said vector is at least 1.5 times that of a second plurality of cells comprising a second vector encoding a Cas9 nuclease; wherein said second plurality of cells are K562 cells transfected with said second vector. The first vector can encode the Cas9 fused to an exonuclease and a gRNA. The exonuclease can be selected from the group consisting of MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean. A donor polynucleotide can be introduced into the first plurality of cells. The method can comprise making an edit to an abnormal locus of a gene by said Cas9-fused to an exonuclease. The donor polynucleotide can

comprise an integration cassette further comprising a functional locus of said gene. The viability can be measured by resazurin assay. The exonuclease can be ExoI. The abnormal locus can be an abnormal locus of a HBB gene. The donor polynucleotide can encode a functional locus of said HBB gene. The fusion protein complex can encode at least one nuclear localization signal (NLS). The first vector encoding the fusion protein complex can have at least 80% sequence identity with any one of SEQ ID NO: 2-18. The first vector can be delivered by electroporation. The donor polynucleotide can comprise a mutated protospacer adjacent motif (PAM) sequence located at the immediate 3' end of a cleavage site, wherein said mutated PAM sequence comprises 5'-NCG-3' or 5'-NGC-3'. The fusion protein complex can be unable to cleave said mutated PAM sequence. The donor polynucleotide can be single-stranded DNA. The donor polynucleotide can be double-stranded DNA.

**[0005]** Disclosed herein is a polypeptide, comprising a first functional fragment, a second functional fragment comprising a Cas nuclease, and a linker peptide, wherein said first functional fragment is coupled to a first end of the linker peptide and the second functional fragment is coupled to a second end of said linker peptide; and when a first complex comprising said polypeptide and a ribonucleic acid (RNA) molecule is administered to a first plurality of cells, a reduced toxicity is observed in said first plurality of cells compared to said toxicity observed in a second plurality of cells when a second complex comprising a Cas9 nuclease and said RNA molecule is administered to said second plurality of cells. The first functional fragment can comprise an exonuclease wherein the exonuclease is selected from the group consisting of MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean. The RNA molecule can be a guide RNA. The exonuclease can be a human Exo1 enzyme. The N-terminal of the human Exo1 enzyme can be coupled to said C-terminal of said linker which is coupled to said C-terminal of said Cas nuclease. The human Exo1 enzyme can comprise SEQ ID NO: 1. The human Exo1 enzyme can comprise a fragment that has a 80% sequence identity of SEQ ID NO:1. The human Exo1 enzyme can comprise a fragment that has a 90% sequence identity of SEQ ID NO:1. The human Exo1 enzyme can comprise a fragment that has a 95% sequence identity of SEQ ID NO:1. The second functional fragment can comprise a Cas9 enzyme. The Cas9 enzyme can comprise a N-terminal nuclear localizing sequence (NLS) and a C-terminal NLS. The Cas9 enzyme can comprise a N-terminal nuclear localizing sequence (NLS). The Cas9 enzyme can comprise a C-terminal nuclear localizing sequence (NLS). The linker peptide can be selected from a group consisting of FL2X, SLA2X, AP5X, FL1X, SLA1X. The linker peptide can be SLA2X. The peptide can comprise 5 to 200 amino acids. The reduced toxicity can be quantified by measuring resorufin

accumulation. After administration of said first complex, the first plurality of cells can have at least two times a number of viable cells compared to said second plurality of cells after administration of said second complex wherein the number of viable cells is quantified by a resorufin assay. After administration of the first complex, the first plurality of cells has at least two times said amount of HDR edited cells when compared to the second plurality of cells after administration of the second complex as quantified by a cellular HDR assay. The cellular HDR assay can comprise IHC, qPCR or deep sequencing.

**[0006]** Disclosed herein is a polynucleotide encoding the aforementioned polypeptide and the RNA molecule. The first end of the linker peptide can be a 3' end and the second end of the linker peptide can be a 5' end. The first end of said linker peptide can be a 5' end and the second end of said linker peptide can be a 3' end. The RNA molecule can be a guide RNA (gRNA). The polynucleotide can comprise a homology directed repair (HDR) template. The gRNA can be selected from sequences listed in Table 2. The HDR template can be single-strand DNA. The HDR template can be double-strand DNA. The polynucleotide can be formulated in a liposome. The liposome can comprise a polyethylene glycol (PEG), a cell-penetrating peptide, a ligand, an aptamer, an antibody, or a combination thereof.

**[0007]** Disclosed herein is a vector comprising a nucleotide sequence of the aforementioned polypeptide. The vector can comprise a promoter. The promoter can be a CMV or a CAG promoter. The vector can be selected from a group consisting of retroviral vectors, adenoviral vectors, lentiviral vectors, herpesvirus vectors, and adeno-associated viral vectors. The vector can be an adeno-associated viral vector. Disclosed herein is a virus-like particle (VLP) comprising the aforementioned vector. Disclosed herein is a kit comprising the aforementioned polypeptide formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.

**[0008]** Disclosed herein is a kit comprising the aforementioned polynucleotide formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.

**[0009]** Disclosed herein is a kit comprising the aforementioned vector formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.

**[0010]** Disclosed herein is a method for inducing homologous recombination of DNA in a cell, comprising contacting the DNA with the aforementioned polypeptide.

**[0011]** Disclosed herein is a method for inducing HDR in a cell in vitro or ex vivo, comprising delivering the aforementioned polynucleotide into a cell. The cell can be a human cell, a non-human mammalian cell, a stem cell, a non-mammalian cell, an invertebrate cell, a plant cell, or a single-eukaryotic organism.

**[0012]** Disclosed herein is a method, comprising: contacting a first of plurality of cells with an aforementioned polynucleotide and a second plurality of cells with a second polynucleotide encoding a wild-type Cas9 enzyme; and inducing a site-specific cleavage at an intended locus followed by HDR in the first plurality of cells and the second plurality of cells; and recovering at least 30-90% more cells in the first plurality of cells compared to the second plurality of cells. The method can further comprise measuring cell viability by measuring an amount of resorufin produced in the first plurality of cells and the second plurality of cells. The first plurality of cells can have 2-5 times an amount of viable cells as quantified by a resorufin assay when compared to the second plurality of cells. The first plurality of cells and the second plurality of cells can comprise a human cell, a non-human mammalian cell, a stem cell, a non-mammalian cell, a invertebrate cell, a plant cell, or a single-eukaryotic organism. The human cell can be a T cell, a B cell, a dendritic cell, a natural killer cell, a macrophage, a neutrophil, an eosinophil, a basophil, a mast cell, a hematopoietic progenitor cell, a hematopoietic stem cell (HSC), a red blood cell, a blood stem cell, an endoderm stem cell, an endoderm progenitor cell, an endoderm precursor cell, a differentiated endoderm cell, a mesenchymal stem cell (MSC), a mesenchymal progenitor cell, a mesenchymal precursor cell, or a differentiated mesenchymal cell. The differentiated endoderm cell can be a hepatocytes progenitor cell, a pancreatic progenitor cell, a lung progenitor cell, or a tracheae progenitor cell. The differentiated mesenchymal cell can be a bone cell, a cartilage cell, a muscle cell, an adipose cell, a stromal cell, a fibroblast, or a dermal cell.

**[0013]** Disclosed herein is a method for treating a single gene disorder in a subject, comprising: culturing a plurality of primary cells obtained from said subject; administering the aforementioned polynucleotide to a plurality of primary cells, wherein the gRNA is configured to recognize a locus of the gene that causes said disorder and the HDR template is configured to provide a functioning sequence of the gene; and inducing a site-specific cleavage at the locus followed by HDR, wherein the functioning sequence of said gene is inserted at the locus. The method can further comprise selecting primary cells in which said functioning sequence of the gene is inserted at the locus; and reintroducing the selected primary cells back into the subject. The subject can be a mammal. The mammal can be a human. The plurality of primary cells can be selected from a group comprising T cells, B cells, dendritic cells, natural killer cells, natural killer cells, macrophages, neutrophils, eosinophils, basophils, mast cells, hematopoietic progenitor cells, hematopoietic stem cells (HSCs), red blood cells, blood stem cells, endoderm stem cells, endoderm progenitor cells, endoderm precursor cells, differentiated endoderm cells, mesenchymal stem cells (MSCs), mesenchymal progenitor cells, mesenchymal precursor cells, differentiated mesenchymal cells, hepatocytes progenitor cells, pancreatic progenitor cells, lung

progenitor cells, tracheae progenitor cells, bone cells, cartilage cells, muscle cells, adipose cells, stromal cells, fibroblasts, and dermal cells. The gene that causes said single gene disorder can be selected from Table 3.

**[0014]** Disclosed herein is a method for treating sickle cell anemia caused by an abnormal HBB gene in a subject, comprising: culturing a plurality of primary cells obtained from said subject; administering the aforementioned polynucleotide to the plurality of primary cells, wherein the gRNA is configured to recognize a locus of said HBB gene that causes the disorder and the HDR template is configured to provide a functioning sequence of said HBB gene; and inducing a site-specific cleavage at said locus followed by HDR, wherein the functioning sequence of said HBB gene is inserted at the locus. The method can further comprise selecting primary cells in which said functioning sequence of said HBB gene is inserted at said locus; and reintroducing said selected primary cells back into said subject. The subject can be a mammal. The mammal can be a human. The primary cell can be a hematopoietic stem cell. The primary cell can be a CD34+ hematopoietic stem cell. The primary cell can be a CD34+ hematopoietic stem cell. The vector can comprise plasmid PX330. The cell can be a CD34+ hematopoietic stem cell.

**[0015]** Disclosed herein is a method for treating sickle cell anemia caused by an abnormal HBB gene in a subject, comprising: culturing a plurality of primary cells obtained from the subject; administering the aforementioned polynucleotide to the plurality of primary cells, wherein the gRNA is configured to recognize a locus of the HBB gene that causes the disorder and the HDR template is configured to provide a functioning sequence of the HBB gene; and inducing a site-specific cleavage at the locus followed by HDR, wherein the functioning sequence of the HBB gene is inserted at the locus. The method can further comprise selecting primary cells in which the functioning sequence of the HBB gene is inserted at the locus; and reintroducing the selected primary cells back into the subject. The subject can be a mammal. The mammal can be a human. The primary cell can be a CD34+ hematopoietic stem cell.

**[0016]** Disclosed herein is a method, comprising: contacting a first of plurality of cells with a first complex comprising the aforementioned polynucleotide and a RNA molecule; inducing a site-specific cleavage followed by HDR in the first plurality of cells, wherein a percentage of cells of the first plurality of cells edited by HDR quantified by a cellular HDR assay is at least two times higher compared to a percentage of cells of a second plurality of cells contacted with a second complex comprising a polynucleotide encoding a wild-type Cas9 enzyme and the RNA molecule. The cellular HDR assay can comprise IHC. The cellular HDR assay can comprise qPCR. The cellular HDR assay can comprise nucleic acid sequencing.

## INCORPORATION BY REFERENCE

[0017] 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

[0018] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0019] Some 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:

[0020] **FIG. 1** shows embodiments of fusion proteins comprising hExo1 enzyme and Cas9 enzyme linked together through different linkers.

[0021] **FIG. 2** shows an embodiment of an intended target site and a HDR template.

[0022] **FIG. 3** shows an embodiment of conducting a resazurin reduction assay. Column 1-8 correspond to Cas9-HR fusion proteins 1-8 described in **FIG.1** respectively.

[0023] **FIG. 4** shows a normalized fold change of resorufin fluorescence of cells transfected with RNP plasmids, GFP plasmids, and control plasmids before puromycin selection.

[0024] **FIG. 5** shows a normalized fold change of resorufin fluorescence of cells transfected with wild type Cas9 enzyme plasmids treated with either dimethyl sulfoxide (DMSO) or pifithrin- $\alpha$  (PFT- $\alpha$ ).

[0025] **FIG. 6A** shows an embodiment of an intended target site with three gRNA sequences (G1, G2, and G3; SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23 respectively) designed to target Exon 1 of the HBB gene.

[0026] **FIG. 6B** shows a normalized fold change of resorufin fluorescence of cells transfected with RNP plasmids with three gRNA sequences (G1, G2, and G3; SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23 respectively) designed to target Exon 1 of the HBB gene.

[0027] **FIG. 6C** shows a Cas9 HBB-G3 reverse Sanger sequence trace (SEQ ID NO: 161).

[0028] **FIG. 7** shows an embodiment of conducting a resazurin reduction assay. Column 1-9 correspond to Cas9-HR fusion proteins 1-9 of fusion proteins described in **FIG.1** respectively.



[0029] **FIG. 8** shows a normalized fold change of resorufin fluorescence of cells transfected with RNP plasmids with different gRNA sequences, GFP plasmids, and two different control plasmids to control cells.

[0030] **FIGs. 9A-B** show a normalized fold change of resorufin fluorescence of cells transfected with RNP plasmids. **FIG. 9A** shows the G2 (SEQ ID NO: 22) and G3 (SEQ ID NO: 23) gRNA targeting the exon 1 of the HBB gene. **FIG 9B** shows that RNP plasmids with the seventh fusion protein (**FIG. 1**) and G3 gRNA have less cellular toxicity compared to RNP plasmids with the unmodified Cas9 and G2 gRNA.

[0031] **FIG. 10** shows a normalized fold change of resorufin fluorescence of cells transfected with different RNP plasmids targeting exon 1 of HBB gene.

[0032] **FIG. 11A** is a diagram of Plasmid PX330 which contains a constitutive promoter for mammalian Cas9 expression, along with U6 promoter driven gRNA expression.

[0033] **FIG. 11B** is an example of the experimental set up wherein cells are seeded and after two days of growth cellular toxicity is quantified.

[0034] **FIG. 11C** is a graph showing reduced cellular toxicity in A549 cells as shown in the **FIG. 11B** experimental set up and a diagram of the gRNA targeting intergenic region on Chromosome 12 depicted above the graph.

[0035] **FIG. 11D** is a graph showing that treatment with alpha-pifithrin (10 micromolar) reduces Cas9 induced cellular toxicity in A549 cells.

[0036] **FIG. 12A** is a diagram of the Puromycin resistance repair template (RT).

[0037] **FIG. 12B** shows the method used to quantify HDR and INDEL rates of hExo-Cas9 fusions in A549 cells.

[0038] **FIG. 12C** is a graph depicting the toxicity of various constructs tested via a resazurin assay.

[0039] **FIG. 12D** depicts the method of the resazurin assay.

[0040] **FIG. 12E** is a depiction of the genomic region of cells successfully integrated by the Puro-RT.

[0041] **FIG. 12F** is a graph of the survival of K562 cells transfected with either Cas9-HR8 (8) or Cas9 (NT) with G2 or G3 RNA after three days of puromycin treatment.

[0042] **FIG. 12G** is an agarose gel of the amplification products of the primers depicted in Fig. 12E showing stable integration of the repair template using Cas9-HR8 (fusion protein 8 of **FIG. 1**) and Cas9 (NT) with gRNA G2 or G3 in the genome.

[0043] **FIG. 13A** shows the genomic region, including the first two exons of HBB targeted to edit the Human Hemoglobin Beta (HBB) gene and a graph depicting data from the toxicity screen of HBB gRNA guides in A549 cells.

[0044] **FIG. 13B** shows sanger sequencing of the HBB genomic region in the HBB-G3 treated A549 cells (SEQ ID NO: 161).

[0045] **FIG. 13C** is a diagram of the wild-type HBB sequence (SEQ ID NO: 162) and the SSRT-G3 sequence (SEQ ID NO: 163) which introduces the sickle cell (E6V) an missense mutation which results in an EcoRI site and four silent mismatch mutations (bolded nucleotides **a, a, a, and g** on single-stranded repair template, SSRT G3) with the HBB-G3 gRNA highlighted by the bar from above. Mutations are designed to prevent gRNA binding upon successful repair

[0046] **FIG. 13D** depicts a HBB editing experiment in which K562 cells or A549 cells are electroporated with Cas9+SSRT-G3, Cas9-HR 1-9+SSRT-G3 or SSRT-G3 alone.

[0047] **FIG. 14** illustrates toxicity assessment of two transfection methods, lipofectamine and calcium phosphate (CalPhos) as determined by transfecting A549 cells with HBB-G3 gRNA and Cas9-HR fusion proteins 4 and 5 as depicted in **FIG. 1**.

[0048] **FIG. 15** illustrates toxicity assessment by transfecting A549 cells with HBB repair templates of **FIG. 13A**. Resazurin levels are measured on day 2 after the transfection.

[0049] **FIG. 16A** shows an agarose gel of EcoRI digestion assay of Cas9-HR fusion protein 8 of **FIG. 1** integrating the HBB repair template into the genome of K562 cells. Arrows indicate the EcoRI digested products. There are no EcoRI digested products in lanes of Cas9 only (NT), SSRT, and Con (no Cas9).

[0050] **FIG. 16B** shows an agarose gel of EcoRI digestion assay of Cas9-HR fusion proteins 4, 5, 6, 7, and 8 of **FIG. 1** integrating the HBB repair template into the genome of K562 cells. Arrows indicate the EcoRI digested products. There are no EcoRI digested products in NT and Con lanes.

[0051] **FIG. 16C** shows a western blotting of Cas9-HR fusion proteins 4, 5, 6, 7, and 8 of **FIG. 1**, Cas9 only (NT), and Con (no Cas9). Arrow indicates detection of Cas9 in Cas9-HR fusion proteins and NT lanes.

[0052] **FIG. 16D** shows successful expression and purification in E. Coli of Cas9-HR 3. Successful expression and purification of Cas9 (lanes 8-14) is also shown to aid comparison.

[0053] **FIG. 16E** shows an immunohistochemistry (IHC) of the same transfected cells from **FIG. 16C**. Arrows indicates that Cas9-HR fusions and Cas9 are localized to the nucleus of the cells.

[0054] **FIG. 17A** illustrates the construct for a full H2B knock-in experiment.

[0055] **FIG. 17B** illustrates p53-dependent decrease of cellular toxicity induced by Cas-HR fusion proteins 4, 5, 6, and 8 of **FIG. 1**, Cas9 only (NT), and Con (no Cas9) in epithelial lung cancer cell lines. A549 cells are positive for p53 activity, while H1299 cells are negative for p53 activity. Toxicity as determined by normalized resazurin levels (y-axis) has shown that absence of p53 in H1299 cells yields lower cellular toxicity.

[0056] **FIG. 17C** illustrates the assessment of successful GFP tagging of H2B as diagrammed in **FIG. 17A** in K562 cells. Arrows indicate successful tagging of H2B with GFP as shown by detection of GFP in the nucleus.

[0057] **FIG. 18A** illustrates the schematic difference between Cas9 only model and Cas9-HR model. The presence of an Exonuclease domain fundamentally changes the predicted in-vitro cleavage pattern. Exo1 has a significant preference for phosphorylated 5' termini vs non-phosphorylated. Therefore, it can be expected when using PCR products or other pieces of DNA lacking 5'-phosphorylated termini that endonuclease cleavage via Cas9 can dominate initially, whereas after cleavage the two fragments each can possess 5'-phosphorylated termini, which result in rapid degradation via the hExo1.

[0058] **FIG. 18B** illustrates an exemplary digestion pattern based on **FIG. 18A**. Only Cas9-HR3 + gRNA and Cas9-HR3 can produce the digested products which demonstrate successful in-vitro nuclease activity. Additionally, though hExo1 strongly prefers phosphorylated 5'-termini, hExo1 can still bind and resect unphosphorylated 5'-termini, so a small amount degradation without gRNAs when compared to Cas9.

[0059] **FIG. 18C** illustrates an actual agarose example of **FIG. 18A** and **FIG. 18B**. Lanes 1 and 2 show Cas9-HR3 targeting either HBB-G1 or HBB-G3, Lanes 3 and 4 show Cas9 (NT) targeting either HBB-G1 or HBB-G3, Lane 5 is untreated DNA.

[0060] **FIG. 18D** illustrates a similar experiment as **FIG. 18C** and differs only by conducting the experiment after leaving enzymes for 2 weeks at 4°C in order to compare protein stability. Lane 1 is digestion pattern from the combination of Cas9-HR3 and gRNA HBB-G1. Lane 2 is digestion pattern from the combination of Cas9 and gRNA HBB-G1. Lane 3 is digestion pattern from the combination of Cas9-HR3 and HBB-G3. Lane 4 is digestion pattern from the combination of Cas9 and HBB-G3. Lane 5 is digestion pattern from Cas9-HR only. Lane 6 is digestion pattern from Cas9 only. Lane 7 is the control, where there is neither Cas9 nor gRNA.

[0061] **FIGs. 19A-G** illustrates induction of genomic integration of the H2BmNeon fusion via Cas9-HR 4, Cas9-HR 8, Cas9 only (NT) and Control without Cas9 (Con). **FIG. 19A** illustrates design of H2B integration detection primers. Two sets of primers are designed to bind outside of the 5' and 3' ends of the repair template annealing to sequences only present in the genome, not

in the RT, while the others anneal to sequences specific to the repair template, and are not present in the unmodified cells. **FIG. 19B** illustrates an agarose gel showing PCR products amplified by the 5' primers, indicating successful tagging of endogenous H2B with GFP. **FIG. 19C** illustrates an agarose gel showing PCR products amplified by the 3' primers, indicating successful tagging of endogenous H2B with GFP. **FIG. 19D** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product amplified by the 5' primers. Figure discloses SEQ ID NOS 164-165, respectively, in order of appearance. **FIG. 19E** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product amplified by the 3' primers. Figure discloses SEQ ID NOS 166-167, respectively, in order of appearance. **FIG. 19F** illustrates sequencing alignment of the PCR product amplified by the 5' primers and discloses SEQ ID NOS 155, 154, 153, and 160, respectively, in order of appearance. **FIG. 19G** illustrates sequencing alignment of the PCR product amplified by the 3' primers and discloses SEQ ID NOS 158, 157, 156, and 159, respectively, in order of appearance.

**[0062]** **FIG. 20** illustrates designs for additional Cas9-HR fusion proteins with expanded functionalities.

#### **DETAILED DESCRIPTION**

**[0063]** A brief description about the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) system is included. The CRISPR/Cas enzyme system first found in bacteria and archaea is an immune defense against viral infection. During viral infection, segments of viral DNA are integrated into CRISPR locus. These segments of integrated viral DNA are transcribed into guide RNA (gRNA), which is sequentially complementary to the viral genome. gRNA directs the Cas enzymes to the gRNA targeted viral genome, where Cas proteins cleave the viral genome, thus defending against viral infection.

**[0064]** The CRISPR system typically comprises a gRNA that is specific to the target DNA sequence and a non-specific Cas 9 protein. Generally, the gRNA includes two distinct segments- CRISPR RNA (crRNA) and transactivating CRISPR RNA (tracrRNA). The crRNA is complementary to the target DNA sequences, and thus recognize the sequence to be cleaved. And the tracrRNA functions as a scaffold for the crRNA-Cas9 interaction. Guide RNA naturally form a duplex molecule, with the crRNA and tracrRNA fragments annealed together. Cas proteins have been investigated and engineered as a tool for genetic editing by generating site-specific double strand breaks (DSBs). Custom designed gRNA directs the Cas proteins to generate DSB at any nucleic acid loci that are complementary to the sequence of gRNA. Cas proteins have been shown to successfully introduce nucleotide changes, deletions, insertions, and substitutions in eukaryotic cells.

**[0065]** The use CRISPR and Cas9 proteins for editing nucleic acids are limited by the endogenous repair mechanism of the cell. DSBs are preferentially repaired by NHEJ. Unintended insertions and deletions at sites of repair associated with NHEJ render development of genetic-based therapy undesirable. Alternatively, if the generated DSBs are resected so that long (<200bp) 3' overhangs are generated, the endogenous repair pathway is forced to use HR. Targeted error-free insertions and deletions of anywhere from 1-1000s of bp of DNA can be achieved by addition of a polynucleotide (template sequence) comprising homology arms flanking the desired insertion or deletion.

**[0066]** Homology directed repair is error free, and results in the ability to insert or delete specific sequences of DNA in a given genome.

**[0067]** Further, the HDR reduces cellular toxicity, which is caused by DSBs introduced by CRISPR and Cas9 enzyme system. The cellular toxicity is dependent on the p53 tumor suppressor pathway, as inhibition or loss of p53 function greatly reduces cellular toxicity in both Human Pluripotent Stem Cells (hPSCs) and in immortalized Retinal Pigment Epithelium (RPE) cells. Since permanent loss of p53 functionality has some severe effects on cells including genomic instability, altered cellular homeostasis, and increased rates of cancer in-vivo, one solution is transient inhibition of p53 by either small molecule or overexpression of dominant negative inhibitors. However, the transient inhibition of p53 in vivo is challenging and could produce undesirable side effects. Therefore, generating a non-toxic Cas9 enzyme is desirable for in vivo applications.

**[0068]** Disclosed herein are compositions and methods related to the selective targeting and editing endogenous nucleic acid segment in both normal cell and in cell associated with genetic diseases with reduced cellular toxicity. Targeted endogenous nucleic acids are cleaved, digested, and edited through HDR. gRNA directs a protein fusion complex comprising of the Cas protein moiety and a human Exo1 enzyme to a specific endogenous nucleic acid segment, where the protein fusion complex introduces cleavage and digestion, leaving 3' or 5' overhangs on the targeted endogenous nucleic acid segment. The overhangs allow for increased rates of HDR when the cell is further presented with a polynucleotide fragment that shares some degrees of sequence homology as the targeted and digested endogenous nucleic acid segment.

**[0069]** Disclosed herein are compositions wherein the targeted endogenous nucleic acids are located in known disease loci. Targeted known disease loci are cleaved, digested, and edited through HDR. gRNA directs a protein fusion complex comprising the Cas protein moiety and a human Exo1 enzyme to a specific known disease locus where the protein fusion complex introduces cleavage and digestion, leaving 3' or 5' overhangs on the targeted endogenous nucleic

acid segment. The overhangs allow for increased rates of HDR when the cell is further presented with a polynucleotide fragment that shares some degrees of sequence homology as the targeted and digested endogenous nucleic acid segment.

### **Fusion protein composition**

**[0070]** Some aspects of the compositions and methods disclosed herein involve at least one modified polypeptide comprising a programmable endonuclease such as a Cas9 or other CRISPR-related programmable endonucleases coupled to a fragment of an exonuclease such as human Exo1 exonuclease or other exonucleases, such as MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean, to reduce cellular toxicity relative to that of an unmodified programmable endonuclease such as Cas9 enzyme in the CRISPR-Cas9 system.

### Cas9 Protein

**[0071]** The polypeptide (fusion protein) comprises a programmable endonuclease such as Cas9, other CRISPR-related programmable endonucleases, other site-specific endonucleases, or a fragment thereof and an exonuclease such as human Exo1 exonuclease or a fragment thereof covalently connected by a peptidyl linker. As used herein, the “Cas9,” “Cas9 domain,” or “Cas9 fragment” refers to an RNA-guided nuclease comprising a Cas9 protein, or a fragment thereof, e.g., a protein comprising an active DNA cleavage domain of Cas9. A Cas9 nuclease is sometimes referred to as a casn1 nuclease or a CRISPR (clustered regularly interspaced short palindromic repeat)-associated nuclease. Cas9 nuclease sequences and structures are well known to those of ordinary skill in the art. Cas9 orthologs have been described in various species, including, but not limited to, *S. pyogenes* and *S. thermophilus*. Wild type (unmodified) Cas9 can be from any of the sequences listed below in Table 1. The Cas9 protein sequences listed in Table 1 is not meant to be limiting. Additional suitable Cas9 nucleases and protein sequences will be apparent to a person of ordinary skill in the art.

**Table 1.** Peptide sequences of various Cas9.

SEQ ID NO: 2	NCBI Reference Sequence: NC_017053.1 ( <i>Streptococcus pyogenes</i> )	MDKKY <u>SIGLDIGTNSVGWAVITDDYK</u> VPSKKFKVLGNTDR <u>HSIKKNLIGALLFGSGETA</u> EATRLKRTARRRYTRRKNRICY LQEIFSNEMAKVDDSSFFHRLEESFLVEEDKKHERHPIFGNIV DEVAYHEKYPTIYHLRKKLADSTDKADLRLIYLALAHMIK FRGHFLIEGDLNPDNSVDKLFQQLVQIYNQLFEENPINASR VDAKAILSARLSKSRLENLIAQLPGEKRNGLFGNLIALSLG LTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYA DLFLAAKNLSDAILLSDILRVNSEITKAPLSASMIKRYDEHH QDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQ EEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIP HQIHLGELHAILRRQEDFYFLKDNREKIEKILTRIPYYVGP LARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIER
--------------	--	---

		<p>MTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEG  MRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE  CFDSVEISGVEDRFNASLGAYHDLLKIIKDKDFLDNEENEDI  LEDIVLTLTLFEDRGMIEERLKYAHLFDDKVMKQLKRRR  YTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQ  IHDDSLTFKEDIQKAQVSGQG<u>HSLHEQIANLAGSPA</u><u>IKKGIL</u>  <u>QTVKIVDELVKVMGHKPENIVEMARENQTTQKGOKNSRE</u>  <u>RMKRIE</u><u>EGIKELGSQILKEHPVENTQLONEKLYLYYLQNGR</u>  <u>DMYVDQELDINRLSDYDVDHIVPQSF</u><u>IKDDSIDNKVLTRSD</u>  <u>KNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL</u>  <u>TKAERGGLSEL</u><u>DKAGFIKRQLVETROITKHVAQILDSRMNT</u>  <u>KYDENDKLIREVKVITLKS</u><u>KLVSDFR</u><u>KDFOFYK</u><u>VREINNYH</u>  <u>HAHDAYLNAVVG</u><u>TALIKKYPKLESEFVYGDYKVYDVRKM</u>  <u>IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIE</u>  <u>TNGETGEIVWDKGRD</u><u>FATVRKVL</u><u>SMPQVNIVKKTEVQTG</u>  GFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVL  VVAKVEKGKSKKLKSVKELLGITIMERSSEKPNIDFLEAK  GYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNL  LALPSKYVNFLYLASHYEKLGKSPEDNEQKQLFVEQHKHY  LDEIIEQISEFSKRVLADANLKVLSAYNKHHRDKPIREQAE  NIIHLFTLTNLGAPAAFKEYFDTTIDRKRYTSTKEVLDATLIH  QSITGLYETRIDLSQLGGD (single underline: HNH domain;  double underline: RuvC domain)</p>
<p>SEQ ID NO: 3</p>	<p><i>Streptococcus</i> <i>thermophilus</i></p>	<p>MLFNKCIISINLDFSNKEKCMTPYSIGLDIGTNSVGWAVIT  DNYKVPSKMKVLGNTS  KKYIKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRIL  YLQEIFSTEMATLDDAFFQ  RLDDSFVLPDDKRDSKYPIFGNLVEEKVYHDEFPTIYHLRK  YLADSTKKADLRLVYLALA  HMIKYRGHFLIEGEFNSKNNDIQKNFQDFLDTYNAIFESDLS  LENSKQLEEIVKDKISKL  EKKDRILKLFPGKNSGIFSEFLKLIVGNQADFRKCFNLDEK  ASLHFSKESYDEDLETLL  GYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSA  MIKRYNEHKEDLALLKEYI  RNISLKYNEVFKDDTKNGYAGYIDGKTNQEDFYVYLKNL  LAEFEGADYFLEKIDREDFL  RKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLLAKNKERI  EKILTFRIPYYVGPLARGN  SDFAWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLY  LPEEKVLPKHSLLYETFN  VYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFKDKRKV  TDKDIEYLHAIYGYDGIEL  KGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIIEIHTLTI  FEDREMIKQRLSKFE  NIFDKSVLKKLSRRHYTGWGLSAKLINGIRDEKSGNTILD  YLIDDGISNRNFMQLIHDD  ALSFKKKIQKAQIIGDEDKGNIEVVKSLPGSPAIIKKGILQSI  KIVDELVKVMGGRKPES  IVVEMARENQYTNQGKSNSQQLKRLEKSLKELGSKILKE  NIPAKLSKIDNNALQNDRLY</p>

		<p> LYYLQNGKDMYTGDDLDIDRLSNYDIDHIIPQAFLKDNSID  NKVLVSSASNRGKSDDFPS  LEVVKRRTFWYQLLKSRLISQRKFDNLTKAERGGLLPED  KAGFIQRQLVETRQITKHVA  RLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELY  KVREINDFHHAHDAYLNAV  IASALLKKYPKLEPEFVYGDYPKYNSFRERKSAATEKVYFYFYS  NIMNIFKKSISLADGRVIE  RPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVE  EQNHGLDRGPKGLFNANLS  SKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFAVLVKGTI  EKGAKKKITNVLEFQGISI  LDRINYRKDKLNFLEKGYKDIELIIEPKYSLFELSDGSRR  MLASILSTNNKRGEIHKG  NQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEEEL  FYYILEFNENYVGAKKNGK  LLNSAFQSWQNHSDIELCSSFIGPTGSEKGLFELTSRGSAA  DFEFLGVKIPRYRDYTPS  SLLKDATLIHQSVTGLYETRIDLAKLGEG </p>
SEQ ID NO: 4	<i>Francisella tularensis subsp. novicida</i> (strain U112)	<p> MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGVY  ELSKDSYTLLMNNRTARRH  QRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLFNR  RGFSFITDGYSPEYLNIVP  EQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKL  MQKILEFKLMKLCCTDIKDD  KVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQG  NLKELSYHHDKYNIQEFEL  KRHATINDRILDITLLTDDLDIWNFNFEKFDKNEEKLQNQ  EDKDHIQAHLHHFVFAVVK  IKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNFCENL  HNKKYSNLSVKNLVNLIGNL  SNLELKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGE  WRVGVKDKQDKKDGAKYSYKDL  CNELKQKVTKAGLVDFLELDPDPCRTIPPYLDNNNRKPCKQ  SLILNPKFLDNQYPNWQQY  LQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQ  QIASGQRDYKDLARILQF  IFDRVKASDELLNEIYFQAKKLKQKASSELEKLESSKLD  EVIANSQLSQILKSQHTNG  IFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHK  YNNTGRFDDDNQLLYCNHK  PRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFISKWL  EHIRGFKKACEDSLKIQKDN  RGLLNHKINIARNTKKGKCEKEIFNLICKIEGSEDKKGNYKH  GLAYELGVLLFGEPNEASK  PEFDRKIKKFNISYSAQIQQIAFAERKGNANTCAVCSADN  AHRMQQIKITEPVEDNKDK  IILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIK  QVLSAKHQLHIPITESN  AFEFEPALADVKGKSLKDRRKKALERISPENIFKDKNNRIK  EFAKGISAYSGANLTDGDF </p>



		DGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKNKGN RIFCLRDLADNYKCLKQFETT DDLEIEKKIADTIWDANKKDFKFGNYRSFINLTPQEQKAFR HALFLADENPIKQAVIRAI NNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFD YFGIPTIGNRGIAEIRQLY EKVDSDIQAYAKGDKPQASYSHLIDAMLAF CIAADEHRND GSIGLEIDKNYSLYPLDKNT GEVFTKDIFSQIKITDNEFSDKKLVKKAIEGFNTHRQMTR DGIYAENYLPILIHKELNE VRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFVDKPIS IDIQISTLEELRNILTTNN IAATAEYYYYINLKTQKLHEYYIENYNTALGYKKYSKEMEF LRSLAYRSERVKIKSIDDVK QVLDKDSNFIIGKITLPFKKEWQRLYREWQNTTIKDDYEF KSFFNVKSITKLHKKVRKD FSLPISTNEGKFLVKRKTWDNFIYQILNDSRSDRADGTPFI PAFDISKNEIVEAIIISF TSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDI GIATIQYKIDNNSRPKVR VKLDYVIDDDSKINYFMNHSLKSRYPDKVLEILKQSTIIEF ESSGFNKTIKEMLGMKLA GIYNETSNN
SEQ ID NO: 5	<i>Staphylococcus aureus</i>	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVE NNEGRRSKRGARRLKRRR RHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLS EEEFSAALLHLAKRRGVHN VNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKD GEVRGSINRFKTSYVKEA KQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPF GWKDIKEWYEMLMGHCTYF PEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYEYK FQIENVFKQKKKPTLKQIA KEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIE NAELLDQIAKILTIYQS SEDIQEELTNLSEL TQEEIEQISNLKGYTGTHNLSLKAINLI LDELWHTNDNQIAIFNR LKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINA IHKKYGLPNDIIIELAR EKNSKDAQKMINEMQKRNRQTNERIEEIIIRTGKENAKYLI EKIKLHDMQEGKCLYSLEA IPLEDLLNPNFYEV DHIIPRSVSFDNSFNKVLVKQEENSK KGNRTPFQYLSSSDSKIS YETFKKHILNLA KGKGRISKTKKEYLLEERDINRFSVQKDFI NRNLVDTRYATRGLMNNL RSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH HAEDALIINANADFIFKEWKK LDKAKKVMENQMFEKQAESMPEIETEQEYKEIFITPHQIK HIKDFKDYKYSHRVDKKNP RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLI NKSPEKLLMYHHDPQTYQKL

		KLIMEQYGDEKNPLYKYEEETGNYLTKYSKKDNGPVIKKI KYYGNKLNHLDDITDDYPNS RNKVVKLSLKPYPYFDVYLDNGVYKFVTVKNLDVIKKENY YEVNSKCYEEAKKLLKISNQA EFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYRE YLENMNDKRPPRIIKTI ASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG
SEQ ID NO: 6	<i>Streptococcus thermophilus</i> (strain ATCC BAA-491 / LMD- 9)	MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTS KKYIKKNLLGVLLFDSGITAE GRRLKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRL DDSFLVPDDKRDSKYPIFG NLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAH MIKYRGHFLIEGEFNSKNND IQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKK DRILKLFPGEKNSGIFSE FLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLETLLGY IGDDYSDFLAKKLYDAI LLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNI SLKTYNEVFKDDTKNGYA GYIDGKTNQEDFYVYLKLLAEFEGADYFLEKIDREDFLRK QRTFDNGSIPYQIHLQEMR AILDKQAKFYFPLAKNKERIEKILTFRIPYYVGPLARGNSDF AWSIRKRNEKITPWNFED VIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYN ELTKVRFIAESMRDYQFL DSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKG IEKQFNSSLSTYHDLLNII NDKEFLDDSSNEAIIIEIHTLTIFEDREMIKQRLSKFENIFDK SVLKKLSRRHYTGWVK LSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALS FKKKIQKAQIIGDEDKGN IKEVVKSLPGSPAIIKKGILQSIKIVDELVKVMGGRKPESIVV EMARENQYTNQGKSNSQQ RLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLY YLQNGKDMYTGDLDIDRL SNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDVPSLEV VKKRKTFWYQLLKSLLIS QRKFDNLTKAERGGSPEDKAGFIQRQLVETRQITKHVARL LDEKFNNKKDENNRAVRTV KIITLTKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVVA SALLKKYPKLEPEFVYGDY PKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLI EVNEETGESVWNKESDL ATVRRVLSYPQVNVVKKVEEQNHGLDRGKPKGLFNANLS SKPKPNSNENLVGAKEYLDPK KYGGYAGISNSFTVLVKGTEIEKGAKKKITNVLEFQGISILDR INRDKLNFLEKGYKD IELIHELKYSLEFSDGSRMLASILSTNNKRGEIHKGNQIFL SQKFVKLLYHAKRISN TINENHRKYVENHKKEFEELFYIILEFNENYVGAKKNGKL LNSAFQSWQNHSIDELCSSF

		IGPTGSERKGLFELTSRGSAADEFGLGVKIPRYRDYTPSSLL KDATLIHQSVTGLYETRI DLAKLGEG
SEQ ID NO: 7	<i>Actinomyces naeshundii</i> (strain ATCC 12104 / DSM 43013 / JCM 8349 / NCTC 10301 / Howell 279)	MWYASLMSAHHLRVGIDVGTHSVGLATLRVDDHGTPIELL SALSHIHDSGVGKEGKKDHD TRKKLSGIARRARLLHHRRTQLQQLDEVLRDLGFPIPTPG EFLDLNEQTDPYRVWRVRA RLVEEKLPPELRGPAISMAVRHIARHRGWRNPYSKVESLLS PAEESPFMKALRERILATT GEVLDDGITPGQAMAQVALTHNISMARGPEGILGKLHQSDN ANEIRKICARQGVSPDVCKQ LLRAVFKADSPRGSASRVAPDPLPGQGSFRRAPKCDPEFQ RFRIISIVANLRISSETKGE NRPLTADERRHVVTFLTEDSQADLTWVDVAEKLGVHRRD LRGTAVHTDDGERSAARPPID ATDRIMRQTKISSLKTWWEADSEQRGAMIRYLIEDPTDS ECAEIIAELPEEDQAKLDSL HLPAGRAAYSRESLTALSDHMLATTDDLHEARKRLFGVDD SWAPPAEAINAPVGNPSVDR TLKIVGRYLSAVESMWGTPEVIHVEHVRDGFTSERMADER DKANRRRYNDNQEAMKKIQR DYGKEGYISRGDIVRLDALELQGCACLYCGTTIGYHTCQLD HIVPQAGPGSNNRRGNLVA VCERCNRSKSNTPFAVWAQKCGIPHVGVEAIGRVRGWR KQTPNTSSEDLTRLKKEVIAR LRRTQEDPEIDERSMESVAWMANELHHRIAAYPETTMVMV YRGSITAAARKAAGIDSRIN LIGEKGRKDRIDRRHHAVDASVVALMEASVAKTLAERSSL RGEQRLTGKEQTWKQYTGST VGAREHFEMWRGHMLHLTELFNERLAEDKVYVTQNIRLR LSDGNAHTVNPSKLVSHRLGD GLTVQQIDRACTPALWCALTREKDFDEKNGLPAREDRAIR VHGHEIKSSDYIQVFSKRKK TDSRDETPFGAIAVRGGFVEIGPSIHHARIYRVEGKKPVYA MLRVFTHDLLSQRHGDLF SAVIPPQSISMRCAPKLRKAITGNATYLGWVVVGDELEI NVDSFTKYAIGRFLEDFPN TTRWRICGYDTNSKLTKPIVLAEGLENPSSAVNEIVELK GWRVAINVLTKVHPTVRR DALGRPRYSSRSNLPTSWTIE
SEQ ID NO: 8	<i>Neisseria meningitidis</i> <i>serogroup C</i> (strain 8013)	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLG VRVFERAEVPKTGDSLAM ARRLARSVRRLTRRRAHRLLRTRLLKREGVLQAANFDEN GLIKSLPNTPWQLRAAALDR KLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLK GVAGNAHALQTGDFRTPAEL ALNKFEKESGHIRNQRSDYSHTFSRKDLQAEILLFEKQKEF GNPHVSGGLKEGIETLLM TQRPALSGDAVQKMLGHCTFEPAPKAANKNTYTAERFIWL TKLNNLRILEQGSRPLTDT

		<p>ERATLMDEPYRKS SKLTYAQARKLLGLEDTAFFKGLRYGKD  NAEASTLMEMKAYHAISRAL  EKEGLKDKK SPLNLSPELQDEIGTAFSLFKTDEDITGRLKDR  IQPEILEALLKHISFDKF  VQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEE  KIYLPPIPADEIRNPVVLRA  LSQARKVINGV VRRYGSPARIHIETAREV GKSFKDRKEIEK  RQEENRKDREKAAAKFREY  FPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEK  GYVEIDHALPFSRTWDDSF  NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARV  ETSRFPRSKKQRILLQKFDED  GFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN  GQITNLLRGFWGLRKVRAEND  RHHALDAVVACSTVAMQQKITRFVRYKEMNAFDGKTID  KETGEVLHQKTHFPQPWEFFA  QEV MIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSRPEAV  HEYVTPLFVSRAPNRKMSG  QGHMETVKSARKLDEGVSVLRVPLTQLKLDLEK MVNRE  REPKLYEALKARLEAHKDDPA  KAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNH  NGIADNATMVRVDVFEKGDKYY  LVPIYSWQVAKGILPDRAVVQ GKDEEDWQLIDDSFNKFSL  HPNDLVEVITKKARMF GYF  ASCHRG TGNINIRIHDLDHKIGKNGILEGIGVKTALS FQKYQ  IDELGKEIRPCRLK RPP  VR</p>
SEQ ID NO: 9	<i>Listeria innocua</i> <i>serovar 6a</i> (strain ATCC BAA-680 / CLIP 11262)	<p>MKKPYTIGLDIGTNSVGWAVLTDQYDLV KRKMKIAGDSE  KKQIKKNFWGVRLFDEGQTAA  DRRMARTARRRIERRRNRISYLQGIFAEEMSKTDANFFCRL  SDSFYVDNEKRNSRHPFFA  TIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHII  KYRGNFLIEGALDTQNTS  VDGIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEK  VTRKEKLERILKLYPGEKS  AGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLE  SLLALIGDEYAELFVAAK  NAYSAVVLSSITVAETETNAKLSASMIERFDTHEEDLGELK  AFIKLHLPKH YEEIFSNT  EKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIKIE  KENFLRKQRTFDNGAIPHQL  HLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLA  NGQSEFAWLTRKADGEIR  PWNIEEKVDFGKSAVDFIEKMTNKD TYLPKENVLPKHSLC  YQKYL VYNELTKVRYINDQG  KTSYFSGQEKEQIFNDL FKQKRKVKKKDLELFLRNMSHVE  SPTIEGLEDSFNSSYSTYHD  LLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIEQLQ  QFSDVLDGVVLK KLERRHY  TGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQ  LINDSNLSFKSII EKEQVTTA</p>

		<p>DKDIQSIVADLAGSPAIKKGILQSLKIVDELVSVMGYPPQTI  VVEMARENQTTGKGKNNNS  RPRYSLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQN  GKDMYTGQDLDIHNL SNYD  IDHIVPQSFITDNSIDNLVL TSSAGNREKGDDVPPLEIVRKRK  VFEKLYQGNLMSKRKF  DYLTKAERGGGLEADKARFIHRQLVETRQITKNVANILHQR  FNYEKDDHGNTMKQVRIVT  LKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANT  LLKVYPQLEPEFVYGDYHQFD  WFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKY  LDTVKKVMSYRQMNIVKKTE  IQKGEFSKATIKPKGNSSKLIPRKTNWDPMKYGGLDSPNM  AYAVVIEYAKGKNKLVFEKK  IIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYE  CEEGRRRMLASANEAQKGN  QQVLPNHLVTLHHAANCEVSDGKSLDYIESNREMFAELL  AHVSEFAKRYTLAEANLNKI  NQLFEQNKEGDIK AIAQSFVDLMAFNAMGAPASFKFFETTI  ERKRYNNLKELLNSTIYQ  SITGLYESRKRLDD</p>
SEQ ID NO: 10	<i>Pasteurella multocida</i> (strain Pm70)	<p>MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRI  FERAEVPKTGESLALSRR  LARSTRRLIRRAHRLLLAKRFLKREGILSTIDLEKGLPNQA  WELRVAGLERRLSAIEWG  AVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQL  LQSDDYRTPAELALKKFAKE  EGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCK  EHIQQYMTELLMWQKPALSG  EAILKMLGKCTHEKNEFKAAKHTYSAERFVWLTKLNNLRI  LEDGAERALNEEERQLLINH  PYEKSKLTYAQVRKLLGLSEQAIFKHLRYSKENAESATFME  LKAWHAIRKALENQGLKDT  WQDLAKKPDLLDEIGTAFSLYKTDEDIQQYL TNKVPNSVIN  ALLVSLNFDKFIELSLKSL  RKILPLMEQGKRYDQACREIYGHYGEANQKTSQLLPAIPA  QEIRNPVVLRTLSQARKVI  NAIRQYGS PARVHIETGRELGKSFKERREIQKQQEDNRTKR  ESAVQKFKELFSDFSSEP  KSKDILKFRLYEQQHGKCLYSGKEINIHLNEKGYVEIDHA  LPFSRTWDDSFNNKVLVLA  SENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAA  KKQRLLTQVIDDNKFIDRNLN  DTRYIARFLSNYIQENLLL VGKNKKNVFTPNGQITALLRSR  WGLIKARENNNRHHALDAI  VVACATPSMQQKITRFIRFKEVHPYKIENRYEMVDQESGEII  SPHFPEPWAYFRQEVNIR  VFDNHPD TVLKEMLPDRPQANHQFVQPLFVS RAPTRKMSG  QGHMETIKSAKRLAEGISVL  RIPLTQLKPNLLENMVNKEREPALYAGLKARLAEFNQDPA  KAFATPFYKQGGQVKAIRV</p>

		<p>EQVQKSGVLVRENNGVADNASIVRTDVFIKNNKFFLVPIYT  WQVAKGILPNKAIVAHKNE  DEWEEMDEGAKFKFSLFPNDLVELKTKKEYFFGYIYGLDR  ATGNISLKEHDGEISKGKDG  VYRVGVKLALSFEKYQVDELGKNRQICRPQQRQPVR</p>
SEQ ID NO: 11	<p><i>Corynebacterium</i>  <i>diphtheriae</i>  (strain ATCC  700971 / NCTC  13129 / Biotype  gravis)</p>	<p>MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSG  LDPDEIKSAVTRLASSGIA  RRTRRLYRRKRRRLQQLDKFIQRQGWPIELEDYSDPLYP  WKVRAELAASYIADEKERGE  KLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREE  IKRASGQVPETATVGQMV  TLCELGTLKLRGEGGVL SARLQQSDYAREIQEICRMQEIGQ  ELYRKIIDVVFAAESPKGS  ASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRV  DGEKRILSVEEKNLVFDHLV  NLTPKKEPEWVTIAEILGIDRGQLIGTATMTDDGERAGARP  PTHDTNRSIVNSRIAPLVD  WWKTASALEQHAMVKALSNAEVDDFDSPEGAKVQAFFA  DLDDDVHAKLDSLHLPVGRAAY  SEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGE  PVGNPVAVDRVLKTVSRWL  ESATKTWGAPERVIIHVREGFVTEKRAREMDGDMRRRAA  RNAKLFQEMQEKLNVQGKPS  RADLWRYQSVQRQNCQAYCGSPITFSNSEMDHIVPRAGQ  GSTNTRENLVAVCHRCNQSK  GNTPF AIWAKNTSIEGVSVKEAVERTRHWVTDGMRSTDF  KKFTKAVVERFQRATMDEEI  DARSMESVAWMANELRSRVAQHFA SHGTTVRVYRGS LTA  EARRASGISGKLFKFFDGVGKS  RLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQE  APQWREFTGKDAEHRAAWRV  WCQKMEKLSALLTEDLRDDR VVMSNVRLRLGNGSAHKE  TIGKLSKVKLSSQLSVSDIDK  ASSEALWCALTREPGFDPKEGLPANPERHIRVNGTHVYAG  DNIGLFPVSAGSIALRGGYA  ELGSSFHARVYKITS GK KPAFAMLRVYTIDLLPYRNQDLF  SVELKPQTMSMRQAEKKLR  DALATGNAEYLGWL VVDEL VVDTSKIATDQVKA VEAE L  GTIRRWVVDGFFSPSKLRLRP  LQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVV  RRDSLGRVRL ESTAHL PVT  WKVQ</p>
SEQ ID NO: 12	<p><i>Campylobacter</i>  <i>jejuni subsp.</i>  <i>jejuni serotype</i>  O:2 (strain ATCC  700819 / NCTC  11168)</p>	<p>MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGE  SLALPRRLARSARKRLAR  RKARLNHLKHLIANEFKLN YEDYQSFDESLAKAYKGS LISP  YELRFRALNELL SKQDFAR  VILHIARRGYDDIKNSDDKEKGAILKAIKQNEEKLAN YQS  VGEYLYKEYFQKFKENSKE  FTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSK KF  EEEVLSVAFYKRALKDFS</p>

		<p>HLVGNC SFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNT  EGILYTKDDLNALLNEVLK  NGTLTYKQTKKLLGLSDDYEFKGEKGT YFIEFKKYKEFIKA  LGEHNLSQDDLNEIAKDIT  LIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALK  LVTPLMLEGKKYDEACNE  LNLKVAINEDKKDFLPAFNETYKDEVTNPVVLRAIKEYR  KVLNALLKKYGKVHKINIEL  AREVGKNHSQRAKIEKEQENENYKAKKDAELECEKLGLKIN  SKNILKLRLFKEQKEFCAYS  GEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQ  NQEKL NQTPFEAFGNDSAK  WQKIEVLAKNLP TTKKQKRILDKNYKDKEQKNFKDRNLND  TRYIARLV LNYTKDYLDLPL  SDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAK  DRNNHLHHAIDAVIIAYANNS  IVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGF  RQKVLDKIDEIFVSKPER  KKPSGALHEETFRKEEFYQSYGGKEGVLKALELGKIRKV  NGKIVKNGDMFRVDIFKHKK  TNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDE  NYEFCFSLYKDSLILIQTKD  MQEPEFVYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKN  ANEKEVIAKSIGIQNLKVF  EKYIVSALGEVTKAEFRQREDFKK</p>
SEQ ID NO: 13	<i>Rhodobacteracea</i> <i>e bacterium</i>	<p>MRLGLDIGTNSIGWWLCETDRADARVRINGVLAGGVRFIS  DGRDPKSRASLAVDRRAARA  MRRRRDRYLRRRATLMKVLANAGLMPSTPEEAKALELLD  PYELRATGLDQILPLTHLGRA  LFHINQRRGFKSNRKTWGDNESGKIKDATARLDLAILAN  GARTYGEFLHKRRQRAVDPR  HVPTVTRL SIANRDGPDGKEEAGYDFYPDRKHLEEEFRKL  WAAQANFHPELTEDLHDLI  FEKIFYQRPLKEPKVGLCLFTSEERLPKAHPLTQARVLYETV  NQLRVIADGRETRRLTLE  ERDQIIYVLDNKKPTVSLKSMAMKLPALARTLKL RDGERF  TLETGVRDAIACDPVRSSLS  HPDRFGPRWSTL DATAQWEVVS RVRKVQSEAEHAALVDW  LMQAYSIDRNHAEATANAPLP  EGFGRLGQTATTSILERLKADVVTYAEAVAACGWHHSDQ  RTGECLDRLPYYGEVLD RHVI  PGTYDANDDEVTRYGRITNPTVHIGLNQLRRLVNR IIETYG  KPDQIVLELARELKQSEQQ  KRDAIKRIRDTEAAKKRSEKLEELGIEDNGRNRMLLRLWE  DLNPEDAMRRFCPYTGERI  SATMIFDGSCDVDHILPYSRTLDDSFANRTLCLKEANREKR  NQTPWKAWGDAPKWD TIEA  KLKNLPENKRWRFPDAMERFEGEKDFLDRALVDTQYLA  RISR TYMDTLFSEGGHVWVVP  GRLTEMLRRHWGLNSLLSDKDRGAVKAKNR TDHRHHAID  AAVVAATDRSLLNRISRAAGQ</p>

		<p>GEAAGQSAELIARDTPPPWEGFRDDLRLVQLDKIIVSHRADH  GRIDREGRKQGRDSTAGQL  HNDTAYGVVDAMTVVSRTPLLSLKPSDIAVTPKGKNIRDP  QLQKALEIATRKGEGKAFAE  ALRQFAEKAGAYQGLRRVRLIETLQESARVEIGTRSEGGPL  KAYKGDSNHCYELWRLPDG  KVKPQVVTTYEAHAGIEKRPHPAAKRLLRTFKRDMVALER  NGETVICYVQKFNQAGILFL  ASHLESNADARDRDPNDSFTLFRMSPGPMHKAGIRRVSV  EIGRLRDGGAETH</p>
SEQ ID NO: 14	<i>Campylobacter coli</i>	<p>MKIIGFNLGIANIGWALRENDEIIDCGVRVFDIPENPKNGNS  LALERRENKARMKIVKRK  KARMLATKTFLKKEFNVDLSKLFLIGSTQSIYELRTKALSSL  ISKEELSAIILHIAKHRG  YDDSALKNENGTIIEALNKNKEAMLKFKSVGEYFYKNFVQ  NKEVKKIRNTTEDYSNSVPR  SLLKQELDLILDKQKELGLIKNADFKAKLFEIIFKRPLKDFS  NKIGNCIEFFENEKRAAK  NTISACEFVALGKVVNLLKSIEKDIGIVYEKDSINEIMSILD  KTSISYKKIRDILNLPQ  DINFKGLDYSKNNVENSCLVLDLKKLNEFKKALGDGFTNLD  KDILDSIATDITLTKDTATL  KEKLKNYNVLNAEQIEKLSLVFNDHINLSLKALKQIPLM  YEGKRYDEACELCNFTIAK  NQEKNEYLPFEKTRFAKDISSPVVIRAICEFRKLLNDIIRRY  GSVHKIHLELTRDFGIS  FNDRKKIIEIEQNEQSRIKALETIKELKLEETSKNIQIVRLF  DQKGICPYSGLKMDLK  CLDELVIDYIRPYNRSLDDSYSNKVLTFKKNLNDLKQGKTPF  EAFGEDEKLWAEINERIKE  YNGKKRKFIFDKFFKDKKPFDFTEQTLQDTRWLTKLVASY  LNEYLSFLPISEDENTALGY  GEKGSKQHVILSSGMITQMLRNFWYLGFKNHKDYKNNAM  DAIIVAFITNSIIFTFNNFKK  ELDLAKAEFYANKISESDYLLKRKFLPPFSGFKEQALEKVK  NIFVSHSLKIKNKGTLHEL  TPLKIKELKNTYGDLDLAVKLGKIRKYNDKYYANAKGSLV  RTDLFVDKENKFHAVSIYKA  DFSTKLPNKTPATTSNGETKEGIEMNENYNFCMSLYKNTP  IGVKKIKGMKESIICYHGF  NTSGSKITYKKHDNNYHNLSEDEMVFVRKNDKESIVVGKI  LEIKKYSISPSGELSLIENE  KRKWF</p>
SEQ ID NO: 15	<i>Ignavibacteria bacterium</i> ADurb.Bin266	<p>MKNILGLDLGTNSIGWALIDKENNKIIDMGSRIIPMSQDILG  EFGKGNISQTAERTNYR  SIRRLRERYLLRRERLHRVLNILEFLPKHYSDQIDFETRLGK  FKEDTEPKIAYKSTIDET  NSKSRFDFIFKKSFAEMLEDFHQQPELFANDNKIPYDWTI  YFLRKKALTKKIEKEELAW  ILLNFNQKRGYYQLREELEEDTNKKEYVVSLLKVIKIVKGE  DKKNKNRNWYSISLENGWV</p>



		<p>YNATFSTEPQWLMTEKEFLVTEELDENGQVKIVKDKKSDK  EGKEKRRRIPLPSFDEINLM  SKSEPDRIYKKIKAKTETAISNSGKTVGEYIYENLLQNPSQK  IRGKLIRTIERKFYKEEL  KQILQKQKEFHPELQNDLNDYDCVRELYKNNEG HQFLLSK  RDFIHLDDIIFYQRPLKS  QKSLISNCTFEFKKYNVGNEEKIKYLKAIPKSHPLYQEFRF  WQWIYNLRVYRKDDDQDVT  NDYLNDPEKYADLFEFLSNRKEIDQKALLKYFKLKESTHR  WNFVEDKKYPCFETRTLST  RLEKVKDLPPNFLTDQTELQLWHIYSVTDKIEFEKALSTFA  KRNKLDVTTFFVENFKKFP  PFKSEYGSYSGKALKKLLPLMRSGRYWKWDDIDEKTKTRI  DKIITGEFDEDIKNKVREKS  INLT TENHFQGLQVWLASYIVYDRHAEAATINKWDTIEHLE  NYIKEFKQHSLRNPIVEQV  TLEALRVIKDIWKQFGKSAENFFDEIHIELGREMKN TADER  KRLTSQINDNENTNVRKA  LLAELKNDSNIENVRPFSPIQQELLKIYEDGVLNSEIEIPDDIS  KISKTAQPSSSELQRY  KLWLEQKYRSPYTGQVIPLAKLFTTDYEIEHIIPQSRYFDDS  FNNKVICEAAVNKLKDNQ  TGLEFIKNHHGEIVQTVFDNKVKIFEENDYRDFVKTHYIKN  RSKRNKLLMEEIPDKMIER  QINDTRYITKFISALLSNIVRAENNDEGLNSKNLIQVNGKITS  LLRQDWGINDIWN DLIL  PRFLRMNQITNSDAFTRYN DKYQKYLPTVPLELSKNYQSK  RIDHRHHALDALIACATRD  HVNLLNKNYAKSKERYDLNRKRLFEKV VYTHPKTGEKIE  REIPKNFIKPWDTFTVDTKN  FLDTIVVSFKQNLRIINKATNQYQKWVKLNGRNVKKEVKQ  SGINWAIRKPLHKETVAGKV  ELKRIKVPKGKILTATRKNLDTSF DIKTIESITDTGIQKILKN  YLSAKGNDPTIAFSPEG  IEEMNKNITRYNNGKPHRPIYKARIFELGSKFILGLTGNKKA  KYVEAAKGTNLFYAIYVD  ENNKRSFETIPLNIVIERQKQGLSSVPENDDKGNKLLFYLS P  NDLVYVPDEDEIINESYL  DVSNLSNEQKKRLYNVND FSSTCYFTPNRIAKAIAPKEVDL  NYDNNKKKLFSGSYDTKTAS  VNGIQIKDICIKLKADRLGNISKANR</p>
<p>SEQ ID NO: 16</p>	<p><i>Fructobacillus sp.</i> <i>EFB-N1</i></p>	<p>MGYNIGLDIGTGSV GWAALTDEGKLARAKGKNLIGVRLFD  SAQSAAQRRSYRTTRRLSR  RKWRLRLENIFSDEMGMIDENFFARLKYSYVHPKDEVNN  AHYYGGYLFPTQQETHDFHE  KFQTIYHLRLKLMIEDCKFDLREIYLAMHHIVKYRGHFLNS  QSKMTIGDSYNPRDFQQA I  QNYAEAKGLIWSLND AQEMTDVLVGQAGFGLSKKAKAER  LLSAFSFDTKEDKKA IQAILA  GIVGNTTDFTKIFNRERSGDELKKWKLKLDSEAFDEQSQAI  VDELDDDEMELFN AIRQAF</p>

		<p>DGFTLMDLLGDQTSISAAMVKRYQQHHDDLKMOVKEIAKK  QGLSHQDFSKIYTAFLKDDTD  KGMKALLDKADLADDVLVEIQQRIESHDFLPKQRTKANSV  IPYQLHLAELEKIIENQGKY  YPFLDFTFNKAGETINKLVELVKFRVPYYVGPMTAADV  EKAGGDATNHWVKRNEGYEK  SPVTPWNFDQVFNRDQAAQDFIDRLTGTDTYLIGEPDLLKN  SLKYQLFTVLNELLNVKIN  GHKIDEKTKHVLIQDLFKSKKTVSEKAIKDYYSQGMGEIQ  IVGLADKTKFNSNLSSYID  LSKTFDAEFMENPANQELLENIIQIQTVFEDVKIAERELQKL  ALPDEQVQQLAKTHYTGW  GNLSDKLLSTPIIQEGSQKVSILNKLQTTSKNFMSIITDNKFG  VQQWIEQNTAETADSI  QDRIDELTTAPANRGIKQAFNVLFDIQKAMGEEPNRVYLE  FAKETQNSVRTNSRYNRLK  DLYKSKTLDVVKALKEELESQKSSLQSERIGDRLYLFLQ  QGKDMYTGQPINIDKLSTD  YDIDHIIPQAYTKDDSIDNRVLVSRPENARKSDSATYTTEVQ  QSAGGLWKSLKNAGFISQ  KKYDRLTKGGDYSGQKTGFIARQLVETRQIIKNVASLIES  EFSQTKAVAIRSEITADMR  RLVAIKKHREINSFHAFDALLITAAGQYMQARYPDRDGA  NVYNEFDYYTNTYLKELRQS  SSSSQVRRLLKPFQVVGTMAGNENWSEDDTQYLRHVMN  FKNILTTRRNDKDNALNKET  IYAVDPKAKLIGTNKKRQDVSLYGGYIYPYSAYMTLVRAN  GKNLLVKVTISAAEKIKSGQ  IELSEYVQQRPEVKKFEKILINKLAIGQLVNNDGNLIYLTYSY  EFYHNAKQLWLPTEEADL  ISQLNKDSSDEDLIKGFIDILTSPAILKRFPFYELDLKKLVNIR  DKFIAVENKFDILMVIL  KALQLDAAQQKPKVCMIDKKSADWKDYRQRGGIKLSDTSEI  IYQSTTGIFEKRVKISNLL</p>
SEQ ID NO: 17	<i>Pedobacter glucosidilyticus</i>	<p>MTKHILGLDLGTNSIGWAIHQVDNANNVPIQIIAMGSRIPLD  SNDRDQFQKGOAISKNK  DRTTARTQRKGYDRKQLKKSDDFKYSLKKILEKLDIFPTEE  LMKLPTLDLWKLRSDAVSN  IEDITPKQLGRILYMLNQKRGYKSARSEANADKKD TDYVA  EVKGRYTQLKDKGQTLGQYF  YKELSDANQNNTYRVKEKVYPREAYIEEFDIINVQKSK  HSFLTDEVIHSLRNEIYYQ  RKLKSQKGLVSICEFEGFETTYFDKKTQQDKTIFTGPKVAP  RTSPLFQFCKIWEVVNNIS  LTKNPEGSKYKWSDRIPTIEEKQTIANYLQENENLSFIELL  KILQLKKEQVYANKQILK  GIQGNTTFSAIHKIIGNSEHLKFDIETIPSKHFVAVLVDKKTGE  ILDERDSLELNSALEQE  PFYQLWHTIYSIKDLDECKKALIKRFNFEEIEAEKLSKIDFN  KQAFGNKSNKAMRKMLPY</p>

		<p>LMLGYNQSEAESFAGYNRRLTKEEKSKNVSDPELQLLAKN  SLRQPVVEKILNQMINVVNA  IIEKYGKPEEIRVELARELKQSKDEREDADKQNGFNKLLNE  LVATKLTELGLPTTKHYIQ  KYKFIFPAKDKNWKEAQVANQCIYCGDTFNLTEALSGDNF  DVDHIVPKALLFDDSQANKV  LVHRSCNSTKTNNTAYDYITKKGSQALNDYVARVDDWFK  RGIISYGKMQRLKVSFEEYQE  RKKIGKETEADKRIWENFIDRQLRETAYIAKKAKEILEKVC  HNVTSTEGNVTAKLRQLWG  WDNVLNMLQLPKYKELEKKTQTFTQLKEWTSDHGNRK  HQKEEIIWTKRDDHRHHAIDA  LVIACQGGFIQRINTLSSSDVKDEMCKELEEDKTVYNERL  TLENYLLEKKPFSTEEIE  KEADKILVSFKAGKKVATLSKYKATGINEIKGVLVPRGPLH  EQSVYGKIKVIEKDKPLKY  LFENSDKIVNPLIKHLVKTRLLENENNAQAALVTLKNKPIL  LNNKQTEILEKASCYNEAT  VLKYKLQSLKASQIDDIVDEKIKFLIKERLSKFGNKEKEAFK  DILWFNEKKQIPITSIRL  FARPDANNLQVIKKHEKGKNIGFVLSGNNHHIAIYEDKNN  KLIQHICDFWHAVERKRNNI  PVLIEDTSTIWNHLINEDFSESFLNKLPNDSLKLFSLQQNE  MFILGLPKEQSEEAIKSN  NKSLLSKHLYLVWSITDGDYFFRHHLETKNTELKKIDGSKE  SKRYLRLSTKSLVDLNPIK  VRLNHLGEITKIGE</p>
SEQ ID NO: 18	<i>Geobacillus thermodenitrifica ns</i>	<p>MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTG  ESLALPRRLARSARRRL  RRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLR  VEALDRKLNDELARILLH  LAKRRGFRSNRK SERTNKENSTMLKHIEENQSILSSYRTVA  EMVVKDPKFSLHKRNKEDN  YTNTVARDDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWA  SQRPFASKDDIEKKVGFC  TFEPKEKRAPKATYTFQSFTVWEHINKLRLVSPGGIRALTD  DERRLIYKQAFHKNKITFH  DVRTLLNLPDDTRFKGLLYDRNTTLKENEKVRFLELGAYH  KIRKAIDSVYGGAAKSRFP  IDFDTFGYALTMFKDDDIRSYLRNEYEQNGKR MENLADK  VYDEELIEELLNLSFSKFGH  LSLKALRNILPYMEQGEVYSTACERAGYTFTGPKKKQKTV  LLPNIPPIANPVVMRALTQA  RKVVNAIHKKYGSPVSIHIELARELSQSFDERRKMQKEQEG  NRKKNETAIRQLVEYGLTL  NPTGLDIVKFKLWSEQNGKCAYSLQPIEIERLLEPGYTEVD  HVIPYSRSLDDSYTNKVLV  LTKENREKGNRTPAEYLGLGSERWQQFETFLTNKQFSKK  KRDRLRLHYDENEENEFKN  RNLNDTRYISRFLANFIREHLKFA DSDDKQKVYTVNGRITA  HLRSRWNFNKNREESNLHH</p>

	AVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPPQFPQ PWPHFADDELQARLSKNPKE SIKALNLGNYDNEKLESQPVFVSRMPKRSITGAAHQETLR RYIGIDERSGKIQTVVKKK LSEIQDKTGHFPMYGKESDPRTYEAIRQRLLLEHNNDPKKA FQEPLYKPKKNGELGPIIR TIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCV PIYTIDMMKGILPNKAIEP NKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGE EIKIKDLFAYYQTIDSSN GGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRG EKRVGVASSSHSKAGETIR PL
--	---

**[0072]** Further, in some embodiments, fragments of Cas9 or other programmable nuclease that retain DNA cleaving function can be used to generate the fusion proteins. For example, a Cas9 or other programmable nuclease polypeptide fragment is at least about 70% identical, at least about 80% identical, at least about 90% identical, at least about 95% identical, at least about 96% identical, at least about 97% identical, at least about 98% identical, at least about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to a wild type Cas9. In some embodiments, the Cas9 fragment may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to a wild type Cas9.

**[0073]** The Cas9 enzymes or other programmable nuclease disclosed herein also comprises at least one nuclear localization signal (NLS), which is an amino acid sequence that attaches to a protein for import into the cell nucleus by nuclear transport. Generally, the NLS comprises one or more short sequences of positively charged lysines or arginines exposed on the protein surface. These types of classical NLSs can be further classified as either monopartite or bipartite. The major structural difference between the two is that the two basic amino acid clusters in bipartite NLSs are separated by a relatively short spacer sequence (hence bipartite - 2 parts), while monopartite NLSs are not. In some embodiments, the NLS comprises sequence PKKKRKV (SEQ ID NO: 19) of the SV40 Large T-antigen (a monopartite NLS). In other embodiments, the NLS of nucleoplasmin comprises sequence KR[PAATKKAGQA]KKKK (SEQ ID NO: 20). There are also many other types of non-classical NLSs. Different types of NLSs disclosed herein are not meant to be limiting and a person of ordinary skill in the art is able to select a NLS to attach to a Cas9 protein. In some embodiments, the Cas9 protein comprises an N-terminal NLS. In other embodiments, the Cas9 protein comprises a C-terminal NLS. In yet other embodiments, the Cas9 protein comprises both N-terminal and C-terminal NLSs.

**[0074]** In some embodiments, the other CRISPR-related programmable endonucleases often includes CRISPR-associated (Cas) polypeptides or Cas nucleases including Class 1 Cas polypeptides, Class 2 Cas polypeptides, type I Cas polypeptides, type II Cas polypeptides, type III Cas polypeptides, type IV Cas polypeptides, type V Cas polypeptides, and type VI CRISPR-associated (Cas) polypeptides, CRISPR-associated RNA binding proteins, or a functional fragment thereof. Further, Cas polypeptides suitable for use with the present disclosure often include Cpf1 (or Cas12a), c2c1, C2c2 (or Cas13a), Cas13, Cas13a, Cas13b, c2c3, Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a, Cas8a1, Cas8a2, Cas8b, Cas8c, Csn1, Csx12, Cas10, Cas10d, Cas1O, Cas1Od, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Cse1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx1O, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, and Cul966; any derivative thereof; any variant thereof; and any fragment thereof.

**[0075]** Additionally, other site-specific endonucleases that are suitable for the fusion protein composition disclosed herein often comprise zinc finger nucleases (ZFN); transcription activator-like effector nucleases (TALEN); meganucleases; RNA-binding proteins (RBP); recombinases; flippases; transposases; Argonaute (Ago) proteins (e.g., prokaryotic Argonaute (pAgo), archaeal Argonaute (aAgo), and eukaryotic Argonaute (eAgo)); or any functional fragment thereof.

#### hExo1 protein

**[0076]** A programmable nuclease is often tethered to an exonuclease domain so as to effect the results disclosed herein. A number of exonuclease / programmable exonuclease combinations are consistent with the disclosure herein. With respect to the exonuclease, certain exemplary exonucleases suitable for use as part of the fusion protein in present application include MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean. Additional suitable exonucleases are also contemplated. In certain embodiments, human Exo1 (hExo1) is used herein as a part of the fusion protein. Full length hExo1 can be divided into roughly two regions: the N-terminal nuclease region (1-392) (SEQ ID NO: 1) MGIQGLLQFI KEASEPIHVR KYKGQVVAVD TYCWLHKGAI ACAEKLAKGE PTDRYVGFVCM KFVNMLLSHG IKPILVFDGC TLPSKKEVER SRRERRQANL LKGKQLLREG KVSEARECFT RSINITHAMA HKVIKAARSQ GVDCLVAPYE ADAQLAYLNK AGIVQAIITE DSDLLAFGCK KVILKMDQFG NGLEIDQARL GMCRQLGDVF TEEKFRYMCILSGCDYLSSL RGIGLAKACK VLRLANNPDI VKVIKKIGHY LKMNITVPED YINGFIRANN TFLYQLVFDIP IKRKLIPLNA

YEDDVDPETL SYAGQYVDDS IALQIALGNK DINTFEQIDD YNPDTAMPAH  
SRSHSWDDKT CQKSANVSSI WHRNYSRPE SGTVSDAPQL KE), and the C-terminal  
MLH2/MSH1 interaction region (393-846). In some embodiments, the N-terminal nuclease  
region of hExo1 (SEQ ID NO: 1) is used to covalently link to a Cas9 with at least one NLS via a  
peptidyl linker. In other embodiments, a fragment of SEQ ID NO: 1 or other exonuclease domain  
that retains the nuclease function is used herein. For example, the fragment is at least about 70%  
identical, at least about 80% identical, at least about 90% identical, at least about 95% identical,  
at least about 96% identical, at least about 97% identical, at least about 98% identical, at least  
about 99% identical, at least about 99.5% identical, or at least about 99.9% identical to SEQ ID  
NO: 1. In some embodiments, the fragment may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,  
15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40,  
41, 42, 43, 44, 45, 46, 47, 48, 49, 50 or more amino acid changes compared to SEQ ID NO: 1 or  
other untruncated or unmutated domain. The N-terminal nuclease region of the hExo1 is  
exemplary, and additionally suitable Exo1 or other exonuclease sequences can be utilized for the  
purpose disclosed herein by a person of ordinary skill in the art.

**[0077]** An exonuclease such as a hExo1 peptide is connected to a programmable endonuclease  
such as a Cas9 peptide and at least one NLS in some cases using a linker. In some embodiments,  
the linker is a linker peptide. The linker peptides not only serves to connect the protein moieties,  
but in some cases also provides many other functions, such as maintaining cooperative inter-  
domain interactions or preserving biological activity (Gokhale RS, Khosla C. Role of linkers in  
communication between protein modules. *Curr Opin Chem Biol.* 2000;4: 22–27; Ikebe M,  
Kambara T, Stafford WF, Sata M, Katayama E, Ikebe R. A hinge at the central helix of the  
regulatory light chain of myosin is critical for phosphorylation-dependent regulation of smooth  
muscle myosin motor activity. *J Biol Chem.* 1998; 273: 17702–17707; and Chen XY, Zaro J, and  
Shen WC. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* 2014;  
65, 1357-1369 are incorporated herein). The linker peptides can be grouped into small, medium,  
and large linkers with average length of less than or up to  $4.5 \pm 0.7$ ,  $9.1 \pm 2.4$ , and  $21.0 \pm 7.6$   
residues or greater, respectively, although examples anywhere within the set defined by these  
three ranges are also contemplated. In some embodiments, the linker peptide comprises 5 to 200  
amino acids. In other embodiments, the linker peptide comprises 5 to 25 amino acids. In certain  
embodiments, the linker peptide is selected from the group consisting of FL2X (encoded by SEQ  
ID NO: 122 (ggtctccttaaacctgtcttgt)), SLA2X (encoded by SEQ ID NO: 123  
(GGAGGTGGAGGCTCTGGTGGAGGCGGATCA)), AP5X (encoded by SEQ ID NO: 124  
(GCAGAGGCTGCAGCCGCTAAGGCC)), FL1X (encoded by SEQ ID NO: 125

(GCAGAGGCTGCAGCCGCTAAGGAGGCAGCTGCCGCTAAGGCC)), SLA1X, (encoded by SEQ ID NO: 126 (GCACCTGCTCCAGCGCCCGCACCAGCTCCC)) and any combinations thereof. In some embodiments, the linker peptide is SLA2X. Again, these disclosed linker peptides are not meant to be limiting. A person of ordinary skill in the art would be able to select an appropriate linker peptide.

**[0078]** The fusion protein disclosed herein can be fused together directly post-translationally or translated from a polynucleotide (fusion nucleotide) that encodes the disclosed fusion protein in a common open reading frame. In some embodiments, a first nucleic acid sequence encoding hExo1 or the N-terminal nuclease region thereof is ligated to one end of a second nucleic acid sequence encoding a selected linker peptide. Further, the other end of the second nucleic acid sequence is ligated with a third nucleic acid sequence encoding Cas9 enzyme with at least one NLS. Generally, stop codons of the first, second, and third nucleic acid sequences are removed. In some embodiments, the first, second and third nucleic acid sequences are codon optimized or engineered for more efficient transfection or expression in a target cell. Similarly, in some instances, intronic sequences are removed.

**[0079]** FIG. 20 illustrates exemplary fusion proteins with various arrangements of nucleases, Cas9, and other functional domains connected by linkers (L1, L2, and L3). Additional non-limiting examples of the fusion proteins include: hExo1-Cas9-DN1s (or reverse orientation DN1s-Cas9-HR); hExo1-Cas9-DN1s-Geminin(1-110) (or DN1s-Cas9-HR-Geminin); hExo1-Cas9-Geminin(1-110) (or Cas9-hExo1-Geminin); hExo1-Cas9-PCV (or PCV-Cas9-hExo1); hExo1-Cas9-PCV-Geminin(1-110) (or PCV-Cas9-hExo1-Geminin); and hExo1-Cas9-CtIP(1-296) (or CtIP-Cas9-hExo1).

**[0080]** In some embodiments, hExo1-Cas9-DN1s (or reverse orientation DN1s-Cas9-HR) can be a fusion of hExo1(1-352) via linker 1(FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS (noted as NLS in FIG. 20) subsequently fused via linker 2 (either TGS or other) to a fragment of human p53 (1231-1644) with an NLS sequence added at the C-Terminus. In some embodiments, Cas9-HR and Cas9-DN1s can be acting at different steps in the homologous recombination pathway. In some embodiments, the HR-Cas9-DN1s can have increased error free editing efficiency relative to either Cas9-HR or Cas9-DN1s. In some embodiments, cellular toxicity can be greatly reduced relative to increase seen with Cas9-DN1s when compared to Cas9 alone.

**[0081]** In some instances, hExo1-Cas9-DN1s-Geminin(1-110) (or DN1s-Cas9-HR-Geminin) can be fusion of hExo1(1-352) via linker 1(FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS subsequently fused via linker 2 (either TGS or other) to a fragment of

human p53 (1231-1644). DN1s can either have an NLS added to its C-Terminus, which can then be fused to Geminin(1-110) via L3 (any sequence), or fused to Geminin with an NLS sequence at its C-Terminus, which can be fused to DN1s via L3. In some embodiments, the cellular toxicity of the hExo1-Cas9-DN1s-Geminin(1-110) (or DN1s-Cas9-HR-Geminin) can be reduced compared to Cas9. In some embodiments, error free editing efficiency of hExo1-Cas9-DN1s-Geminin(1-110) (or DN1s-Cas9-HR-Geminin) can be increased compared to Cas9. In some embodiments, the error free editing efficiency of hExo1-Cas9-DN1s-Geminin(1-110) (or DN1s-Cas9-HR-Geminin) can be increased compared to Cas9 due to post-translational regulation via geminin of hExo1-Cas9-DN1s-Geminin restricting nuclease activity to S/G2 phase, when endogenous HR is highest in the cell.

**[0082]** In some embodiments, hExo1-Cas9-Geminin(1-110) (or Cas9-hExo1-Geminin) can be a fusion of hExo1(1-352) via linker 1(FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS, possessing or lacking a C-terminal NLS sequence subsequently fused via linker 2 (either TGS or other) to a fragment of Geminin (1-110) either possessing or lacking a C-terminal NLS sequence. In some embodiments, hExo1-Cas9-Geminin(1-110) (or Cas9-hExo1-Geminin) comprises reduced cellular toxicity and increased error free editing efficiency compared to Cas9.

**[0083]** In some embodiments, hExo1-Cas9-PCV (or PCV-Cas9-hExo1 can be a fusion of hExo1(1-352) via linker 1 (FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS, possessing or lacking a C-terminal NLS sequence subsequently fused via linker 2 (either TGS or other) to PCV. In some embodiments, PCV can bind to a specific ssDNA sequence thereby tethering the repair template to the Cas9 complex. In some embodiments, hExo1-Cas9-PCV comprises increased error free editing efficiency compared to Cas9. In some embodiments, hExo1-Cas9-PCV comprises reduced cellular toxicity compared to Cas9.

**[0084]** In some embodiments, hExo1-Cas9-PCV-Geminin(1-110) (or PCV-Cas9-hExo1-Geminin) can be a fusion of hExo1(1-352) via linker 1(FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS, possessing or lacking a C-terminal NLS sequence subsequently fused via linker 2 (either TGS or other) to PCV, which can then be fused to a fragment of Geminin (1-110). In some embodiments, hExo1-Cas9-PCV-Geminin(1-110) (or PCV-Cas9-hExo1-Geminin) comprises higher error free editing efficiency compared to Cas9. In some embodiments, hExo1-Cas9-PCV-Geminin(1-110) (or PCV-Cas9-hExo1-Geminin) comprises higher error free editing efficiency compared to Cas9 due to restriction of nuclease activity to S/G2 phase.



**[0085]** In some embodiments, hExo1-Cas9-CtIP(1-296) (or CtIP-Cas9-hExo1) can be a fusion of hExo1(1-352) via linker 1(FL1X, AP5X or other) to Cas9 possessing or lacking an N-terminal FLAG+NLS, possessing or lacking a C-terminal NLS sequence subsequently fused via linker 2 (either TGS or other) to CtIP. In some embodiments, CtIP can improve error free editing efficiency compared to Cas9 without CtIP. In some embodiments, CtIP can improve error free editing efficiency compared to Cas9 via binding downstream of blocked DSBs (double-strand breaks) and resecting back towards the break using 3'-5' exonuclease activity.

**[0086]** Escheria Coli (E. Coli) version of Exo I

**[0087]** In certain embodiments, the Escheria Coli (E. Coli) version of Exo I (E.Coli ExoI) is used herein as a part of the fusion protein. E.coli Exo1 possesses 3'to 5' exonuclease activity as opposed to the 5' to 3' exonuclease activity of hExo1. The E.Coli ExoI Cas9 fusion can generate much longer deletions than traditional Cas9.

#### Nucleic acid sequence

**[0088]** Some nucleotide constructs consistent with the disclosure comprise nucleic acid encoding an exonuclease such as hExo1. Further, some nucleotide constructs consistent with the disclosure comprise nucleic acid encoding a programmable endonuclease such as a Cas9 or other CRISPR-related programmable endonucleases. In some embodiments, the nucleic acid sequence encoding hExo1 or the N-terminal nuclease region thereof is non-naturally occurring, but the hExo1 or the N-terminal nuclease region thereof encoded by it has an amino acid sequence that is naturally occurring. In some instances, the nucleic acid sequence is different from a naturally occurring hExo1 or the N-terminal nuclease region thereof nucleic acid sequence but encodes a polypeptide identical to hExo1 or the N-terminal nuclease region thereof owing to codon degeneracy. Similarly, the third nucleic acid sequence encoding Cas9 enzyme with at least one NLS is non-naturally occurring, but the Cas9 protein encoded by it has an amino acid sequence that is naturally occurring. In some instances, the nucleic acid sequence is different from a naturally occurring Cas9 nucleic acid sequence but encodes a polypeptide identical to Cas9 owing to codon degeneracy.

#### **Ribonucleoprotein (RNP)**

**[0089]** A ribonucleoprotein (RNP) typically comprises at least two parts: one part comprises a programmable endonuclease such as a Cas9 or other CRISPR-related programmable endonucleases; and the other part comprises a gRNA or other specificity-conveying nucleic acid. Often, a wild type Cas9 enzyme or other Cas or non-Cas programmable endonuclease can be one part of the CRISPR-Cas9 system. The modified Cas9 protein coupled to a fragment of hExo1 via

a linker peptide can also be one part of the CRISPR-Cas9 system. Further, the modified Cas9 protein and a gRNA can form a ribonucleoprotein (RNP).

### gRNA

**[0090]** A ribonucleic acid that comprises a sequence for guiding the ribonucleic acid to a target site on a gene and another sequence for binding to an endonuclease such as Cas9 enzyme is used herein. Often, the ribonucleic acid is a gRNA. In some embodiments, the gRNA is a synthetic gRNA (sgRNA). The gRNA directs the fusion protein complex to a targeted nucleotide sequence of the DNA molecule. The gRNA is a short synthetic RNA composed of a scaffold sequence necessary for Cas-binding and a user-defined about 20 nucleotide spacer that defines the genomic target to be modified. In certain embodiments, a spacer of a gRNA can be designed to recognize the exon 1 of HBB gene. Thus, one can change the genomic target of the Cas protein by simply changing the target sequence present in the gRNA.

**[0091]** There are several ways to deliver gRNA into cells. One is to deliver gRNA into the cells as plasmid DNA. In some embodiments, the nucleic acids encoding the fusion proteins can be cloned into one plasmid or other suitable vectors with a nucleic acid sequence encoding a designed gRNA targeting a gene of interest.

**[0092]** A list of representative gRNA constituents is provided below.

**Table 2.** A list of gRNA sequences.

Seq ID No.	Gene Name	Guide Name	Guide Sequence 5'-3'
SEQ ID NO: 21	HBB	HBB-1	GTAACGGCAGACTTCTCCTC
SEQ ID NO: 22	HBB	HBB-2	GTCTGCCGTTACTGCCCTGT
SEQ ID NO: 23	HBB	HBB-3	GAGGTGAACGTGGATGAAGT
SEQ ID NO: 24	HBG1	HBG1-1	TATCTGTCTGAAACGGTCCC
SEQ ID NO: 25	HBG1	HBG1-2	GCTAAACTCCACCCATGGGT
SEQ ID NO: 26	HBG1	HBG1-3	CAAGGCTATTGGTCAAGGCA
SEQ ID NO: 27	BCL11A	BCL11A-1	AAATAAGAATGTCCCCCAAT
SEQ ID NO: 28	BCL11A	BCL11A-2	CACAAACGGAAACAATGCAA
SEQ ID NO: 29	BCL11A	BCL11A-3	AATATCATTCTGTTCAAAA
SEQ ID NO: 30	CCR5	CCR5-1	TAATAATTGATGTCATAGAT
SEQ ID NO: 31	CCR5	CCR5-2	TGACATCAATTATTATACAT
SEQ ID NO: 32	CCR5	CCR5-3	CTTTTTATTTATGCACAGGG
SEQ ID NO: 33	CXCR4	CXCR4-1	ATCCCCTCCATGGTAACCGC
SEQ ID NO: 34	CXCR4	CXCR4-1	ACTTACACTGATCCCCTCCA
SEQ ID NO: 35	PPP1R12C	PPP1R12C-1	GGAGAGGATGGCCCGGCGGC

SEQ ID NO: 36	PPP1R12C	PPP1R12C-2	ATGGCCCGGCGGCTGGCCCG
SEQ ID NO: 37	PPP1R12C	PPP1R12C-3	GGATGGCCCGGCGGCTGGCC
SEQ ID NO: 38	HPRT	HPRT-1	TAGGTATGCAAATAAATCA
SEQ ID NO: 39	HPRT	HPRT-2	CATACCTAATCATTATGCTG
SEQ ID NO: 40	HPRT	HPRT-3	TAAATTCTTTGCTGACCTGC
SEQ ID NO: 41	HPRT	HPRT-4	TGTAGCCCTCTGTGTGCTCA
SEQ ID NO: 42	HPRT	HPRT-5	AACTAGAATGACCAGTCAAC
SEQ ID NO: 43	HPRT	HPRT-6	GATGATCTCTCAACTTTAAC
SEQ ID NO: 44	FactorVIII	FactorVIII-1	CACTAAAGCAGAATCGCAA
SEQ ID NO: 45	FactorVIII	FactorVIII-2	TGCCTTTACCTTGCGTCCAC
SEQ ID NO: 46	FactorVIII	FactorVIII-3	CCTGTCAGTCTTCATGCTGT
SEQ ID NO: 47	FactorVIII	FactorVIII-4	TCTGCTAGGTCCTACCATCC
SEQ ID NO: 48	FactorIX	FactorIX-1	CTTTCACAATCTGCTAGCAA
SEQ ID NO: 49	FactorIX	FactorIX-2	AAATTCTGAATCGGCCAAAG
SEQ ID NO: 50	FactorIX	FactorIX-3	CGGCCAAAGAGGTATAATTC
SEQ ID NO: 51	FactorIX	FactorIX-4	ATTCTTTATAGACTGAATTT
SEQ ID NO: 52	LRRK2	LRRK2-1	GCTCAGTACTGCTGTAGAAT
SEQ ID NO: 53	LRRK2	LRRK2-2	TGCTCAGTACTGCTGTAGAA
SEQ ID NO: 54	HTT	HTT-1	GAAGGACTTGAGGGACTCGA
SEQ ID NO: 55	HTT	HTT-2	AGCGGCTGTGCCTGCGGCGG
SEQ ID NO: 56	HTT	RHO-1	GCGTACCACACCCGTCGCAT
SEQ ID NO: 57	HTT	RHO-2	CGAGTACCCACAGTACTACC
SEQ ID NO: 58	HTT	RHO-3	CCTGTGGTCCTTGGTGGTCC
SEQ ID NO: 59	CTFR	CTFR-1	ATATTTTCTTTAATGGTGCC
SEQ ID NO: 60	CTFR	CTFR-2	TCTGTATCTATATTCATCAT
SEQ ID NO: 61	SFTPBB	SFTPBB-1	GTGGTACCTCTGGTGGCGGG
SEQ ID NO: 62	SFTPBB	SFTPBB-2	GCTAGCTGTGGCAGTGGCCC
SEQ ID NO: 63	PD1	PD1-1	GAAGGTGGCGTTGTCCCCTT
SEQ ID NO: 64	PD1	PD1-2	ATGTGGAAGTCACGCCCGTT
SEQ ID NO: 65	CTLA-4	CTLA4-1	CCTTGGATTTTCAGCGGCACA
SEQ ID NO: 66	CTLA-4	CTLA4-2	TGCATACTCACACACAAAGC
SEQ ID NO: 67	CTLA-4	CTLA4-3	AGCTGTTTCTTTGAGCAAAA
SEQ ID NO: 68	HLA-A	HLA-A-1	CGGCTCCATCCTCTGGCTCG

SEQ ID NO: 69	HLA-A	HLA-A-2	CCTTCACATTCCGTGTCTCC
SEQ ID NO: 70	HLA-A	HLA-A-3	CCTGCGCTCTTGGACCGCGG
SEQ ID NO: 71	HLA-A	HLA-A-4	CTGAGCCGCCATGTCCGCCG
SEQ ID NO: 72	HLA-B	HLA-B-1	GCAGGAGGGGCCGGAGTATT
SEQ ID NO: 73	HLA-B	HLA-B-2	TGGACGACACCCAGTTCGTG
SEQ ID NO: 74	HLA-B	HLA-B-3	CTCTCCGCTGCTCCGCCTCA
SEQ ID NO: 75	HLA-B	HLA-B-4	GATCTGAGCCGCCGTGTCCG
SEQ ID NO: 76	HLA-C	HLA-C-1	GTAGAACAAAAAAAAAAGACC
SEQ ID NO: 77	HLA-C	HLA-C-2	TGGGCACTGTTGCTGVCTGG
SEQ ID NO: 78	HLA-C	HLA-C-3	GAGAGACTCATCAGAGCCCT
SEQ ID NO: 79	HLA-C	HLA-C-4	CTTCCTCCTACACATCATAG
SEQ ID NO: 80	HLA-C	HLA-C-5	TAGCGGTGACCACAGCTCCA
SEQ ID NO: 81	HLA-DPA	HLA-DPA-1	GAAGGAGACCGTCTGGCATC
SEQ ID NO: 82	HLA-DPA	HLA-DPA-2	TCAAACATAAACTCCCCTGT
SEQ ID NO: 83	HLA-DPA	HLA-DPA-3	AATCTGTTCTGGGCAGGAAG
SEQ ID NO: 84	HLA-DPA	HLA-DPA-4	CCCTGCAGTCATAGAAGTCC
SEQ ID NO: 85	HLA-DQ	HLA-DQ-1	TGTGGAGGTGAAGACATTGT
SEQ ID NO: 86	HLA-DQ	HLA-DQ-2	TCGCTCTGACCACCGTGATG
SEQ ID NO: 87	HLA-DRA	HLA-DRA-1	TGTGGAAGTGAAGAGAGCCCA
SEQ ID NO: 88	HLA-DRA	HLA-DRA-2	CCAGTACCTCCAGAGGTAAC
SEQ ID NO: 89	HLA-DRA	HLA-DRA-3	GATGAGCGCTCAGGAATCAT
SEQ ID NO: 90	LMP-7	LMP-7-1	GCCACTGTCCATGACCCCGT
SEQ ID NO: 91	LMP-7	LMP-7-2	GTGGAGAACATATTTCTGA
SEQ ID NO: 92	LMP-7	LMP-7-3	TGGGCCATCTCAATCTGAAC
SEQ ID NO: 93	LMP-7	LMP-7-4	TGCTGGAAGTGAAGGCGAG
SEQ ID NO: 94	TAP1	TAP1-1	TCATCCAGGATAAGTACACA
SEQ ID NO: 95	TAP1	TAP1-2	GATCAATGCTCGGGCCAACG
SEQ ID NO: 96	TAP1	TAP1-3	ACGCCACTGCCTGTCGCTGA
SEQ ID NO: 97	TAP2	TAP2-1	TGAGGAAGCAAAGTCCCCAG
SEQ ID NO: 98	TAP2	TAP2-2	AGCCGCGTCCACCAGCAGCA
SEQ ID NO: 99	TAPBP	TAPBP-1	TCCTGAAAGGGTTGAACTGT
SEQ ID NO: 100	TAPBP	TAPBP-2	TTTCCGGTCCATGGGCCCA
SEQ ID NO: 101	CUTA	CUTA-1	CTCGGGGTAGCAACAAAAGG

SEQ ID NO: 102	CUTA	CUTA-2	GCCATGGTCAGCAAGACTCG
SEQ ID NO: 103	DMD	DMD-1	TGGCAAAGTCTCGAACATCT
SEQ ID NO: 104	DMD	DMD-2	ATTCGGGGATGCTTCGCAAA
SEQ ID NO: 105	DMD	DMD-3	CTATTATGAAGAATCAAAGC
SEQ ID NO: 106	DMD	DMD-4	CAGTTTTAAAAGACAGGACA
SEQ ID NO: 107	GR/NR3C1	NR3C1-1	CCTGAGCAAGCACACTGCTG
SEQ ID NO: 108	IL2RG	IL2RG-1	CTAGGTTCTTCAGGGTGGGA
SEQ ID NO: 109	IL2RG	IL2RG-2	GTCCTGACAGGGGAGAAAGA
SEQ ID NO: 110	IL2RG	IL2RG-3	TTAGGTTCTCTGGAGCCCAG
SEQ ID NO: 111	IL2RG	IL2RG-4	GTTAGGTTCTCTGGAGCCCA
SEQ ID NO: 112	RFX5	RFX5-1	AAGGATACTTGGACTGGCCC
SEQ ID NO: 113	RFX5	RFX5-2	TCGAGCTTTGATGTCAGGAA
SEQ ID NO: 114	AR/NR3C4	NR3C4-1	ACAGGCTACCTGGTCCTGGA
SEQ ID NO: 115	AR/NR3C4	NR3C4-2	TCTCCCCAAGCCCATCGTAG
SEQ ID NO: 116	AR/NR3C4	NR3C4-3	ACTCTCTTCACAGCCGAAGA
SEQ ID NO: 117	AR/NR3C4	NR3C4-4	TAGCCCCCTACGGCTACT
SEQ ID NO: 118	AR/NR3C4	NR3C4-5	AAGATCCTTTCTGGGAAAGT
SEQ ID NO: 119	AR/NR3C4	NR3C4-6	CATGGTGAGCGTGGACTTTC
SEQ ID NO: 120	TGFBR1	TGFBR1-1	TTGCTTGTTCAAGAAACAAT
SEQ ID NO: 121	TGFBR1	TGFBR1-2	ATTGTGTTACAAGAAAGCAT

#### HDR template sequence

**[0093]** Genome stability necessitates the correct and efficient repair of DSBs. In eukaryotic cells, mechanistic repair of DSBs occurs primarily by two pathways: Non-Homologous End-Joining (NHEJ) and Homology Directed Repair (HDR). NHEJ is the canonical homology-independent pathway as it involves the alignment of only one to a few complementary bases at most for the re-ligation of two ends, whereas HDR uses longer stretches of sequence homology to repair DNA lesions. HDR is the more accurate mechanism for DSB repair due to the requirement of higher sequence homology between the damaged and intact donor strands of DNA. The process is error-free if the DNA template used for repair is identical to the original DNA sequence at the DSB, or it can introduce very specific mutations into the damaged DNA.

**[0094]** As addressed above, HDR methods provide the great freedom in genomic engineering, allowing for as little as single base mutations and up to insertions or deletions of kilo-bases (kb) of DNA. In eukaryotes, HDR rate is governed by the competition between two different

pathways: Homologous Recombination (HR) and Non-Homologous End Joining (NHEJ). The competition between these two pathways begins by competitive binding by either MRN/CtIP complex or Ku 70/80 heterodimer. If MRN/CtIP bind first, they recruit other proteins, including Exonuclease I (ExoI), which possess 5'→3' exonuclease activity<sup>20</sup>. 5' end resection of double strand DNA breaks by either Exo1 or Dna2 at each side of the break commits the DSB to be repaired by the HR pathway. Alternatively, if the Ku 70/80 heterodimer binds, it can then recruit other NHEJ pathway members, including DNA Ligase IV, and eventually repairs the double strand break via NHEJ.

**[0095]** HDR template sequences are needed to be delivered into cells when delivering the CRISPR-Cas9 system to the cells. HDR templates used to create specific mutations or insert new elements into a gene require a certain amount of homology surrounding the target sequence that will be modified. In some embodiments, the 5' and 3' homology arms start at the CRISPR-induced DSB. In general, the insertion sites of the modification can be very close to the DSB, ideally less than 10bp away if possible. In some embodiments, the 5' and 3' homology arm of the HDR template sequences are at least 80% identical to the targeted sequence. Further, in some embodiments, single stranded donor oligonucleotide (ssDON) is utilized for smaller insertions. Each homology arm of the ssDON may comprise about 30-80 bp nucleotide sequence. The length of the homology arm is not meant to be limiting and the length can be adjusted by a person of ordinary skill in the art according to a locus of gene interest and experimental system. For larger insertions such as fluorescent proteins or selection cassettes, double stranded donor oligonucleotide (dsDON) can be utilized as HDR template sequence. In some embodiments, each homology arm of the ssDON may comprise about 800-1500 bp nucleotide sequence. To prevent Cas9 enzyme cleaving the HDR template, in some embodiments, a single base mutation can be introduced in the Protospacer Adjacent Motif (PAM) sequence of the HDR template.

#### **Methods for delivery**

**[0096]** Several different methods are used to deliver ribonucleoproteins and ssDON or other nucleic acids to a cleavage site, such as transfection. Transfection methods can be used to deliver CRISPR-Cas9 or other programmable endonuclease components to cells. Some of exemplary methods can be used to deliver the disclosed modified CRISPR-Cas9 system to cells and additional methods consistent with the disclosure known to a person of ordinary skill in the art can choose a particular method depending on the type of cells and the format of the CRISPR-Cas9 components.

**[0097]** Delivery can be broken into two major categories: cargo and delivery vehicle. Regarding CRISPR/Cas9 cargoes, three approaches are commonly available: (1) DNA plasmid

encoding both the Cas9 protein or other programmable endonuclease and the guide RNA, (2) mRNA for Cas9 or other programmable endonuclease translation alongside a separate guide RNA, and (3) Cas9 protein or other programmable endonuclease with guide RNA (ribonucleoprotein complex). The delivery vehicle used will often dictate which of these three cargos can be packaged, and whether the system is usable in vitro and/or in vivo.

**[0098]** Vehicles used to deliver the gene editing system cargo can be classified into three general groups: physical delivery, viral vectors, and non-viral vectors. The most common physical delivery methods are microinjection, electroporation, and nucleofection. Electroporation enables delivery of the CRISPR machinery in cell types that are difficult to transform using lipid-based delivery systems. Application of a controlled, short electric pulse to the cells forms pores in the cell membrane, allowing entry of foreign material. Nucleofection is a variant of electroporation, in which the electric pulse is optimized such that the nuclear membrane of the cells also forms pores. The CRISPR components are thus directly delivered inside the nucleus. Microinjection is commonly used to inject the Cas9 or other programmable endonuclease and gRNA ribonucleoprotein complex in embryos, although it can also be used in cells. Zebrafish, mouse, and most recently human embryos have been manipulated using this technique.

**[0099]** Viral delivery vectors include specifically engineered adeno-associated virus (AAV), and full sized adenovirus and lentivirus vehicles. Especially for in vivo work, viral vectors have found favor and are the most common CRISPR/Cas9 delivery vectors. AAV, of the Dependovirus genus and Parvoviridae family, is a single stranded DNA virus that has been extensively utilized for gene therapy. While LVs and AdVs are clearly distinct, the way they are utilized for delivery of CRISPR/Cas9 components is quite similar. In the case of LV delivery, the backbone virus is a provirus of HIV; for AdV delivery, the backbone virus is one of the many different serotypes of known AdVs. Both LV and AdV can infect dividing and non-dividing cells; however, unlike LV, AdV does not integrate into the genome. This is advantageous in the case of CRISPR/Cas9-based editing for limiting off-target effects. As is the case with AAV particles, both LV and AdV can be used in in vitro, ex vivo, and in vivo applications, which eases both efficacy and safety testing. In terms of mechanism, this class of CRISPR/Cas9 delivery is like AAV delivery described above. Full viral particles containing the desired Cas9 and sgRNA are created via transformation of HEK 293 T cells. These viral particles are then used to infect the target cell type. The biggest difference between LV/AdV delivery and AAV delivery is the size of the particle; both LVs and AdVs are roughly 80–100nm in diameter. Compared with the 20nm diameter of AAV, larger insertions are better tolerated in these systems. When considering CRISPR/Cas9, additional packaging space for differently-sized Cas9

constructs or several sgRNAs for multiplex genome editing is a significant advantage over the AAV delivery system.

**[0100]** A viral vector can be a modified viral vector, alternatively, it can be an unmodified vector. Often, the modified viral vector is a genetically modified vector. The modified viral vector can show reduced immunogenicity, an increase in the persistence of the vector in the blood stream, or impaired uptake of the vector by macrophages and antigen presenting cells.

**[0101]** The modified viral vector can further comprise a polymer, a lipid, a peptide, a magnetic nanoparticle (MNP), an additional compound, or a combination thereof. The polymer, lipid, or magnetic nanoparticle can be attached to a capsid of the viral vector. The polymer can be a polyethylene glycol (PEG). The polymer can be N-[2-hydroxypropyl] methacrylamide (HPMA), poly(2-(dimethylamino)ethyl methacrylate) (pDMAEMA), or arginine-grafted bioreducible polymers (ABPs). The peptide can be a cell-penetrating peptide, a cell adhesion peptide, or a peptide which binds to a receptor on a cell. The cell can be a tumor cell. Any suitable cell-penetrating peptide can be used. Examples of cell-penetrating peptides include, but are not limited to a polylysine peptide and a polyarginine peptide. The cell adhesion peptide can be an arginylglycylaspartic acid (RGD) peptide. An additional compound can be a compound which binds to a receptor on a cell, such as folic acid.

**[0102]** In some instances, the modified viral vector is a genetically modified vector. The genetically modified vector can have reduced immunogenicity, reduced genotoxicity, increased loading capacity, increased transgene expression, or a combination thereof. In some instances, the genetically modified viral vector is a pseudotyped viral vector. The pseudotyped viral vector can have at least one foreign viral envelope protein. The foreign viral envelope protein can be an envelope protein from a lyssavirus, an arenavirus, a hepadnavirus, a flavivirus, a paramyxovirus, a baculovirus, a filovirus, or an alphavirus. The foreign viral envelope protein can be the glycoprotein G of a vesicular stomatitis virus (VSV). In some instances, the foreign viral envelope protein is a genetically modified viral envelope protein. The genetically modified viral envelope protein can be a non-naturally occurring viral envelope protein.

**[0103]** In some embodiments, the viral vectors are virus-like particles (VLPs). VLPs resemble viruses but are non-infectious because they do not contain viral genetic materials. VLPs have been produced from components of a wide variety of virus families including Parvoviridae (e.g. adeno-associated virus), Retroviridae (e.g. HIV), Flaviviridae (e.g. Hepatitis C virus) and bacteriophages. VLPs can be produced in multiple cell culture systems including bacteria, mammalian cell lines, insect cell lines, yeast and plant cells.



**[0104]** With respect to non-viral vector delivery vehicles, lipid nanoparticles/liposomes can be used herein. A lipid can be a cationic lipid, an anionic lipid, or neutral lipid. The lipid can be a liposome, a small unilamellar vesicle (SUV), a lipidic envelope, a lipidoid, or a lipid nanoparticle (LNP). The lipid can be mixed with the nucleic acid to form a lipoplex (a nucleic acid-liposome complex). The lipid can be conjugated to the nucleic acid. The lipid can be a non-pH sensitive lipid or a pH-sensitive lipid. The lipid can further comprise a polythethylene glycol (PEG).

**[0105]** The cationic lipid can be a monovalent cationic lipid, such as N-[1-(2,3-dioleoyloxy)propyl]-N,N,N-trimethylammonium chloride (DOTMA), [1,2-bis(oleoyloxy)-3-(trimethylammonio)propane] (DOTAP), or  $3\beta$ [N-(N', N'-dimethylaminoethane)-carbonyl] cholesterol (DC-Chol). The cationic lipid can be a multivalent cationic lipid, such as Di-octadecyl-amido-glycyl-spermine (DOGS) or {2,3-dioleoyloxy-N-[2(sperminocarboxamido)ethyl]-N,N-dimethyl-1-propanaminium trifluoroacetate} (DOSPA).

**[0106]** The anionic lipid can be a phospholipid or dioleoylphosphatidylglycerol (DOPG). Examples of phospholipids include, but are not limited to, phosphatidic acid, phosphatidylglycerol, or phosphatidylserine. In some instances, the anionic lipid further comprises a divalent cation, such as  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Mn}^{2+}$ , and  $\text{Ba}^{2+}$ .

**[0107]** The cationic lipid or the anionic lipid can further comprise a neutral lipid. The neutral lipid can be dioleoylphosphatidyl ethanolamine (DOPE) or dioleoylphosphatidylcholine (DOPC). In some instances, the use of a helper lipid in combination with a charged lipid yields higher transfection efficiencies.

**[0108]** The liposome can further comprise a polymer, a lipid, a peptide, a magnetic nanoparticle (MNP), an additional compound, or a combination thereof. The polymer, lipid, or magnetic nanoparticle can be attached to the liposome or integrated into the liposomal membrane. The polymer can be a polyethylene glycol (PEG). The polymer can be N-[2-hydroxypropyl] methacrylamide (HPMA), poly (2-(dimethylamino)ethyl methacrylate) (pDMAEMA), or arginine-grafted bioreducible polymers (ABPs). The peptide can be a cell-penetrating peptide, a cell adhesion peptide, or a peptide which binds to a receptor on a cell. The cell can be a tumor cell. Any suitable cell-penetrating peptide can be used. Examples of cell-penetrating peptides include, but are not limited to a polylysine peptide and a polyarginine peptide. The cell adhesion peptide can be an arginylglycylaspartic acid (RGD) peptide. An additional compound can be a compound which binds to a receptor on a cell, such as folic acid.

## **Kit**

[0109] Disclosed herein are kits and articles of manufacture for use with one or more methods and compositions described herein. The kit can comprise a polynucleotide composition described herein formulated in a compatible pharmaceutical excipient and placed in an appropriate container.

[0110] The kit can include 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. A container can be formed from a variety of materials such as glass or plastic.

[0111] The kit can include an identifying description, a label, or a package insert. The label or package insert can list contents of kit or the immunological composition, instructions relating to its use in the methods described herein, or a combination thereof. The label can be 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. In some instances, the label is used to indicate that the contents are to be used for a specific therapeutic application.

[0112] The kit herein can further comprise one or more reagents that used to deliver the polynucleotide sequences to cells, tissues, or organs.

## **Applications**

[0113] The disclosed RNPs can be introduced into cells using one of the delivery methods disclosed herein to induce homologues recombination of DNA in the cells. Further, the disclosed RNPs can be introduced into cells using one of the delivery methods disclosed herein to induce HDR in cells in vitro or ex vivo. The DNA molecule is contacted with the RNPs. The modified Cas9 protein guided by a gRNA introduces a DSB by cleaving at a location as determined by the hybridization of the gRNA with the DNA molecule. The hExo1 peptide partially digests the cleaved DNA molecule, leaving a 3' or 5' overhang. The HDR template sequences comprising some degrees of sequence homology as the digested DNA molecule promotes and serves as the template for HDR. After HDR, the DNA molecule in the cell comprises a sequence that is identical to the HDR template at the region where homologous recombination occurs.

[0114] By inducing HDR in cells, the cellular toxicity caused by wild type Cas9 protein along with gRNAs is decreased. Cellular toxicity can be measured by several cell viability assays. In some embodiments, tetrazolium reduction assay is used. A variety of tetrazolium compounds

have been used to detect viable cells. The most commonly used compounds include: MTT, MTS, XTT, and WST-1. These compounds fall into two basic categories: 1) MTT which is positively charged and readily penetrates viable eukaryotic cells and 2) those such as MTS, XTT, and WST-1 which are negatively charged and do not readily penetrate cells. The latter class (MTS, XTT, WST-1) are typically used with an intermediate electron acceptor that can transfer electrons from the cytoplasm or plasma membrane to facilitate the reduction of the tetrazolium into the colored formazan product. For example, viable cells with active metabolism convert MTT into a purple colored formazan product with an absorbance maximum near 570 nm. When cells die, they lose the ability to convert MTT into formazan, thus color formation serves as a useful and convenient marker of only the viable cells.

**[0115]** In other embodiments, resazurin reduction assay is used. Resazurin is a cell permeable redox indicator that can be used to monitor viable cell number with protocols similar to those utilizing the tetrazolium compounds. Resazurin can be dissolved in physiological buffers (resulting in a deep blue colored solution) and added directly to cells in culture in a homogeneous format. Viable cells with active metabolism can reduce resazurin into the resorufin product which is pink and fluorescent. The quantity of resorufin produced is proportional to the number of viable cells which can be quantified using a microplate fluorometer equipped with a 530 nm or 560 nm excitation / 590 nm emission filter set. The wavelength can be adjusted according to different types of cells and experimental designs. Resorufin also can be quantified by measuring a change in absorbance; however, absorbance detection is not often used because it is far less sensitive than measuring fluorescence.

**[0116]** Further, the disclosed RNPs herein are used to treat diseases where the causes of the diseases are traced to a locus of chromosomal abnormality. In certain embodiments, a biological sample is obtained from a subject afflicted with a disease. DNA is extracted from the biological sample and sequenced to determine the locus of chromosomal abnormality. Primary cells harboring the chromosomal abnormality are isolated from the subject and cultured *ex vivo*. The RNPs are delivered into the said cultured primary cells using one of the delivery methods disclosed herein. The HDR template sequences are also delivered into the cultured primary cells. In some embodiments, the gRNA moiety comprises at least 10 nucleotides complementary to the targeted locus of chromosomal abnormality. The HDR template sequences comprise an integration cassette flanked by a 5' homology region and a 3' homology region, wherein the 5' homology region and the 3' homology region exhibit at least 80% identity to adjacent segments of the targeted locus. The integration cassette of the HDR template comprises a wild type sequence that corresponds to the locus of chromosomal abnormality as detected in the primary

cells. Upon delivering of the RNPs, the gRNA directs the protein fusion complex to the targeted locus, where the modified Cas protein moiety creates a DSB by cleaving said targeted locus as recognized by the gRNA. The nuclease moiety partially digests the cleaved locus of chromosomal abnormality, leaving a 3' overhang. The presence of the HDR template sequences promotes endogenous repair through HDR. Primary cells with wild type sequence replacing chromosomal abnormality are screened and selected for reintroducing back into the subject.

**[0117]** In some embodiments, primary cells are selected from the group comprising T cells, B cells, dendritic cells, natural killer cells, natural killer cells, macrophages, neutrophils, eosinophils, basophils, mast cells, hematopoietic progenitor cells, hematopoietic stem cells (HSCs), red blood cells, blood stem cells, endoderm stem cells, endoderm progenitor cells, endoderm precursor cells, differentiated endoderm cells, mesenchymal stem cells (MSCs), mesenchymal progenitor cells, mesenchymal precursor cells, differentiated mesenchymal cells, hepatocytes progenitor cells, pancreatic progenitor cells, lung progenitor cells, tracheae progenitor cells, bone cells, cartilage cells, muscle cells, adipose cells, stromal cells, fibroblasts, and dermal cells.

**[0118]** Further, in some embodiments, the gRNA is configured to recognize exon 1 of the human HBB gene. The HDR template is configured to have 5' and 3' arm homology with a functional human HBB gene. In other embodiments, the gRNA is configured to recognize a region of CFTR and the HDR template is designed to have 5' and 3' arm homology with a functional CFTR gene.

**[0119]** Please see a list of single gene disorders with the mutated locus of gene respectively listed in Table 3. Examples of human monogenic diseases, modes of inheritance, and associated genes.

**Table 3.**

<b>Disease</b>	<b>Type of Inheritance</b>	<b>Gene Responsible</b>
Phenylketonuria (PKU)	Autosomal recessive	Phenylalanine hydroxylase (PAH)
Cystic fibrosis	Autosomal recessive	Cystic fibrosis conductance transmembrane regulator (CFTR)
Sickle-cell anemia	Autosomal recessive	Beta hemoglobin (HBB)
Albinism, oculocutaneous, type II	Autosomal recessive	Oculocutaneous albinism II (OCA2)
Glucocorticoid Resistance Syndrome	Autosomal dominant	Glucocorticoid Receptor (GR)
Huntington's disease	Autosomal dominant	Huntingtin (HTT)

Myotonic dystrophy type 1	Autosomal dominant	Dystrophia myotonica-protein kinase (DMPK)
Hypercholesterolemia, autosomal dominant, type B	Autosomal dominant	Low-density lipoprotein receptor (LDLR); apolipoprotein B (APOB)
Neurofibromatosis, type 1	Autosomal dominant	Neurofibromin 1 (NF1)
Polycystic kidney disease 1 and 2	Autosomal dominant	Polycystic kidney disease 1 (PKD1) and polycystic kidney disease 2 (PKD2), respectively
Hemophilia A	X-linked recessive	Coagulation factor VIII (F8)
Hemophilia B	X-linked recessive	Coagulation factor IX (F9)
LRRK2 Linked Parkinson's Disease	Autosomal Dominant	Leucine-Rich Repeat Kinase 2(LRRK2)
Muscular dystrophy, Duchenne type	X-linked recessive	Dystrophin (DMD)
Pulmonary Surfactant Metabolism Disorder 1	Autosomal Recessive	SFTB-B, ABCA3
Hypophosphatemic rickets, X-linked dominant	X-linked dominant	Phosphate-regulating endopeptidase homologue, X-linked (PHEX)
Rett's syndrome	X-linked dominant	Methyl-CpG-binding protein 2 (MECP2)
Spermatogenic failure, nonobstructive, Y-linked	Y-linked	Ubiquitin-specific peptidase 9Y, Y-linked (USP9Y)
X-linked severe combined immunodeficiency (XSCID)	X-linked recessive	Interleukin 2 receptor subunit gamma (IL2RG)

[0120] Moreover, the disclosed RNPs herein are used to introduce genetic modification to confer immunity against diseases. A biological sample is obtained from a subject. DNA is extracted and the locus for the targeted genetic modification is sequenced. Primary cells the subjected are isolated and cultured ex vivo. RNPs and the HDR template sequences are delivered into said cultured primary cells. The gRNA moiety directs the RNPs to the targeted locus to initiate the formation of DSB and DNA digestion to generate the 3' overhang. The HDR template comprises an integration cassette flanked by a 5' homology region and a 3' homology region, wherein the 5' homology region and the 3' homology region exhibit at least 80% identity to adjacent segments of the targeted loci. The integration cassette comprises a wild type

sequence that is different from the subject's sequence at the targeted locus. The presence of the polynucleotide promotes endogenous repair through HDR. Primary cells harboring wild type sequence encoded by the polynucleotide are screened and selected for reintroducing back into the subject.

### **Certain Definitions**

**[0121]** 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.

**[0122]** As used herein, the terms “polypeptide,” “peptide” and “protein” are often used interchangeably herein in reference to a polymer of amino acid residues. A protein, generally, refers to a full-length polypeptide as translated from a coding open reading frame, or as processed to its mature form, while a polypeptide or peptide informally refers to a degradation fragment or a processing fragment of a protein that nonetheless uniquely or identifiably maps to a particular protein. A polypeptide can be a single linear polymer chain of amino acids bonded together by peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. Polypeptides can be modified, for example, by the addition of carbohydrate, phosphorylation, etc. Proteins can comprise one or more polypeptides.

**[0123]** As used herein, the terms “fragment,” “domain,” or equivalent terms refer to a portion of a protein that has less than the full length of the protein and maintains the function of the protein. Further, when the portion of the protein is blasted against the protein, the portion of the protein sequence would align at least with 80% identity to part of the protein sequence.

**[0124]** As used herein, the terms “polynucleotide,” “nucleic acid,” “oligonucleotide,” or equivalent terms, refer to molecules that comprises a polymeric arrangement of nucleotide base monomers, where the sequence of monomers defines the polynucleotide. Polynucleotides can include polymers of deoxyribonucleotides to produce deoxyribonucleic acid (DNA), and polymers of ribonucleotides to produce ribonucleic acid (RNA). A polynucleotide can be single or double stranded. When single stranded, the polynucleotide can correspond to the sense or antisense strand of a gene. A single-stranded polynucleotide can hybridize with a complementary portion of a target polynucleotide to form a duplex, which can be a homoduplex or a heteroduplex. The length of a polynucleotide is not limited in any respect. Linkages between nucleotides can be internucleotide-type phosphodiester linkages, or any other type of linkage. A polynucleotide can be produced by biological means (e.g., enzymatically), either in vivo (in a

cell) or in vitro (in a cell-free system). A polynucleotide can be chemically synthesized using enzyme-free systems. A polynucleotide can be enzymatically extendable or enzymatically non-extendable.

**[0125]** As used herein, the terms “vector,” “vehicle,” “construct” and “plasmid” are used in reference to any recombinant polynucleotide molecule that can be propagated and used to transfer nucleic acid segment(s) from one organism to another. Vectors generally comprise parts which mediate vector propagation and manipulation (e.g., one or more origin of replication, genes imparting drug or antibiotic resistance, a multiple cloning site, operably linked promoter/enhancer elements which enable the expression of a cloned gene, etc.). Vectors are generally recombinant nucleic acid molecules, often derived from bacteriophages, or plant or animal viruses. Plasmids and cosmids refer to two such recombinant vectors. A “cloning vector” or “shuttle vector” or “subcloning vector” contain operably linked parts that facilitate subcloning steps (e.g., a multiple cloning site containing multiple restriction endonuclease target sequences). A nucleic acid vector can be a linear molecule, or in circular form, depending on type of vector or type of application. Some circular nucleic acid vectors can be intentionally linearized prior to delivery into a cell.

**[0126]** As used herein, the term “gene” generally refers to a combination of polynucleotide elements, that when operatively linked in either a native or recombinant manner, provide some product or function. The term “gene” is to be interpreted broadly, and can encompass mRNA, cDNA, rRNA and genomic DNA forms of a gene. In some uses, the term “gene” encompasses the transcribed sequences, including 5' and 3' untranslated regions (5'-UTR and 3'-UTR), exons and introns. In some genes, the transcribed region will contain “open reading frames” that encode polypeptides. In some uses of the term, a “gene” comprises only the coding sequences (e.g., an “open reading frame” or “coding region”) necessary for encoding a polypeptide. In some aspects, genes do not encode a polypeptide, for example, ribosomal RNA genes (rRNA) and transfer RNA (tRNA) genes. In some aspects, the term “gene” includes not only the transcribed sequences, but in addition, also includes non-transcribed regions including upstream and downstream regulatory regions, enhancers and promoters. The term “gene” encompasses mRNA, cDNA and genomic forms of a gene.

**[0127]** As used herein, the terms “subject,” “individual,” or “patient” are often used interchangeably herein. A “subject” can be a biological entity containing expressed genetic materials. The biological entity can be a plant, animal, or microorganism, including, for example, bacteria, viruses, fungi, and protozoa. The subject can be tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro. The subject can be a mammal. The

mammal can be a human. The subject may be diagnosed or suspected of being at high risk for a disease. The disease can be cancer. In some cases, the subject is not necessarily diagnosed or suspected of being at high risk for the disease.

**[0128]** As used herein, the term “in vivo” is used to describe an event that takes place in a subject’s body.

**[0129]** As used herein, the term “ex vivo” is used to describe an event that takes place outside of a subject’s body. An “ex vivo” assay is not performed on a subject. Rather, it is performed upon a sample separate from a subject. An example of an ‘ex vivo’ assay performed on a sample is an ‘in vitro’ assay.

**[0130]** As used herein, the term “in vitro” is used to describe an event that takes place contained in a container for holding laboratory reagent such that it is separated from the living biological source organism from which the material is obtained. In vitro assays can encompass cell-based assays in which cells alive or dead are employed. In vitro assays can also encompass a cell-free assay in which no intact cells are employed.

**[0131]** “Treating” or “treatment” refers 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 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. A therapeutic benefit may refer to eradication or amelioration of symptoms or of an underlying disorder being treated. Also, a therapeutic benefit can be achieved with the eradication or amelioration of one or more of the physiological symptoms associated with the underlying disorder such that an improvement is observed in the subject, notwithstanding that the subject may still be afflicted with the underlying disorder. A prophylactic effect includes delaying, preventing, or eliminating the appearance of a disease or condition, delaying or eliminating the onset of symptoms of a disease or condition, slowing, halting, or reversing the progression of a disease or condition, or any combination thereof. For prophylactic benefit, a subject at risk of developing a particular disease, or to a subject reporting one or more of the physiological symptoms of a disease may undergo treatment, even though a diagnosis of this disease may not have been made.

**[0132]** Certain ranges are presented herein with numerical values being preceded by the term “about.” The term “about” is 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, such as a number that is within 10% of the value of the number that it precedes. In determining whether a number is near to or approximately a specifically recited number, the near or



approximating un-recited number may 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, is encompassed within the methods and compositions described herein are. The upper and lower limits of these smaller ranges may 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 are also included in the methods and compositions described herein.

**[0133]** Mention is frequently made to “Cas9” throughout the disclosure. It is understood that, although Cas9 is a particular embodiment, additional programmable endonucleases are also contemplated, such as Cas12 or others. Accordingly, mention of Cas9 should not always be read to exclude alternate or other programmable endonuclease.

**[0134]** Similarly, “hEXO1” is frequently referred to. It is understood that, although hEXO1 is a particular embodiment, additional programmable endonucleases are also contemplated. Accordingly, mention of hEXO1 should not always be read to exclude alternate or other exonuclease.

**[0135]** 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.

### **Figure Descriptions**

**[0136]** **FIG. 1** shows 9 Cas9-HR fusion proteins. The first fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS via linker FL2X. The second fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS via linker SLA2X. The third fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS via linker AP5X. The fourth fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS via linker FL1X. The fifth fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS via linker SLA1X. The sixth fusion protein comprises a hExo1

(SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS and a C-terminal NLS via linker FL2X. The seventh fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS and a C-terminal NLS via linker SLA2X. And the eight fusion protein comprises a hExo1 (SEQ ID NO:1) coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS and a C-terminal NLS via linker AP5X. The ninth fusion protein comprises a hExo1 (SEQ ID NO:1) directly coupled to a Cas9 (anyone of SEQ ID NO:2-18) with a N-terminal NLS and C-terminal NLS.

**[0137]** **FIG. 2** shows an embodiment of an intended target site for nucleotides cleaving and replacing. The intended target site is about 1kb to the 3' end of the human H2B gene on chromosome 6. This figure also shows an embodiment of a HDR template, which contains a puromycin antibiotic resistance cassette coupled to a CMV promoter at the 5' end and coupled to SV40 poly(A) at the 3' end. Further, the HDR template contains 5' and 3' homology regions to the intended target site described above. A single G->C mutation introduced in the PAM sequence to prevent RNP cutting of the HDR template.

**[0138]** **FIG. 3** shows an embodiment of experiment design. Cells are cultured in a 96-well plate and each well is seeded with about  $2.5 \times 10^4$  cells. Each column of the 96-well plate receives a treatment with different plasmid respectively. Because each column contains 8 wells, each treatment has 8 replicates. The cells in the first column are transfected with plasmids encoding the first fusion protein as shown in **FIG. 1**; the cells in the second column with plasmids encoding the second fusion protein as shown in **FIG. 1**; the cells in the third column with plasmids encoding the third fusion protein as shown in **FIG. 1**; the cells in the fourth column with plasmids encoding the fourth fusion protein as shown in **FIG. 1**; the cells in the fifth column with plasmids encoding the fifth fusion protein as shown in **FIG. 1**; the cells in the sixth column with plasmids encoding the sixth fusion protein as shown in **FIG. 1**; the cells in the seventh column with plasmids encoding the seventh fusion protein as shown in **FIG. 1**; and the cells in the eighth column with plasmids encoding the eighth fusion protein as shown in **FIG. 1**. Further, the cells in the ninth column are transfected with plasmids encoding unmodified Cas9 enzymes. The cells in the tenth column are transfected with plasmids encoding GFPs. The cells in the eleventh column are controls without any treatment.

**[0139]** **FIG. 4** shows a bar graph displaying normalized fold changes of the measured resorufin fluorescence before puromycin selection. The y-axis displays numbers of normalized fold change and the x-axis displays treatments with plasmids encoding the 8 fusion proteins respectively, unmodified Cas9, and GFP and non-transfected controls. It is expected that cells from the control treatment to have the greatest resorufin fluorescence due to minimal cellular

toxicity. The control treatment's resorufin fluorescence measurement is normalized to 1, accordingly, their normalized fold change number is 1. Every other treatment's resorufin fluorescence measurement is compared to the control treatment to obtain the normalized fold change number. It is also expected that all the treatments have some degree of cellular toxicity, therefore, each treatment can have a normalized fold change number smaller than 1. For example, the treatment with wild type Cas9 displays the smallest fold change number compared to the control treatment, which means that the wild type Cas9 transfected cells have the least amount of resorufin fluorescence. In contrast, treatments with plasmids encoding the seventh fusion protein and GFP have similar and the largest fold change number, which indicates that that the transfected cells have the second greatest amount of resorufin fluorescence.

**[0140]** **FIG. 5** shows a bar graph displaying normalized fold change of the measured resorufin fluorescence at day 2 after the cells are transfected with plasmids encoding wild type Cas9 enzymes. The left bar displays a normalized fold change number of cells treated with DMSO and the right bar displays a normalized fold change number of cells treated with PFT- $\alpha$ , which specifically block transcriptional activity of the tumor suppressor p53. The right bar displays a higher number than the left bar, which means the cells treated with PFT- $\alpha$  have increased resorufin fluorescence measurements, therefore PFT- $\alpha$  reduces cellular toxicity. This indicates that the cellular toxicity associated with CRISPR-Cas9 system seen in A549 cells is at least partially dependent on p53, which is the main factor driving Cas9 mediated cellular toxicity seen in other human cell types. A549 cells are positive for p53 activity.

**[0141]** **FIG. 6** shows a bar graph displaying normalized fold change of resorufin fluorescence of cells transfected with RNP plasmids with different gRNA sequences to control cells. Panel **A** of **FIG. 6** shows three gRNA sequences (G1, G2, and G3) designed to target Exon 1 of the HBB gene. In Panel **B** of **FIG. 6**, the y-axis displays numbers of normalized fold change and the x-axis displays columns of NT HBB-G1, NT HBB-G2, NT HBB-G3, and Controls. The control's resorufin fluorescence measurement is normalized to 1, accordingly, their normalized fold change number is 1. The NT HBB-G3 has the smallest normalized fold change number, which indicates it has the least resorufin fluorescence. Panel **C** of **FIG. 6** shows Cas9 HBB-G3 reverse sequence trace indicating generation of INDELS, linking toxicity to nuclease cleavage activity.

**[0142]** **FIG. 7** shows an embodiment of experiment design. Cells are cultured in a 96-well plate and each well is seeded with about  $2.5 \times 10^4$  cells. Each column of the 96-well plate receives a treatment with different plasmid respectively. Because each column contains 8 wells, each treatment has 8 replicates. The cells in the first column are transfected with plasmids encoding the first fusion protein as shown in **FIG. 1**; the cells in the second column with plasmids

encoding the second fusion protein as shown in **FIG. 1**; the cells in the third column with plasmids encoding the third fusion protein as shown in **FIG. 1**; the cells in the fourth column with plasmids encoding the fourth fusion protein as shown in **FIG. 1**; the cells in the fifth column with plasmids encoding the fifth fusion protein as shown in **FIG. 1**; the cells in the sixth column with plasmids encoding the sixth fusion protein as shown in **FIG. 1**; the cells in the seventh column with plasmids encoding the seventh fusion protein as shown in **FIG. 1**; the cells in the eighth column with plasmids encoding the eighth fusion protein as shown in **FIG. 1**; and the cells in the ninth column with plasmids encoding the ninth fusion protein as shown in **FIG. 1**. Further, the cells in the tenth column are transfected with plasmids encoding unmodified Cas9 enzymes, while the cells in the eleventh column are transfected with plasmids encoding unmodified Cas9 enzymes as well as plasmids encoding hExo1 (1-352). The cells in the twelfth column are transfected with plasmids encoding GFPs. The cells in the thirteenth column are controls without any treatment.

**[0143]** **FIG. 8** shows a bar graph displaying normalized fold change of the measured resorufin fluorescence. The y-axis displays numbers of normalized fold change and the x-axis displays treatments with RNP plasmids encoding the 9 fusion proteins respectively and G3 (SEQ ID NO: 23) gRNA. The x-axis further displays two additional treatments: one transfecting the cells with RNP plasmids encoding unmodified Cas9 and G3 gRNA (Cas9 WT); and the other one transfecting the cells with RNA plasmids encoding unmodified Cas9 and hExo1 separately and G3 gRNA (Cas9 WT + Exo1). Moreover, the x-axis displays GFP treatment group and Control group without any treatment. The Control group's resorufin fluorescence measurement is normalized to 1 as the normalized fold change number is 1. Every other treatment's resorufin fluorescence measurement is compared to the Control group's to obtain the normalized fold change number. As expected, different degrees of cellular toxicity are observed from all the other treatments, with the Cas9 and Cas9 + hExo1 groups showing the smallest normalized fold change numbers, which indicates that the transfected cells from the two positive control groups have the least amount of resorufin fluorescence, and additionally demonstrating the necessity for direct fusion of hExo1 to Cas9 for toxicity reduction.

**[0144]** **FIGs. 9A-B** show a bar graph displaying normalized fold change of the measured resorufin fluorescence. **FIG. 9A** shows the G2 and G3 gRNA targeting the exon 1 of the HBB gene. In panel **FIG. 9B**, the y-axis displays numbers of normalized fold change and the x-axis displays treatments with RNP plasmids encoding the seventh fusion protein with G3 gRNA and RNP plasmids encoding an unmodified Cas9 and G2 gRNA and Control. The Control group's resorufin fluorescence measurement is normalized to 1 as the normalized fold change number is

1. Every other treatment's resorufin fluorescence measurement is compared to the Control groups to obtain the normalized fold change number. The RNP plasmids encoding an unmodified Cas9 and G2 gRNA group displays the lowest normalized fold change number, which indicates the transfected cells from this group have the least amount of resorufin fluorescence.

**[0145]** FIG. 10 shows a normalized fold change of resorufin fluorescence of cells transfected with different RNP plasmids targeting exon 1 of HBB gene with or without a single-stranded Homology Directing Repair Template (HDRT). Both Cas9-HR7 with and without HDRT shows increased resorufin fluorescence (hence decreased cellular toxicity). Additionally, addition of HDRT reduces toxicity in both Cas9-HR7 and wild-type Cas9 (NT), however we are unsure whether this affect is specific (requiring homology arms for HBB exon 1), or if the HDRT is simply competing for transfection with the plasmids encoding Cas9-HR7 and Cas9(NT). Regardless, Cas9-HR7 shows reduced toxicity.

**[0146]** FIG. 11A is a diagram of Plasmid PX330 which contains a constitutive promoter for mammalian Cas9 expression, along with U6 promoter driven gRNA expression. This plasmid was modified to produce the various Cas9-HR versions 1-9.

**[0147]** FIG. 11B is an example of the experimental set up wherein cells are seeded in a 96 well glass bottom well plate. Cellular toxicity is quantified two days post-transfection via conversion of resazurin to resorufin, which is then normalized to a non-transfected control to allow for accurate comparisons across experiments. As indicated above the 96 well plate, each column is a different treatment, with 8 rows providing 8 independent replicates for each treatment.

**[0148]** FIG. 11C is a graph showing reduced cellular toxicity in A549 cells with gRNA targeting intergenic region on Chromosome 12. Cas9-HR constructs 1-8 have significantly less toxicity in A549 cells than unmodified Cas9, as shown by the higher normalized fold change values. The averages of Cas9-HR 1-8, Cas9, Cas9+hExo1, GFP and untransfected control (Con) are normalized to the Con average fluorescence. Importantly, physical coupling of Cas9 and hExo1 is necessary for toxicity reduction, as transfection of both Cas9 and hExo1 does not reduce toxicity relative to Cas9 alone. All experiments are done in duplicate independent well plates (16 replicates total), error bars represent the standard error of the mean.

**[0149]** FIG. 11D is a graph showing that treatment with alpha-pifithrin (10micromolar) reduces Cas9 induced cellular toxicity in A549 cells. As an extension of Figure 5, here it can be seen the addition of DMSO does not change toxicity relative to transfection of Cas9 alone, indicating that the effects seen with PFT- $\alpha$  are specific.

**[0150] FIG. 12A** is a diagram of a Puromycin resistance repair template (RT). It contains a 5' homology arm (5'), a strong constitutive viral promotor (pCMV), a Puromycin Resistance gene (Puro), a poly-A sequence (SC40Pa), and a 3' Homology Arm (3'). Below the repair template is the genomic region targeted by guides Int-G2 and G3. The repair template is designed to integrate in the middle of both guide sequences, thereby preventing further Cas9 cleavage. The integration site is in an intergenic region between H2B-B and H2B-A on Chromosome 6, which allows testing of the ability of Cas9-HR to function in both intergenic and coding regions of the genome. Furthermore, both strands are targeting, testing Cas9-HRs compatibility with both sense and anti-sense orientation.

**[0151] FIG. 12B** shows the method used to quantify toxicity of hExo-Cas9 fusions in A549 cells. A549 cells were plated in 96 well plates, 500ng of each plasmid and 100ng of repair template were transfected via a standard Cal-Phos protocol, as described previously.

**[0152] FIG. 12C** is a graph depicting the toxicity of various constructs tested via a resazurin assay. All Cas9-HR constructs show significantly less toxicity (higher normalized fluorescence) than Cas9-NT, with 8 having no statically significant difference in toxicity than repair template only controls. Additionally, Cas9-HRs targeting both sense and anti-sense showed similar reductions in toxicity, indicating Cas9-HR can function in either orientation.

**[0153] FIG. 12D** depicts the method of the assay to measure HDR activity of Cas9-HR8 and Cas9. This assay identifies the rate of HDR via measuring cellular survival to treatment with Puromycin.

Because this is a survival assay, and A549 cells showed significant p53 dependent cellular toxicity to transfection of Cas9, K562 cells (p53 -/-) were used instead in order to facilitate accurate quantification of HDR rate. K562 cells were aliquoted in 12 well plates after electroporation with 500ng of either Cas9-HR8 or Cas9 and 100ng of repair template. After two days DNA was extracted and Puromycin (0.5mg/mL) selection was initiated. After three days of selection, K562 cells were quantified via resazurin in 96 well plates.

**[0154] FIG. 12E** is a depiction of the genomic regions of cells successfully integrated by the Puro-RT. The left primer pair and right primer pair are designed so that one primer binds in the genomic region outside of the repair template, while the other binds a sequence specific to the puromycin cassette. Successful amplification of both 5' and 3' primer sets strongly indicates correct integration of the transgene.

**[0155] FIG. 12F** shows the survival data of K562 cells transfected with either Cas9-HR8 or Cas9 with G2 or G3 gRNA after three days of puromycin treatment. Data was normalized to cells transfected with a plasmid containing the RT. Cas9-HR8 targeting Sense (Int-G2) and anti-

sense (Int-G3) showed greater than a two-fold increase in normalized resorufin fluorescence relative to wildtype Cas9 (NT) after 3 days of puromycin selection. A two-fold increase in resorufin fluorescence translates to at least a two-fold increase in HDR rate, showing that not only can Cas9-HRs can dramatically reduce toxicity, they also can increase HDR rate and Cas9-HR functions in multiple cell types.

[0156] **FIG. 12G** shows an agarose gel after amplification of the target region by both the 5' and 3' primer pairs showing that the repair template had been successfully integrated by 8<sup>th</sup> fusion protein of **FIG. 1** and Cas-9 control (NT). There was no amplification of the target region when transfected with only GFP construct (GFP) or without the template (Con).

[0157] **FIG. 13A** shows the genomic region, including the first two exons of HBB targeted to edit the Human Hemoglobin Beta (HBB) gene. The inset shows a larger version of Exon 1, with a diagram depicting the gRNAs tested. The graph shows data from the toxicity screen of HBB gRNA guides in A549 cells. Toxicity experiments were performed as in **FIG 11D**, with HBB-G3 showing higher toxicity than either HBB-G1 or HBB-G2.

[0158] **FIG. 13B** shows sanger sequencing of the HBB genomic region in the HBB-G3 treated A549 cells. Sanger sequencing shows characteristic noise following Cas9 cleavage and repair via NHEJ pathways, with the bar indicating the gRNA sequence. In this case the noise is 5' as opposed to 3' due to sequence with the reverse primer. Clear cleavage and repair via NHEJ could not be detected in cells treated with HBB-G1 or HBB-G2.

[0159] **FIG. 13C** is a diagram of the wild-type HBB sequence and the SSRT-G3 sequence which introduces the sickle cell (E6V), an EcoRI site which creates a mis-sense mutation, and four silent mismatch mutations (bolded **a**, **a**, **a**, and **g** nucleotide bases) with the HBB-G3 gRNA highlighted by the bar above. Single strand repair template (SSRT)-G3 is 120bp long, with 60bp arms on either side of the predicted cut site. Mutations are designed to prevent gRNA binding upon successful repair

[0160] **FIG. 13D** depicts a HBB editing experiment in which K562 cells or A549 cells are electroporated with Cas9+SSRT-G3, Cas9-HR 1-9+SSRT-G3 or SSRT-G3 alone. After two days the cells are quantified as in **FIG. 11D**. DNA is then extracted and the HBB locus is amplified with two primer pairs. The outer pair is digested with EcoRI to quantify the HDR editing rate and the inner pair can be used for deep sequencing to provide an independent quantification of HDR rate in addition to INDEL rate, allowing for accurate quantification of the HDR/INDEL ratio.

[0161] **FIG. 14** illustrates toxicity assessment of two transfection methods, lipofectamine and calcium phosphate (CalPhos) as determined by transfecting A549 cells with HBB-G3 gRNA and

Cas9-HR fusion proteins 4 and 5 as depicted in **FIG. 1**. The similar results using either Cal-Phos or lipofectamine seen Cas9-HR4 and 5 strongly indicate that the toxicity effects are not dependent on particular transfection reagents/methods. Additionally, lipofectamine transfection of Cas9(NT) showed increased toxicity relative to Cal-Phos transfection, indicating that Cal-Phos transfection may actually underreport the reduction of toxicity by Cas9-HRs compared to likely more efficient lipofectamine transfection.

**[0162]** **FIG. 15** illustrates toxicity assessment by transfecting A549 cells with SSRT HBB repair templates of **FIG. 13A**. Resazurin levels are measured on day 2 after the transfection. Cas9-HR fusion proteins 4 and 8 are less toxic in A549 cells. SSRTs reduces toxicity cellular toxicity, particularly for NT.

**[0163]** **FIG. 16A** shows an agarose gel of EcoRI digestion assay depicting Cas9-HR fusion protein 8 of **FIG. 1** integrating the HBB repair template into the genome of K562 cells. Arrows depict the EcoRI digested products. There are no detectable EcoRI digested products in lanes of Cas9 only (NT), SSRT, and Con (no Cas9). This shows that Cas9-HRs are flexible in repair template choice, and both SSRTs and double strand (DS)RTs can be used for genomic edits.

**[0164]** **FIG. 16B** shows an additional agarose gel of EcoRI digestion assay depicting Cas9-HR fusion proteins 4, 5, 6, 7, and 8 of **FIG. 1** integrating the HBB repair template into the genome of K562 cells. Arrows depict the EcoRI digested products, indicating successful HDR. As expected given previous toxicity results(**Fig 8**), Cas9-HR4 appears to have the highest HDR rate, with all other Cas9-HRs and Cas9(NT) showing some level of successful HDR when compared to digestion of an untransfected control (Con).

**[0165]** **FIG. 16C** shows a western blotting of Cas9-HR fusion proteins 4, 5, 6, 7, and 8 (as shown in **FIG. 1**), Cas9 only (NT), and Con (no Cas9). Arrow indicates detection of Cas9 in Cas9-HR fusion proteins and NT lanes. While amounts appear lower for fusions 4-7, additional blots and IHC (**FIG. 16E**) show that proper expression and localization of all Cas9-HRs, indicating that the reduction of toxicity is not likely due to a reduction in expression levels. As an example, Cas9-HR4 and 8 are some of the lowest and highest expressors of the Cas9-HRs assayed by western blot. If cellular toxicity was truly reduced by reducing expression, it would be expected that Cas9-HR 4 would have the lowest toxicity at every target tested in the genome. However, that is not the case, as **FIG. 11C and FIG. 12C** show that Cas9-HR4 actually has the highest toxicity of all the Cas9-HRs tested, whereas Cas9-HR8 is among the least toxic. Given these results, it is much more likely that expression levels play no significant role in determining cellular toxicity, further evidenced by the fact that Cas9-HR4 has the greatest reduction of toxicity of all Cas9-HRs tested when targeting HBB exo1(**FIG. 8**), thus showing no clear



correlation between expression level and toxicity. Finally, it is interesting to note that Cas9-HR 4 and 8 appear to show complimentary reductions in toxicity at various sites in genome: if Cas9-HR4 reduces toxicity, Cas9-HR8 does not (or less effectively does so), and vice-versa. This may speak to the different local chromatin environments in these different environments, with the possibility that the different linker identities allow for optimal positioning of the hExo1 domain in different chromatin environments. Therefore, the use of different versions of Cas9-HR may allow for reduction of toxicity (and increase in HDR) at virtually all locations throughout the genome.

**[0166] FIG. 16D** illustrates successful expression and purification of Cas9-HR3 from E. Coli monitored via SDS-PAGE with Coomassie staining. Lane L is ladder. Lanes 1 and 8 are soluble fractions of cell lysate. Lanes 2 and 9 are insoluble lysed cell pellet. Lanes 3 and 10 are flow-throughs of the soluble fractions passing through a Nickle (Ni-NTA) column. Lanes 4 and 11 are elution fractions where proteins bound to the Nickle are eluted. Lanes 5 and 12 are follow-throughs of sulphopropyl (SP) cation exchange chromatography resin. Lanes 6 and 13 are elution fractions eluted with 500mM NaCl. Lanes 7 and 14 are elution fractions eluted with 1M NaCl. Lanes 1-7 are from cells transfected with Cas9-HR3. Lanes 8-14 serve as controls for purification protocol and are from E. Coli expressing only unmodified Cas9. Development of successful E. Coli based protein purification protocol for Cas9-HRs allows for both in-vitro tests of Cas9-HR activity, as well as direct RNP transfection and editing of various eukaryotic organisms.

**[0167] FIG. 16E** illustrates immunohistochemistry (IHC) of same transfected cells from **FIG. 16C**. Arrows indicate that Cas9-HR fusions and Cas9 are localized to the nucleus of the cells. Both detection and proper localization of all Cas9-HR4-8 (5-7 assayed and proper localization seen, data not shown) in the nucleus further demonstrate that the reduction of toxicity by Cas9-HRs is not due to improper localization nor significant reduction in expression levels as assayed by IHC.

**[0168] FIG. 17A** illustrates the design of the repair template for an H2BmNeon knock-in experiment. This experiment allows for accurate quantification of HDR rate via properly localized GFP fluorescence in a non-survival based assay.

**[0169] FIG. 17B** illustrates p53-dependent decrease of cellular toxicity induced by Cas-HR fusion proteins 4, 5, 6, and 8 of FIG. 1, Cas9 only (NT), and Con (no Cas9) in epithelial lung cancer cell lines. A549 cells are positive for p53 activity, while H1299 cells are negative for p53 activity. Toxicity as determined by normalized resazurin levels (y-axis) has shown that absence of p53 in H1299 cells yields lower cellular toxicity. In A549 cells, only Cas9-HR8 shows a

significant decrease in toxicity relative to Cas9(NT) , while Cas9-HR4-7 are similar to NT. However, in H1299 cells, toxicity decreases dramatically for Cas9-HR4-7 and NT to roughly the level seen in A549 with Cas9-HR8, while Cas9-HR8 slightly decreases toxicity even further. As with previous experiments it is anticipated that the different orientation of the hExo1 domain due to different linker identity influences the likelihood of end resection, and therefore commitment to HDR. In this case, Cas9-HR8 has the highest rates of HDR, which is directly tested in Fig. 17C. This also further corroborates the results seen in A549 cells with PFT- $\alpha$ , as it is likely that the loss of p53 function in H1299 vs A549 cells drives the significant reduction in toxicity seen in Cas9-HR4-7 and NT. Additionally, it is noted that Cas9 HR8 has reduced toxicity relative to the other fusion proteins in H1299 cells.

**[0170]** **FIG. 17C** illustrates the assessment of successful tagging of H2B (via GFP+ cell quantification) **FIG. 17A** in K562 cells. Arrows in IHC images indicate correct expression and localization of cells with successful H2BmNeon knock-in. The data from this experiment show again that reduction of toxicity in A549 cells is linked with increase in HDR rate in K562 cells, indicating that reduction of toxicity by Cas9-HRs in p53+ cells may serve as a proxy for HDR rate. Importantly, this is a non-survival based assay which also shows an at least two fold increase (2.5X for Cas9-HR8) in HDR rate compared to Cas9(NT). Additionally, this experiment shows that Cas9-HRs can function equally well in both intergenic (**FIG. 12F**) and coding sequences (this experiment).

**[0171]** **FIG. 18A** illustrates the schematic difference between the experimentally verified Cas9 only model and the theoretical Cas9-HR model. The presence of an Exonuclease domain fundamentally changes the predicted in-vitro cleavage pattern. Exo1 has a significant preference for phosphorylated 5' termini vs non-phosphorylated termini. Therefore, theoretically when using PCR products or other pieces of DNA normally lacking 5'-phosphorylated termini that endonuclease cleavage via Cas9 can dominate initially, whereas after cleavage the two fragments will each possess 5'-phosphorylated termini, which can result in rapid degradation via the hExo1 domain.

**[0172]** **FIG. 18B** illustrates an exemplary digestion pattern based on **FIG. 18A**. Only Cas9-HR3 + gRNA and Cas9-HR3 can produce the digested products which demonstrate successful in-vitro nuclease activity. Additionally, though hExo1 strongly prefers phosphorylated 5'-termini, hExo1 can still bind and resect unphosphorylated 5'-termini, so a small amount degradation without gRNAs may be seen with the addition of Cas9-HRs without gRNA.

**[0173]** **FIG. 18C** illustrates an actual agarose example of **FIG. 18A** and **FIG. 18B**. Genomic DNA was amplified with primers amplifying a roughly 950bp region surrounding HBB Exon 1.

Lanes 1 and 2 show Cas9-HR3 with gRNAs HBB-G1 or HBB-G3, Lanes 3 and 4 show Cas9 (NT) with gRNAs HBB-G1 or HBB-G3, Lane 5 is an untreated control. Cas9 cleavage patterns are as expected based on the verified model, with both HBB-G1 and G3 showing strong cleavage, with a clear reduction of the initial product (950 bp) and accumulation of cleavage products (pairs of bands ~550-300bp). The cleavage pattern of Cas9-HR3 also matches the predicted pattern, with a clear reduction in the intensity of the large initial product (950bp), demonstrating that Cas9-HR3 retains functional guided endonuclease activity. Additionally, compared to Cas9, Cas9-HR3 doesn't produce any intermediately sized cleavage products (650-300bp), likely due to digestion via hExo1 domain. Therefore, these results show that Cas9-HR3 shows both expected enzymatic activities (endo- and exo-nuclease) in-vitro.

**[0174] FIG. 18D** illustrates a similar experiment as **FIG. 18C**, which differs by conducting the experiment after leaving enzymes for 2 weeks at 4°C in order to compare protein stability. Lane 1 is digestion pattern from the combination of Cas9-HR3 and gRNA HBB-G1. Lane 2 is digestion pattern from the combination of Cas9 and gRNA HBB-G1. Lane 3 is digestion pattern from the combination of Cas9-HR3 and HBB-G3. Lane 4 is digestion pattern from the combination of Cas9 and HBB-G3. Lane 5 is digestion pattern from Cas9-HR only. Lane 6 is digestion pattern from Cas9 only. Lane 7 is the control where there is neither Cas9 nor gRNA. These results show that both Cas9-HR3 and Cas9 have similar levels of stability.

**[0175] FIG. 19A** illustrates design of H2B integration detection primers. Two sets of primers are designed to bind outside of the 5' and 3' ends of the repair template annealing to sequences only present in the genome, not in the RT, while the others anneal to sequences specific to the repair template, and are not present in the unmodified cells. Successful amplification of both the 5' and 3' set of primers strongly indicates successful and proper tagging of H2B with mNeon.

**[0176] FIG. 19B** illustrates an agarose gel showing PCR products amplified from gDNA extracted from K562 cells transfected with Cas9-HR4,8 and Cas9NT plus H2BmNeon RT along with an untransfected control (lanes 4,8,NT and Con). Amplification with the 5' primers and gDNA from Cas9-HR 4, 8 and Cas9(NT) all show successful amplification of the 5' product, while Con does not, indicating proper integration of the 5' end of the RT. Additionally, the higher amount of amplified product using gDNA from Cas9-HR8 corresponds to the higher rates of HDR seen in **FIG. 17C**.

**[0177] FIG. 19C** PCR products amplified from gDNA extracted from K562 cells transfected with Cas9-HR4,8 and Cas9NT plus H2B-mNeon RT along with an untransfected control (lanes 4,8,NT and Con). While levels of Cas9-HR8 and Cas9 appear similar, given the significantly higher amplification of these two it is likely that the reaction had proceeded past the exponential

phase, making quantification less reliable. Regardless, amplification with the 3' primers and gDNA from Cas9-HR 4, 8 and Cas9(NT) all show successful amplification of the 3' product, while Con lane shows no specific bands, indicating proper integration of the 3' end of the RT.

**[0178] FIG. 19D** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product amplified by the 5' primers from Cas9-HR8. The top trace shows the 5' sequence of the product, with the white bar showing sequences only present in the genome, while the shaded bar shows sequences present in both the RT and genome. The intervening sequences are cropped out, and the bottom trace shows the 3' end of the product. The shaded bar again represents the H2B ORF, while the white bar represents mNeon. Additionally, the shaded bars show the two silent mutations introduced to prevent additional cleavage after transgene integration. Both Cas9-HR4 and Cas9(NT) traces were the same.

**[0179] FIG. 19E** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product amplified by the 3' primers from Cas9-HR8. The top trace shows the 5' sequence of the product, with the white bar showing mNeon, while the shaded bar shows sequences present in both the RT and genome. The intervening sequences are cropped out, and the bottom trace shows the 3' end of the product. The shaded bar again represents the H2B 3' region, with the dashed line showing the transition from genome and RT to only genomic sequences. Additionally, three arrows show SNPs relative to the reference sequences. Cas9-HR4 contained similar mutations, whereas the Cas9 trace became degraded right after the end of the RT. It is much more likely that these represent bonified SNPs, though it cannot be ruled out that Cas9-HRs may induce some errors around the junction site. Direct sequencing of the control cells would help to resolve this.

**[0180] FIG. 19F** illustrates sequencing alignment of the PCR product amplified by the 5' primers. No errors are seen relative to the expected reference sequence.

**[0181] FIG. 19G** illustrates sequencing alignment of the PCR product amplified by the 3' primers. The only changes relative to the expected sequence are seen outside of the RT sequence, and most likely show cell line specific SNPs relative to the reference sequence.

**[0182] FIG. 20** illustrates additional Cas9-HR fusion proteins with combinations of domains linked by at least two linkers to Cas9. These various fusions could possibly increase HDR rate and/or further decrease cellular toxicity.

## EXAMPLES

**[0183]** The following examples are given for the purpose of illustrating various embodiments as described in the present disclosure and are not meant to be limiting in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the

present disclosure. Changes therein and other uses which are encompassed within the spirit of the present disclosure as defined by the scope of the claims will occur to those skilled in the art.

### **Example 1 – Reduced Cellular Toxicity in A549 Cells**

**[0184]** Referring to **FIG. 1**, several plasmid constructs with different polynucleotides were generated. Each polynucleotide encodes a different fusion protein comprised of a hExo1 fragment (Amino acids 1-352 of SEQ ID NO: 1) linked to a Cas9 (any of SEQ ID NO: 2-18) via a specific linker peptide. Some plasmid constructs encoded Cas9 enzymes with one N-terminal nucleus localizing sequence (NLS) or a C-terminal NLS and some plasmid constructs encoded Cas9 enzymes with both the C-terminal NLS and the N-terminal NLS. All plasmid constructs were sequenced to ensure that no mutations occurred in the polynucleotide sequences. Each of the plasmid constructs also contained a nucleotide sequence encoding a gRNA directed to an intended chromosomal site. The intended chromosomal site is in the intergenic region between VSP33A on the 5' and CLIP1 on the 3' on Chromosome 12. This region has no predicted genes or long non-coding RNA. Once the cells were transfected with the plasmids, Cas9-gRNA ribonucleoproteins (RNPs) were formed inside the cells. Control plasmids were prepared to encode unmodified Cas9 (any of SEQ ID NO: 2-18) enzyme.

**[0185]** Human lung carcinoma A549 cells were cultured and about  $2.5 \times 10^4$  cells were plated in 96-well plates, with 8-16 transfection replicates per individual treatment. Each well was then transfected with 62.5 ng of plasmid DNA using a standard Calcium Phosphate transfection technique and incubated overnight for 16-20 hours. Cells were then allowed to recover for one day. Resazurin reduction assay (**FIG. 3**) was used to estimate the number of viable cells in the 96-well plates. Resazurin is a cell permeable redox indicator that can be used to monitor viable cell number. Resazurin was dissolved in physiological buffers (resulting in a deep blue colored solution) and added directly to cells in culture in the 96-well plates in a homogeneous format. Viable cells with active metabolism can reduce resazurin into the resorufin product which is pink and fluorescent. Further, the quantity of resorufin produced is proportional to the number of viable cells which can be quantified using a microplate fluorometer equipped with a 535 nm excitation / 590 nm emission filter set.

**[0186]** Referring to **FIG. 4**, two days after the plasmid DNA transfection, most cells transfected with the plasmids encoding fusion hExo1-Cas9 proteins had statistically increased cellular viability (about 3-4 folds) compared to cells transfected with control plasmids encoding unmodified Cas9 enzymes. Further, cells treated with HDR templates, a control antibody, or GFP plasmids and cells received no treatment have similar cellular viability.

### **Example 2 – Reduced Cellular Toxicity in A549 cells with gRNA targeting HBB gene**

**[0187]** Similarly to experiments conducted in Example 1, several plasmids containing polynucleotide encoding fusion protein hExo1-Cas9 enzymes (**FIG. 1**) were generated. Each of the plasmid constructs also contained a nucleotide sequence encoding a gRNA directed to recognize exon 1 of the human HBB gene. Compared to the experiments conducted in Example 1, an additional control of transfecting cells with plasmids with a nucleotide sequence encoding wildtype Cas9 enzyme and a nucleotide sequence encoding hExo1 separately was incorporated. Three gRNA sequences as listed in Table 2 directed to recognize one of HBB gene's exon were used. Control plasmids were prepared to encode unmodified Cas9 (any of SEQ ID NO: 2-18) enzyme.

**[0188]** Similar cell culture and transfection protocols were used as in experiments conducted in Example 1. Referring to **FIG. 6**, gRNA G3 (SEQ ID NO: 23) has the highest cellular toxicity compared to gRNA G1 (SEQ ID NO: 21) and gRNA G2 (SEQ ID NO: 22). Referring to **FIG. 8**, cells transfected with RNP plasmids in general had higher percentage of viable cells compared to cells with the two control treatments. Further, **FIG. 9B** shows that RNP plasmids with the seventh fusion protein (**FIG. 1**) and G2 gRNA had less cellular toxicity compared to RNP plasmids with the unmodified Cas9 and G3 gRNA.

### **Example 3 - Treating Sickle Cell Anemia in a Patient**

**[0189]** A biological sample is obtained from a subject afflicted with sickle cell anemia. Genomic DNA is extracted from the biological sample and sequenced to verify a single nucleotide substitution (A to T) in the amino acid 6 codon of the  $\beta$ -globin gene. This mutation converts a glutamic acid codon (GAG) to a valine codon (GTG). Hematopoietic stem cells are isolated from the bone marrow cavity of the patient and cultured *ex vivo*. Nucleic acid vectors encoding the protein fusion complex of the hExo1-Cas9 and the gRNA moiety are delivered into the cultured hematopoietic stem cells. Further, the DNA template sequences with an integration cassette encoding the wild type sequence of exon 1 of  $\beta$ -globin gene are delivered to the cultured hematopoietic stem cells. The gRNA moiety comprises at least 10 nucleotides complementary to the GTG locus of exon 1 of the  $\beta$ -globin gene. The DNA template sequence comprises an integration cassette flanked by a 5' homology region and a 3' homology region, wherein the 5' homology region and the 3' homology region exhibit at least 80% identity to the segments flanking the GTG locus of exon 1. The integration cassette of the polynucleotide comprises a wild type GAG sequence that corresponds to the locus of chromosomal abnormality as detected in the primary cells. Upon delivering of the nucleic acids encoding the RNPs and DNA template sequences into the cultured hematopoietic stem cells, the gRNA directs the engineered hExo1-

Cas9 proteins to the GTG locus, where the Cas9 portion of the engineered hExo1-Cas9 proteins creates a DSB. The hExo1 portion of the engineered hExo1-Cas9 proteins partially digests the cleaved GTG locus of, leaving a 3' overhang. The presence of the DNA template sequences promotes endogenous repair through HDR, where the integration cassette with the correct wild type sequence, GAG, at amino acid 6 of exon 1 of the  $\beta$ -globin gene is inserted into the chromosome of the hematopoietic stem cells. Hematopoietic stem cells with corrected GAG sequence is screened for and selected to be transplanted back into the patient.

#### **Example 4 – Reduced Cellular Toxicity in A549 cells with gRNA targeting intergenic region on Chromosome 12**

[0190] Similarly to experiments conducted in Example 2, several plasmids containing polynucleotide encoding fusion protein hExo1-Cas9 enzymes (**FIG. 11A** and **FIG. 1**) were generated. Each of the plasmid constructs also contained a nucleotide sequence encoding a gRNA directed to recognize an intergenic region on Chromosome 12, of which A549 cells have two copies. Compared to the experiments conducted in Example 3, the control of transfecting cells with plasmids with a nucleotide sequence encoding wildtype Cas9 enzyme and a nucleotide sequence encoding hExo1 separately were also incorporated. Control plasmids were prepared to encode unmodified Cas9 (any of SEQ ID NO: 2-18) enzyme.

[0191] Similar cell culture and transfection protocols were used as in experiments conducted in Example 2. Roughly  $2.5 \times 10^4$  cells were plated in 96 well plates, with 8-16 replicates per individual experiment, as diagramed in **FIG. 11B**. Referring to **FIG. 11C**, cells transfected with PX330 plasmids in general had a much higher percentage of viable cells compared to cells with the two control treatments, 3-4 fold increase in cellular viability. **FIG. 11C** also shows that it is the fusion of hExo1 that causes the decrease in cellular toxicity as the co-expression of Cas9 and hExo1 do not affect cellular toxicity. **FIG. 11D** shows that treatment with alpha pfithrin of cells transfected with wild type Cas9 reduces the toxicity caused by the activity of Cas9. The inactivation of the Cas9 shown in **FIG. 11D** indicates that the cause of toxicity of Cas9 treatment in A549 cells is at least partly due to activation of P53 based on apoptosis, the same as in Ihrey et al. and Haapaniemi et al.

#### **Example 5 – Quantification of HDR and INDELS rates of hExo-Cas9 fusions in A549 cells**

[0192] Cas-9 hExo1 fusions were used to integrate an antibiotic resistance cassette into a locus on Chromosome 6 of A549 cells. The Puromycin resistance repair template is diagramed in **FIG. 12A**. It contains a 5' Homology Arm (5'), a strong constitutive viral promoter (pCMV), a Puromycin Resistance gene (Puro), a poly-A sequence (SV40Pa), and a 3' Homology Arm (3'). Below the repair template shows the genomic region targeted by guides Int-G2 and G-3. The

repair template is designed to disrupt integrate in the middle of both guide sequences, thereby preventing further Cas9 cleavage. The success of the integration is quantified by antibiotic selection. A549 cells have only one copy of Chromosome 6. The target integration site is ~1kb to the 3' end of the human H2B gene on Chromosome 6. The region has no predicted genes.

**[0193]** Similar cell culture and transfection protocols were used as in experiments conducted in Example 4. Roughly  $2.5 \times 10^4$  cells were plated in 96 well plates, with 8-16 replicates per individual experiment, as diagramed in **FIG. 12B**.

**[0194]** **FIG. 12C** shows that Cas9-HR8 with G2 gRNA and G3 RNA, respectively, showed the greatest survival rate of A549 cells in Day 2 as compared to the other fusion proteins and Cas9. Due to this result, Cas9-HR8 was used in the following example.

#### **Example 6 – Quantification of HDR and INDELS rates of hExo-Cas9 fusions in K562 cells**

**[0195]** Compared to the experiment in Example 5, K562 cells were used and Neon (Thermo Fisher) electroporation was used. K562 cells lack P53 function. In light of the results of Example 5, it was important to remove the variable of the activation of P53 by the activity of Cas9 as this would differ between fusion Cas9 and wild type Cas9, introducing the possibility of effecting the results of the antibiotic screen.

**[0196]** K562 cells were electroporated with 500ng of each plasmid and 100ng of repair template as shown in **FIG. 12D**. After two days, DNA is extracted from ~1/10 of surviving cells and used for analysis of Puro RT genomic integration. The following day, 0.5 mg/mL Puromycin is added and after three days cellular survival is quantified with the standard resazurin assay as shown in **FIG. 12D**.

**[0197]** Quantification of toxicity was performed as in Example 5, with the addition of fusion construct 9. **FIG. 12 F** shows a dramatic reduction in cellular toxicity between Cas9-HR 8 (with gRNA G2 and G3) as compared to Cas9 (with gRNA G2 and G3), double the amount of surviving cells.

**[0198]** Successful amplification using primer specific for the genome and specific for the repair template demonstrates successful integration of the repair template and that the reduction in toxicity of the Cas9-HR series of constructs is not due to lack of nuclease activity. There may be indication that the Cas9-HR series has a higher editing efficiency than Cas9.

**[0199]** **FIG. 12E** shows a depiction of the genomic region of cells which successfully integrated the repair template. Transfected cells are quantified on day 2. DNA is extracted from one well of each treatment. After 7 days of 1 microgram/milliliter puromycin treatment, cells are quantified with Resazurin. DNA is extracted from another row of cells. The insertion junctions are amplified with left and right primer pairs. Deep sequencing can be performed to identify



INDEL rates in cells with successful HDR. DNA from Day 2 is used to quantify INDEL rates in HDR-cells by amplification with the left and right primer as seen in **FIG. 12E**.

**[0200]** **FIG.12G** depicts the results of gel electrophoresis on the amplification products, it shows that K562 cells transfected with either Cas9-HR8 (8) or Cas9 with gRNA G2 or G3 (NT) successfully produced amplicons with both primer pairs depicted in **FIG. 12E** while either GFP or transfected cells did not. This indicates that the repair template was successfully integrated.

#### **Example 7– Determining relationship between toxicity and Cas9 activity**

**[0201]** As seen in Example 2, different guide RNAs can have radically different cleavage rates and toxicities. Constructs with unmodified Cas9 and guides targeting regions shown in **FIG. 13A** were transfected into A549 cells using the same method as Example 4. Toxicity was quantified using Resazurin as in Example 1.

**[0202]** DNA was extracted from cells transfected with HBB-G1, HBB-G2, and HBB-G3, amplified with the outer primer pair in **FIG. 13D** and sent for Sanger sequencing. Only HBB-G3 showed significant cleavage as evidenced by the characteristic increase in noise following the cut-site shown in **FIG. 13B**. This indicates that toxicity is a good proxy for Cas9 nuclease activity in A549 cells. Guide RNA HBB-G3 was therefore used in Example 8.

#### **Example 8- Editing known disease loci with Cas9-HR**

**[0203]** Similar to Example 7, K562 cells are used because they lack P53 activity as well as because they share more similarities to hematocytes than A549 cells.

**[0204]** The gRNA of Example 7, HBB-G3, is transfected with Cas9 and Cas9-HR 1-9 respectively to introduce multiple mutations into the HBB locus of K562 cells. The first mutation chosen is Sickle Cell E6V mutation. The Sickle Cell E6V mutation is made along with an additional mutation creating an EcoRI restriction site and two silent mutations designed to prevent re-cutting of the repair template once integrated into the genome, in addition to 60bp homology arms on each side of the predicted cut-site.

**[0205]** Transfection is achieved with electroporation. Two days after electroporation, toxicity assays with Resazurin are conducted as in Example 6. DNA is also harvested and the HBB locus is amplified to prepare for deep sequencing to measure INDELS and HDR rate. Alternatively, DNA can be digested with EcoRI to measure target efficiency. **FIG. 16A** and **FIG. 16B** illustrate that upon integration of the repair template, the genomic locus can now be digested with EcoRI. EcoRI digested amplicons can be observed in Cas9-HR4, Cas9-HR5, Cas9-HR6, Cas9-HR7, and Cas9-HR8 lanes. **FIG. 16C**, **FIG. 16D**, and **FIG. 16E** confirm that Cas9-HR is expressed and localized to the nucleus of the transfected cells.

### **Example 9-Editing CD34+ hematopoietic stem cells**

**[0206]** The experiments of Example 8 are repeated on CD34+ cells. The gRNA from Example 8, HBB-G3, is transfected with Cas9 and Cas9-HR 1-9 respectively to introduce multiple mutations into the HBB locus of K562 cells. The first mutation chosen is Sickle Cell E6V mutation. The Sickle Cell E6V mutation is made along with an additional mutation creating an EcoRI restriction site and two silent mutations designed to prevent re-cutting of the repair template once integrated into the genome, in addition to 60bp homology arms on each side of the predicted cut-site.

**[0207]** Transfection is achieved with electroporation. Two days after electroporation, toxicity assays with Resazurin are conducted as in Example 6. DNA is harvested and the HBB locus is amplified to prepare for deep sequencing to measure INDELS and HDR rate. Alternatively, DNA is digested with EcoRI to measure target efficiency.

### **Example 10-In-vitro nuclease activity of Cas9-HR3**

**[0208]** A 954bp piece of DNA was amplified from wildtype K562 cells using standard Taq DNA polymerase and HBB-out-4-F (5'-aacgatcctgagacttccaca-3' (SEQ ID NO: 127)) and HBB-out-5-R (5'-tgcttaccagctgtgattcc-3' (SEQ ID NO: 128)), T<sub>m</sub>=56 for 35 cycles, and purified using the Qiagen PCR cleanup kit. Next, HBB-G1 (5'-guaacggcagacuucuccuc-3' (SEQ ID NO: 129),IDT) or HBB-G3 (5'-gaggugaacguggaugaagu-3' (SEQ ID NO: 130),IDT) were combined with tracrRNA (IDT) at final concentrations of 1μM each in duplex buffer (IDT). The RNA was heated for 5 minutes at 95°C, then allowed to cool to room temperature. Cas9 or Cas9-HR3 were then combined with either HBB-G1 or HBB-G3 guide RNA complex and amplified DNA at a 10:10:1 molar ratio (30nM: 30nM: 3nM) in 1X Cas9 reaction Buffer (50mM Tris, 100mM NaCl, 10mM MgCl<sub>2</sub>, 1mM DTT, pH7.9) and incubated for 1hr at 37°C, after which 1μL of Proteinase K was added and the reaction was incubated for an additional 20 minutes at 50°C. The samples were then electrophoresed on a standard 1% TAE agarose gel and imaged.

**[0209]** **FIG. 18A** illustrates the mechanistic modeling of the Cas9-HR. Cas9 binds to the intended site, cuts, and then remains bound until digested away with proteinase K. As Cas9-HR possesses additional 5'→3' exonuclease activity, a more complex pattern is expected. Importantly, it has been shown that hExo1 has roughly 10X the affinity for phosphorylated 5'-double strand DNA ends as for unphosphorylated. This leads to two important consequences. First, it is expected that there would be some small digestion of the PCR without addition of any gRNA, which is not generally expected to happen with Cas9. Changing the nature of the primers used to amplify the DNA fragment (either with 5'-phosphates or thioester bonds) can either increase or decrease this degradation respectively. Second, since cleavage of double-stranded-

DNA (dsDNA produces ends with 5'-phosphates, it is expected that either the original Cas9-HR or other unbound Cas9-HR molecules resect the dsDNA in 5'→3' generating a mix of various dsDNA, double stranded and single-stranded (ds::ss) DNA, and ssDNA products. **FIG. 18B** illustrates an anticipated Cas9 and Cas9-HR digestion pattern based on the mechanism of **FIG. 18A**. **FIG. 18C** illustrates an actual agarose example of **FIG. 18A** and **FIG. 18B**. Lanes 1 and 2 show Cas9-HR3 targeting either HBB-G1 or HBB-G3, Lanes 3 and 4 show Cas9 (NT) targeting either HBB-G1 or HBB-G3, Lane 5 is Untreated DNA. **FIG. 18D** illustrates a similar experiment as **FIG. 18C** and differs from **FIG. 18B** by conducting the experiment after leaving enzymes for 2 weeks at 4°C in order to compare protein stability. Lane 1 is digestion pattern from the combination of Cas9-HR3 and gRNA HBB-G1. Lane 2 is digestion pattern from the combination of Cas9 and gRNA HBB-G1. Lane 3 is digestion pattern from the combination of Cas9-HR3 and HBB-G3. Lane 4 is digestion pattern from the combination of Cas9 and HBB-G3. Lane 5 is digestion pattern from Cas9-HR only. Lane 6 is digestion pattern from Cas9 only. Lane 7 is the control where there is neither Cas9 nor gRNA. **FIG. 18C** and **FIG. 18D** demonstrate that the digestion pattern correspond to the mechanism as shown in **FIG. 18A** and **FIG. 18B**.

#### **Example 11- hH2B genomic integration and genomic validation**

[0210] Cas9-HR and Cas were utilized to introduce an hH2b fragment into the H2B genomic locus. Primers were designed so that the genomic primer is outside of the H2B-mNeon repair template (RT), while the other is RT specific (within mNeon) as shown in **FIG 19A**. Sequences for 3' primers are H2B-RT-3'-F: 5'-aggcctttaccgatgtgatg-3' (SEQ ID NO: 131), H2B-RT-3'-R: 5'-acggagtctcgctctgtcac-3' (SEQ ID NO: 132). Sequences for 5' primers are H2B-RT-5'-F: 5'-caaaactgcaaggctgcaata-3' (SEQ ID NO: 133), H2B-RT-3'-R: 5'-gaccaccatgtcaaagtcc-3' (SEQ ID NO: 134)

[0211] After transfection of K562 cells, genomic DNA was extracted from cells transfected with repair template (RT) and either Cas9-HR4, Cas9-HR8, Cas9 (NT), or untransfected (Con). Standard Taq polymerase (Bioneer, T<sub>m</sub>=56,35 cycles) was used to amplify the fragments flanked by the 5' primers or the 3' primers.

[0212] **FIG. 19B** illustrates an agarose gel showing PCR products amplified by the 5' primers. Amplification products were detected for Cas9-HR4,8 and Cas9-NT, but were not detected in the un-transfected control.

[0213] **FIG. 19C** illustrates an agarose gel showing successful specific amplification by the 3' primers in Cas9-HR4, Cas9-HR8, and Cas9 only (NT).

[0214] **FIG. 19D** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product of Cas9-HR8 amplified by the 5' primers. Solid or unfilled bars bellowed the called base

denote identity of DNA (top left bar without fill is genomic, two bars in the middle with stripes are H2B ORF, and the bottom right bar without fill is mNeon), with the vertical grey dashed line of **FIG. 19A** showing the junction between genomic sequences included in the RT vs solely endogenous genomic sequences. Clear transition from H2B to mNeon sequences included in the RT to solely endogenous genomic sequence indicated successful integration of the transgene at the 5' end.

**[0215]** **FIG. 19E** illustrates absorbance of sequence trace from Sanger sequencing of the PCR product of Cas9-HR8 amplified by the 3' primers. Bars above the called bases denote identity of DNA (unfilled bar is mNeon and shaded bar is genomic). Clear transition from mNeon to genomic sequences included in the RT to solely endogenous genomic sequence indicated successful integration of the transgene at the 3' end.

**[0216]** **FIG. 19F** and **FIG. 19G** illustrate alignment of sequencing results the 5' (**FIG. 19F**) and 3' (**FIG. 19G**) PCR products from Cas9-HR4, Cas9-HR8, and NT with the reference sequence. Sequences were aligned using ClustalOmega.

#### **Example 12- Editing Adipose or Pre-Adipose Tissue to Increase Metabolic Flux**

**[0217]** Cells from either undifferentiated or mature adipose tissue are isolated from a patient and transfected with either plasmids encoding any one of the versions of Cas9-HR or purified RNPs. The chosen Cas9-HR(s) can be targeted to sites of the human genome which have been already been shown to be amenable to DNA insertion ("safe harbor sites") or any such novel site identified. Additionally, a repair template containing the cDNA for either Uncoupling Proteins (UCPs) 1, 2, 3 is transfected simultaneously. This transgene contains 5' Homology Arms (HAs) to the chosen integration site, either a ubiquitous or tissue specific enhancer complexed with a basal promoter, either with or without 5' UTR sequence, an ORF consisting of the aforementioned cDNA from either UCP 1, 2 or 3, with or without a 3' UTR sequence, a polyadenylation sequence, and a 3' HA to the chosen integration site. Integration and subsequent reintroduction of the Adipose Tissue expressing this transgene can increase basal metabolism, leading to overall weight loss and decrease in adipose lipid deposit size. Use of Cas9-HRs can lead to reduction in toxicity and increase the number of cells successfully integrated.

#### **Example 13-Editing Human Dermal Cells to Decrease Androgenic Alopecia**

**[0218]** Plasmids encoding Cas9-HR(s) or purified RNPs can be used to transfect either isolated cells or in-situ on the scalp to transfect transgenes expressing either full length or modified Sex Binding Hormone Globulin (SBHG), NRF 2, or SRD5A1,2 or 3. The chosen Cas9-HR(s) can be targeted to sites of the human genome which have been already been shown to be amenable to DNA insertion ("safe harbor sites") or any such novel site identified. These transgenes contain 5'

Homology Arms (HAs) to the chosen integration site, either a ubiquitous or tissue specific enhancer complexed with a basal promoter, either with or without 5'UTR sequence, an ORF consisting of the aforementioned cDNA from either SBHG, NFR2 or SRD5A1,2, or 3, with or without a 3' UTR sequence, a poly-adenylation sequence, and a 3' HA to the chosen integration site. Successful transfection of either in-situ cells or re-introduction of isolated dermal cells can delay or permanently halt hair-loss and result in hair regrowth.

**[0219]** 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 present disclosure. It should be understood that various alternatives to the embodiments described herein may be employed. It is intended that the following claims define the scope of the present disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.

**[0220] References:**

1. Oakes, B. L., Nadler, D. C. & Savage, D. F. Protein engineering of Cas9 for enhanced function. *Methods Enzymol.* 546, 491–511 (2014).
2. Cong, L. et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* 339, 819–823 (2013).
3. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308 (2013).
4. Jinek, M. et al. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* 337, 816–821 (2012).
5. Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* 475, 1955–1964 (2018).
6. Gehrke, J. M. et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* (2018). doi:10.1038/nbt.4199
7. Wang, L. et al. In Vivo Delivery Systems for Therapeutic Genome Editing. *Int. J. Mol. Sci.* 17, (2016).
8. Zhang, J.-P. et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol.* 18, 35 (2017).
9. Li, H. et al. Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv* 178905 (2017). doi:10.1101/178905
10. Canny, M. D. et al. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* 36, 95–102 (2018).

11. Liang, X., Potter, J., Kumar, S., Ravinder, N. & Chesnut, J. D. Enhanced CRISPR/Cas9-mediated precise genome editing by improved design and delivery of gRNA, Cas9 nuclease, and donor DNA. *J. Biotechnol.* 241, 136–146 (2017).
12. Ihry, R. J. et al. p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* 24, 939–946 (2018).
13. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.* 24, 927–930 (2018).
14. Biegging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* 14, 359–370 (2014).
15. Muller, P. A. J. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* 25, 304–17 (2014).
16. Canny, M. D. et al. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *36*, 95–102 (2018).
17. Ceccaldi, R., Rondinelli, B. & D’Andrea, A. D. Repair Pathway Choices and Consequences at the Double-Strand Break. *Trends Cell Biol.* 26, 52–64 (2016).
18. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.* 79, 181–211 (2010).
19. Shibata, A. et al. DNA double-strand break repair pathway choice is directed by distinct MRE11 nuclease activities. *Mol. Cell* 53, 7–18 (2014).
20. Tomimatsu, N. et al. Exo1 plays a major role in DNA end resection in humans and influences double-strand break repair and damage signaling decisions. *DNA Repair* 11, 441–8 (2012).
21. Bolderson, E. et al. Phosphorylation of Exo1 modulates homologous recombination repair of DNA double-strand breaks. *Nucleic Acids Res.* 38, 1821–1831 (2010).
22. Tomimatsu, N. et al. Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nat. Commun.* 5, 3561 (2014).
23. Tomimatsu, N. et al. DNA-damage-induced degradation of EXO1 exonuclease limits DNA end resection to ensure accurate DNA repair. *J. Biol. Chem.* 292, 10779–10790 (2017).
24. Paudyal, S. C., Li, S., Yan, H., Hunter, T. & You, Z. Dna2 initiates resection at clean DNA double-strand breaks. *Nucleic Acids Res.* 45, 11766–11781 (2017).
25. Tomimatsu, N. et al. DNA-damage-induced degradation of EXO1 exonuclease limits DNA end resection to ensure accurate DNA repair. *J. Biol. Chem.* 292, 10779–10790 (2017).
26. Chapman, J. R., Taylor, M. R. G. & Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell* 47, 497–510 (2012).

27. Hu, Z. et al. Ligase IV inhibitor SCR7 enhances gene editing directed by CRISPR–Cas9 and ssODN in human cancer cells. *Cell Biosci.* 8, 12 (2018).
28. Ren, B. et al. Improved Base Editor for Efficiently Inducing Genetic Variations in Rice with CRISPR/Cas9-Guided Hyperactive hAID Mutant. *Mol. Plant* 11, 623–626 (2018).
29. Li, X. et al. Base editing with a Cpf1–cytidine deaminase fusion. *Nat. Biotechnol.* 36, 324–327 (2018).
30. Jiang, W. et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* 28, 855–861 (2018).
31. La Russa, M. F. & Qi, L. S. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol. Cell. Biol.* 35, 3800–9 (2015).
32. Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* 18, 495–506 (2017).
33. Jia, P.-P. et al. Role of human DNA2 (hDNA2) as a potential target for cancer and other diseases: A systematic review. *DNA Repair* 59, 9–19 (2017).
34. Orans, J. et al. Structures of human exonuclease 1 DNA complexes suggest a unified mechanism for nuclease family. *Cell* 145, 212–23 (2011).
35. Xiaoying, C., Jennica, Z. & Wei-Chiang, S. Fusion Protein Linkers: Property, Design and Functionality. *Adv Drug Deliv Rev* 65, 1357–1369 (2014).
36. Präbst, K., Engelhardt, H., Ringgeler, S. & Hübner, H. Basic Colorimetric Proliferation Assays: MTT, WST, and Resazurin. in *Cell Viability Assays: Methods and Protocols* (eds. Gilbert, D. F. & Friedrich, O.) 1–17 (Springer New York, 2017). doi:10.1007/978-1-4939-6960-9\_1
37. Lieber, M., Todaro, G., Smith, B., Szakal, A. & Nelson-Rees, W. A continuous tumor-cell line from a human lung carcinoma with properties of type II alveolar epithelial cells. *Int. J. Cancer* 17, 62–70 (1976).
38. Klein, E. et al. Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *Int. J. Cancer* 18, 421–431 (1976).

The following paragraphs describe and represent aspects of the invention:

1. A method comprising introducing a first vector into a plurality of cells wherein said first vector encodes a fusion protein complex comprising a Cas9 nuclease fused to an exonuclease; wherein a viability of said plurality of cells comprising said vector is at least 1.5 times that of a second plurality of cells comprising a second vector encoding a Cas9 nuclease; wherein said second plurality of cells are K562 cells transfected with said second vector.
2. The method of paragraph 1, wherein said first vector encodes said fusion protein complex and a gRNA.
3. The method of paragraph 1, wherein said exonuclease is selected from the group consisting of MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean.
4. The method of paragraph 2, wherein a donor polynucleotide is introduced into said plurality of cells.
5. The method of paragraph 4, wherein an edit is made to an abnormal locus of a gene by said Cas9-fused to an exonuclease.
6. The method of paragraph 5, wherein said donor polynucleotide comprises an integration cassette further comprising a functional locus of said gene.
7. The method of paragraph 1, wherein said viability is measured by resazurin assay.
8. The method of paragraph 3, wherein said exonuclease is ExoI.
9. The method of paragraph 5, wherein said abnormal locus is an abnormal locus of a HBB gene.
10. The method of paragraph 9, wherein said donor polynucleotide encodes a functional locus of said HBB gene.
11. The method of paragraph 1, wherein said fusion protein complex encodes at least one nuclear localization signal (NLS).
12. The method of paragraph 1, wherein said first vector encoding said fusion protein complex has at least 80% sequence identity with any one of SEQ ID NO: 2-18.
13. The method of paragraph 1, wherein said first vector is delivered by electroporation.
14. The method of paragraph 4, wherein said donor polynucleotide comprises a mutated protospacer adjacent motif (PAM) sequence located at the immediate 3' end of a cleavage site, wherein said mutated PAM sequence comprises 5'-NCG-3' or 5'-NGC-3'.
15. The method of paragraph 14, wherein said fusion protein complex cannot cleave said mutated PAM sequence.
16. The method of paragraph 4, wherein said donor polynucleotide is single-stranded DNA.



17. The method of paragraph 4, wherein said donor polynucleotide is double-stranded DNA.
18. A polypeptide, comprising a first functional fragment, a second functional fragment comprising a Cas nuclease, and a linker peptide, wherein:
  - said first functional fragment is coupled to a first end of the linker peptide and the second functional fragment is coupled to a second end of said linker peptide; and
  - when a first complex comprising said polypeptide and a ribonucleic acid (RNA) molecule is administered to a first plurality of cells, a reduced toxicity is observed in said first plurality of cells compared to said toxicity observed in a second plurality of cells when a second complex comprising a Cas9 nuclease and said RNA molecule is administered to said second plurality of cells.
19. The polypeptide of paragraph 18, wherein said first functional fragment comprises an exonuclease wherein said exonuclease is selected from the group consisting of MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean.
20. The polypeptide of paragraph 19, wherein said RNA molecule is a guide RNA molecule.
21. The polypeptide of paragraph 19, wherein said exonuclease is a human Exo1 enzyme.
22. The polypeptide of paragraph 21 wherein said N-terminal of said human Exo1 enzyme is coupled to said C-terminal of said linker which is coupled to said C-terminal of said Cas nuclease.
23. The polypeptide of paragraph 21, wherein said human Exo1 enzyme comprises SEQ ID NO: 1.
24. The polypeptide of paragraph 21, wherein said human Exo1 enzyme comprises a fragment that has a 80% sequence identity of SEQ ID NO:1.
25. The polypeptide of paragraph 21, wherein said human Exo1 enzyme comprises a fragment that has a 90% sequence identity of SEQ ID NO:1.
26. The polypeptide of paragraph 21, wherein said human Exo1 enzyme comprises a fragment that has a 95% sequence identity of SEQ ID NO:1.
27. The polypeptide of paragraph 18, wherein said second functional fragment comprises a Cas9 enzyme.
28. The polypeptide of paragraph 27, wherein said Cas9 enzyme comprises a N-terminal nuclear localizing sequence (NLS) and a C-terminal NLS.
29. The polypeptide of paragraph 27, wherein said Cas9 enzyme comprises a N-terminal nuclear localizing sequence (NLS).

30. The polypeptide of paragraph 27, wherein said Cas9 enzyme comprises a C-terminal nuclear localizing sequence (NLS).
31. The polypeptide of any of paragraphs 18-30, wherein said linker peptide is selected from a group consisting of FL2X, SLA2X, AP5X, FL1X, SLA1X.
32. The polypeptide of paragraph 31, wherein said linker peptide is SLA2X.
33. The polypeptide of any of paragraph 31, wherein said linker peptide comprises 5 to 200 amino acids.
34. The polypeptide of paragraph 18, wherein said reduced toxicity is quantified by measuring resorufin accumulation.
35. The polypeptide of paragraph 34, wherein after administration of said first complex, said first plurality of cells have at least two times a number of viable cells when compared to said second plurality of cells after administration of said second complex, wherein the number of viable cells is quantified by a resorufin assay.
36. The polypeptide of paragraph 34, wherein after administration of said first complex, said first plurality of cells have at least two times said amount of HDR edited cells when compared to said second plurality of cells after administration of said second complex as quantified by a cellular HDR assay.
37. The polypeptide of paragraph 33, wherein said cellular HDR assay comprises IHC, qPCR or deep sequencing.
38. A polynucleotide encoding said polypeptide of any of paragraphs 17-35 and said RNA molecule.
39. The polynucleotide of paragraph 38, wherein said first end of said linker peptide is a 3' end and said second end of said linker peptide is a 5' end.
40. The polynucleotide of paragraph 38, wherein said first end of said linker peptide is a 5' end and said second end of said linker peptide is a 3' end.
41. The polynucleotide of paragraph 38, wherein said RNA molecule is a guide RNA (gRNA).
42. The polynucleotide of paragraph 38, further comprising a homology directed repair (HDR) template.
43. The polynucleotide of paragraph 38, wherein said gRNA is selected from sequences listed in Table 2.
44. The polynucleotide of paragraph 38, wherein said HDR template is a single-strand DNA.
45. The polynucleotide of paragraph 38, wherein said HDR template is a double-strand DNA.

46. The polynucleotide of paragraph 38, wherein said polynucleotide is formulated in a liposome.
47. The polynucleotide of paragraph 46, wherein said liposome comprises a polyethylene glycol (PEG), a cell-penetrating peptide, a ligand, an aptamer, an antibody, or a combination thereof.
48. A vector comprising a nucleotide sequence of paragraph 38.
49. The vector of paragraph 48, wherein said vector comprises a promoter.
50. The vector of paragraph 49, wherein said promoter is a CMV or a CAG promoter.
51. The vector of any of paragraphs 48-50, wherein said vector is selected from a group consisting of retroviral vectors, adenoviral vectors, lentiviral vectors, herpesvirus vectors, and adeno-associated viral vectors.
52. The vector of paragraph 51, wherein said vector is an adeno-associated viral vector.
53. A virus-like particle (VLP) comprising said vector of any of paragraphs 48-50.
54. A kit comprising a polypeptide of any of paragraphs 18-41 formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.
55. A kit comprising the polynucleotide of paragraph 38 formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.
56. A kit comprising a vector of any of paragraphs 48-52 formulated in a compatible pharmaceutical excipient, an insert with administering instructions, reagents.
57. A method for inducing homologous recombination of DNA in a cell, comprising contacting said DNA with a polypeptide of any of paragraphs 18-35.
58. A method for inducing HDR in a cell in vitro or ex vivo, comprising delivering a polynucleotide of paragraph 38 into a cell.
59. The method of any of paragraphs 57-58, wherein said cell is a human cell, a non-human mammalian cell, a stem cell, a non-mammalian cell, an invertebrate cell, a plant cell, or a single-eukaryotic organism.
60. A method, comprising:
  - contacting a first plurality of cells with a polynucleotide of paragraph 18 and a second plurality of cells with a second polynucleotide encoding a wild-type Cas9 enzyme; and
  - inducing a site-specific cleavage at an intended locus followed by HDR in said first plurality of cells and said second plurality of cells; and
  - recovering at least 30-90% more cells in said first plurality of cells compared to said second plurality of cells.

61. A method of paragraph 60, further comprising measuring cell viability by measuring an amount of resorufin produced in said first plurality of cells and said second plurality of cells.
62. The method of paragraph 61, wherein said first plurality of cells have 2-5 times an amount of viable cells as quantified by a resorufin assay when compared to said second plurality of cells.
63. The method of any of paragraphs 60-62, wherein said first plurality of cells and said second plurality of cells comprise a human cell, a non-human mammalian cell, a stem cell, a non-mammalian cell, an invertebrate cell, a plant cell, or a single-eukaryotic organism.
64. The method of paragraph 63, wherein said human cell is a T cell, a B cell, a dendritic cell, a natural killer cell, a macrophage, a neutrophil, an eosinophil, a basophil, a mast cell, a hematopoietic progenitor cell, a hematopoietic stem cell (HSC), a red blood cell, a blood stem cell, an endoderm stem cell, an endoderm progenitor cell, an endoderm precursor cell, a differentiated endoderm cell, a mesenchymal stem cell (MSC), a mesenchymal progenitor cell, a mesenchymal precursor cell, or a differentiated mesenchymal cell.
65. The method of paragraph 64, wherein said differentiated endoderm cell is a hepatocytes progenitor cell, a pancreatic progenitor cell, a lung progenitor cell, or a tracheae progenitor cell.
66. The method of paragraph 64, wherein said differentiated mesenchymal cell is a bone cell, a cartilage cell, a muscle cell, an adipose cell, a stromal cell, a fibroblast, or a dermal cell.
67. A method for treating a single gene disorder in a subject, comprising:  
culturing a plurality of primary cells obtained from said subject;  
administering a polynucleotide of paragraph 42 to said plurality of primary cells, wherein said gRNA is configured to recognize a locus of said gene that causes said disorder and said HDR template is configured to provide a functioning sequence of said gene; and  
inducing a site-specific cleavage at said locus followed by HDR, wherein said functioning sequence of said gene is inserted at said locus.
68. The method of paragraph 67, further comprising:  
selecting primary cells in which said functioning sequence of said gene is inserted at said locus; and  
reintroducing said selected primary cells back into said subject.
69. The method of either paragraph 67 or paragraph 68, wherein said subject is a mammal.
70. The method of paragraph 69, wherein said mammal is a human.
71. The method of paragraph 67, wherein said plurality of primary cells are selected from a group comprising T cells, B cells, dendritic cells, natural killer cells, natural killer cells,

macrophages, neutrophils, eosinophils, basophils, mast cells, hematopoietic progenitor cells, hematopoietic stem cells (HSCs), red blood cells, blood stem cells, endoderm stem cells, endoderm progenitor cells, endoderm precursor cells, differentiated endoderm cells, mesenchymal stem cells (MSCs), mesenchymal progenitor cells, mesenchymal precursor cells, differentiated mesenchymal cells, hepatocytes progenitor cells, pancreatic progenitor cells, lung progenitor cells, tracheae progenitor cells, bone cells, cartilage cells, muscle cells, adipose cells, stromal cells, fibroblasts, and dermal cells.

72. The method of paragraph 67, wherein said gene that causes said single gene disorder is selected from Table 3.

73. A method for treating sickle cell anemia caused by an abnormal HBB gene in a subject, comprising:

culturing a plurality of primary cells obtained from said subject;

administering a polynucleotide of paragraph 42 to said plurality of primary cells, wherein said gRNA is configured to recognize a locus of said HBB gene that causes said disorder and said HDR template is configured to provide a functioning sequence of said HBB gene; and

inducing a site-specific cleavage at said locus followed by HDR, wherein said functioning sequence of said HBB gene is inserted at said locus.

74. The method of paragraph 73, further comprising:

selecting primary cells in which said functioning sequence of said HBB gene is inserted at said locus; and

reintroducing said selected primary cells back into said subject.

75. The method of either paragraph 73 or paragraph 74, wherein said subject is a mammal.

76. The method of paragraph 75, wherein said mammal is a human.

77. The method of either paragraph 73 or paragraph 74, wherein said primary cell is a hematopoietic stem cell.

78. The method of paragraph 73 wherein said primary cell is a CD34+ hematopoietic stem cell.

79. The method of paragraph 74 wherein said primary cell is a CD34+ hematopoietic stem cell.

80. The vector of paragraph 48 wherein said vector is plasmid PX330.

81. The method of paragraph 58 wherein said cell is CD34+ hematopoietic stem cell.

82. A method for treating sickle cell anemia caused by an abnormal HBB gene in a subject, comprising:

- culturing a plurality of primary cells obtained from said subject;  
administering a polynucleotide of paragraph 22 to said plurality of primary cells,  
wherein said gRNA is configured to recognize a locus of said HBB gene that causes the  
disorder and said HDR template is configured to provide a functioning sequence of said  
HBB gene; and  
inducing a site-specific cleavage at the locus followed by HDR, wherein the  
functioning sequence of said HBB gene is inserted at said locus.
83. The method of paragraph 82, further comprising:  
selecting primary cells in which said functioning sequence of the HBB gene is  
inserted at said locus; and  
reintroducing said selected primary cells back into said subject.
84. The method of either paragraph 82 or paragraph 83, wherein said subject is a mammal.
85. The method of paragraph 84, wherein said mammal is a human.
86. The method of paragraph 82 or paragraph 83 wherein said primary cell is CD34+  
hematopoietic stem cell.
87. A method, comprising:  
contacting a first of plurality of cells with a first complex comprising said  
polynucleotide of paragraph 18 and a RNA molecule;  
inducing a site-specific cleavage followed by HDR in the first plurality of cells,  
wherein a percentage of cells of said first plurality of cells edited by HDR quantified by a  
cellular HDR assay is at least two times higher compared to a percentage of cells of a  
second plurality of cells contacted with a second complex comprising a polynucleotide  
encoding a wild-type Cas9 enzyme and said RNA molecule.
88. The method of paragraph 84, wherein said cellular HDR assay comprises IHC.
89. The method of paragraph 84, wherein said cellular HDR assay comprises qPCR.
90. The method of paragraph 84, wherein said cellular HDR assay comprises nucleic acid  
sequencing.

**CLAIMS:**

1. A composition comprising:
  - i) a polypeptide, comprising a first functional fragment, a second functional fragment comprising a Cas9 nuclease, and a linker peptide, wherein:  
said first functional fragment is coupled to a first end of the linker peptide and the second functional fragment is coupled to a second end of said linker peptide; and  
said first functional fragment comprises an exonuclease wherein said exonuclease is selected from the group consisting of MRE11, EXO1, EXOIII, EXOVII, EXOT, DNA2, CtIP, TREX1, TREX2, Apollo, RecE, RecJ, T5, Lexo, RecBCD, and Mungbean;
  - ii) a guide RNA (gRNA); and
  - iii) a homology directed repair (HDR) template.
2. The composition of claim 1, wherein said exonuclease is a human Exol enzyme.
3. The composition of claim 2, wherein an N-terminal of said human Exol enzyme is coupled to a C-terminal of said linker which is coupled to said C-terminal of said Cas nuclease.
4. The composition of claim 2, wherein said human Exol enzyme comprises at least 80% sequence identity to SEQ ID NO: 1.
5. The composition of claim 4, wherein said Cas9 nuclease comprises an N-terminal nuclear localizing sequence (NLS).
6. The composition of claims 4 or 5, wherein said Cas9 nuclease comprises a C-terminal nuclear localizing sequence (NLS).
7. The composition of any one of claims 1 to 6, wherein said linker peptide is selected from a group consisting of FL2X, SLA2X, AP5X, FL1X, SLA1X.
8. The composition of claim 7, wherein said linker peptide is SLA2X.
9. The composition of claim 1, wherein said linker peptide comprises 5 to 200 amino acids.
10. A polynucleotide encoding the composition of any one of claims 1 to 9.
11. A vector comprising the polynucleotide of claim 10.

28 09 23