



US012612633B2

(12) **United States Patent**
Kim et al.

(10) **Patent No.:** **US 12,612,633 B2**
(45) **Date of Patent:** ***Apr. 28, 2026**

(54) **COMPOSITIONS FOR INDUCING MODIFICATIONS OF TARGET ENDOGENOUS NUCLEIC ACID SEQUENCES IN NUCLEUSES OF EUKARYOTIC CELLS**

(71) Applicant: **TOOLGEN INCORPORATED**, Seoul (KR)

(72) Inventors: **Jin-Soo Kim**, Seoul (KR); **Seung Woo Cho**, Seoul (KR); **Sojung Kim**, Seoul (KR)

(73) Assignee: **ToolGen Incorporated**, Seoul (KR)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **19/002,832**

(22) Filed: **Dec. 27, 2024**

(65) **Prior Publication Data**

US 2025/0122509 A1 Apr. 17, 2025

Related U.S. Application Data

(63) Continuation of application No. 18/932,745, filed on Oct. 31, 2024, which is a continuation of application No. 17/004,338, filed on Aug. 27, 2020, which is a continuation of application No. 14/685,568, filed on Apr. 13, 2015, now Pat. No. 10,851,380, which is a continuation of application No. PCT/KR2013/009488, filed on Oct. 23, 2013.

(60) Provisional application No. 61/837,481, filed on Jun. 20, 2013, provisional application No. 61/803,599, filed on Mar. 20, 2013, provisional application No. 61/717,324, filed on Oct. 23, 2012.

(51) **Int. Cl.**

C12N 9/22 (2006.01)
C12N 9/16 (2006.01)
C12N 15/10 (2006.01)
C12N 15/11 (2006.01)
C12N 15/52 (2006.01)
C12N 15/63 (2006.01)
C12N 15/82 (2006.01)
C12N 15/85 (2006.01)
C12N 15/90 (2006.01)

(52) **U.S. Cl.**

CPC **C12N 15/52** (2013.01); **C12N 9/16** (2013.01); **C12N 9/22** (2013.01); **C12N 15/102** (2013.01); **C12N 15/111** (2013.01); **C12N 15/63** (2013.01); **C12N 15/8216** (2013.01); **C12N 15/85** (2013.01); **C12N 15/907** (2013.01); **C12Y 301/21** (2013.01); **C12N 2310/10** (2013.01); **C12N 2310/20** (2017.05); **C12N 2310/531** (2013.01)

(58) **Field of Classification Search**

CPC **C12N 2310/20**; **C12N 9/22**
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,766,900 A 6/1998 Shillito et al.
5,767,367 A 6/1998 Dudits et al.
5,925,517 A 7/1999 Tyagi et al.
6,040,295 A 3/2000 Rolland et al.
8,697,359 B1 4/2014 Zhang
8,771,495 B2 7/2014 Paneccasio et al.
8,883,233 B2 11/2014 Gilllessen
8,889,559 B2 11/2014 Trapp et al.

(Continued)

FOREIGN PATENT DOCUMENTS

AU 733057 B2 5/2001
AU 2011203213 A1 8/2011

(Continued)

OTHER PUBLICATIONS

Locarnini and Zoulim, Molecular genetics of HBV infection. *Anti-viral Therapy* (2010), 15 Suppl 3:3-14 (Year: 2010).*

Hudson and Ortlund, The structure, function and evolution of proteins that bind DNA and RNA. *Nature Reviews Molecular Cell Biology* (2014), 15: 749-760 (Year: 2014).*

Sternberg et al., Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* (2012), 18: 661-672 (Year: 2012).*

Hwang et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature Biotechnology* (2013), 31: 227-229 (Year: 2013).*

Jinek et al., (2012) "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science*, vol. 337, No. 6096, pp. 816-821.

(Continued)

Primary Examiner — Catherine Konopka

(74) *Attorney, Agent, or Firm* — Gemini Law LLP

(57) **ABSTRACT**

The present disclosure relates to targeted genome editing in eukaryotic cells or organisms. More particularly, the present disclosure provides for compositions and methods that may induce modifications in target endogenous nucleic acid sequences in nucleuses of eukaryotic cells. For example, disclosed herein are methods of modifying a target endogenous nucleic acid sequence in a nucleus of a eukaryotic cell in vitro or ex vivo, comprising: preparing a composition comprising a recombinant Cas9 protein, a single guide RNA, and a Cas9/RNA complex formed by at least a part of said recombinant Cas9 protein and single guide RNA, and introducing a Cas9/RNA complex into the eukaryotic cell in vitro or ex vivo, wherein in the composition, the single guide RNA is at least two-fold molar excess over the recombinant Cas9 protein, wherein the Cas9/RNA complex is a combination of the recombinant Cas9 protein and the single guide RNA, and wherein the Cas9/RNA complex is complexed prior to being introduced into the eukaryotic cell; wherein the single guide RNA is in vitro transcribed RNA or synthetic RNA, wherein the combination of the recombinant Cas9 protein and the single guide RNA produces a modification of the target endogenous nucleic acid sequence in the nucleus of the eukaryotic cell.

18 Claims, 61 Drawing Sheets

Specification includes a Sequence Listing.

(56)

References Cited

U.S. PATENT DOCUMENTS

8,993,233 B2 3/2015 Zhang et al.
 8,999,641 B2 4/2015 Zhang et al.
 9,023,649 B2 5/2015 Mali et al.
 9,260,723 B2 2/2016 Mali et al.
 9,493,779 B2 11/2016 Ainley et al.
 9,637,739 B2 5/2017 Siksnyš et al.
 10,851,380 B2 12/2020 Kim et al.
 2005/0220796 A1 10/2005 Dynan et al.
 2009/0111119 A1 4/2009 Doyon et al.
 2010/0001189 A1 1/2010 Federici
 2010/0076057 A1 3/2010 Sontheimer et al.
 2011/0223638 A1 9/2011 Wiedenheft et al.
 2012/0058102 A1 3/2012 Wilson et al.
 2014/0068797 A1* 3/2014 Doudna C12N 15/907
 435/375

2014/0090113 A1 3/2014 Cogan et al.
 2014/0093913 A1 4/2014 Cost et al.
 2014/0179770 A1 6/2014 Zhang et al.
 2014/0186843 A1 7/2014 Zhang et al.
 2014/0186919 A1 7/2014 Zhang et al.
 2014/0189896 A1 7/2014 Zhang et al.
 2014/0242664 A1 8/2014 Zhang et al.
 2014/0273230 A1 9/2014 Chen et al.
 2014/0310830 A1 10/2014 Zhang et al.
 2014/0342456 A1 11/2014 Mali et al.
 2014/0349405 A1 11/2014 Sontheimer et al.
 2014/0357530 A1 12/2014 Zhang et al.
 2015/0020223 A1 1/2015 Zhang et al.
 2015/0031134 A1 1/2015 Zhang et al.
 2015/0050699 A1 2/2015 Siksnyš et al.
 2015/0067921 A1 3/2015 Cogan et al.
 2015/0067922 A1 3/2015 Yang et al.
 2015/0079681 A1 3/2015 Zhang
 2015/0166615 A1 6/2015 Xia et al.
 2015/0203872 A1 7/2015 Zhang
 2015/0225734 A1 8/2015 Voytas et al.
 2015/0232882 A1 8/2015 Zhang et al.
 2015/0247150 A1 9/2015 Zhang et al.
 2015/0291965 A1 10/2015 Zhang et al.
 2015/0344912 A1 12/2015 Kim et al.
 2015/0351340 A1 12/2015 Bundock et al.
 2015/0356239 A1 12/2015 Zhang et al.
 2016/0017366 A1 1/2016 Chen et al.
 2016/0032274 A1 2/2016 Church et al.
 2016/0046961 A1 2/2016 Jinek et al.
 2016/0153004 A1 6/2016 Zhang et al.
 2016/0153005 A1 6/2016 Zhang et al.
 2016/0153006 A1 6/2016 Zhang et al.
 2016/0168594 A1 6/2016 Zhang et al.
 2016/0175462 A1 6/2016 Zhang et al.
 2016/0186213 A1 6/2016 Zhang et al.
 2021/0047648 A1 2/2021 Kim et al.

FOREIGN PATENT DOCUMENTS

JP 2005006578 A 1/2005
 NZ 228948 A 6/1991
 WO WO1998010084 A1 3/1998
 WO WO2000055378 A1 9/2000
 WO WO2002067966 A1 9/2002
 WO WO2002067996 A2 9/2002
 WO WO2005054494 A2 6/2005
 WO WO2007024029 A1 3/2007
 WO WO2007025195 A1 3/2007
 WO WO2008108989 A2 9/2008
 WO WO2009042164 A1 4/2009
 WO WO2010001189 A1 1/2010
 WO WO2010052341 A1 5/2010
 WO WO2010076939 A1 7/2010
 WO WO2011007193 A1 1/2011
 WO WO2011056186 A1 5/2011
 WO WO2011130346 A1 10/2011
 WO WO2011146121 A1 11/2011
 WO WO2012012738 A1 1/2012

WO WO2012138939 A1 10/2012
 WO WO2013098244 A1 7/2013
 WO WO2013188522 A2 12/2013
 WO WO2014018423 A2 1/2014
 WO WO2014022702 A2 2/2014
 WO WO2014065596 A1 5/2014
 WO WO-2014089290 A1* 6/2014 A61K 38/00
 WO WO2014093635 A1 6/2014
 WO WO2014093694 A1 6/2014
 WO WO2014093709 A1 6/2014
 WO WO2014093718 A1 6/2014
 WO WO2014099750 A2 6/2014
 WO WO2014144155 A1 9/2014
 WO WO2014197568 A2 12/2014
 WO WO2014204725 A1 12/2014
 WO WO2015026883 A1 2/2015

OTHER PUBLICATIONS

Pennisi, E., The CRISPR Craze. *Science*. 2013, 341(6148):833-836.
 Perez-Pinera, P., et al., RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*. 2013, 10(10):973-976.
 Perez-Pinera, P., et al., Advances in Targeted Genome Editing. *Current Opinion in Chemical Biology*. 2012, 16(3-4):268-277. (1 page).
 Perez-Rodriguez, R., Elucidating a novel mechanism of DNA silencing caused by envelope stress in *Escherichia coli*. Dissertation and Abstract first available to the public on Jun. 17, 2010 (156 pages).
 Pillich, H., et al., Activation of the unfolded protein response by *Listeria monocytogenes*. *Cellular Microbiology*, 2012, 14(6):949-964.
 Primo, M. N., et al., Lentiviral vectors for cutaneous RNA managing. *Experimental Dermatology*. 2012, 21(3):162-170.
 Program and Conference Logistics provided to the attendees of the CRISPR 2012: 5th Annual CRISPR Research Meeting held at the University of California, Berkeley, CA (Jun. 2012). (5 pages).
 PShooter™ Vector user guide, Invitrogen by Life Technologies, revision date Mar. 29, 2012 (36 pages).
 Qi, L., et al., RNA Processing Enables Predictable Programming of Gene Expression. *Nat Biotechnol*. 2012, 30(10):1002-1006 and Supplementary Materials.
 Qi, L. S., et al., Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*. 2013, 152(5):1173-1183.
 Ramirez, C. L., et al., Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acid Research*. 2012, 40(12):5560-5568. (9 pages).
 Ran, F. A., et al., Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*. 2013, 154(6):1380-1389.
 Ran, F. A., et al., Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc*. 2013, 8:2281-2308.
 Reyon, D., et al., FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol*. 2012, 30(5):460-465. (8 pages).
 Richter, H., et al., Exploiting CRISPR/Cas: Interference Mechanisms and Applications. *International Journal of Molecular Sciences*. 2013, 14(7):14518-14531. (14 pages).
 Rios, X., et al., Stable Gene Targeting in Human Cells Using Single-Strand Oligonucleotides with Modified Bases. *PLoS One*. 2012, 7(5):e36697. (14 pages).
 Sakurai, K., et al., Silencing of Gene Expression in Cultured Cells Using Small Interfering RNAs. *Curr. Protoc. Cell Biol*. 2010;47:27.1.1-27.1.28. (28 pages).
 Sander, J. D., et al., Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nature Biotechnology*. 2011, 29(8):697-698 with Supplementary Information. (27 pages).
 Sanders, R., 2013, "Cheap and easy technique to snip DNA could revolutionize gene therapy;" UC Berkeley News Center, Jan. 7, 2013, retrieved from the internet at <<http://newscenter.berkeley.edu/2013/01/07/cheap-and-easy-technique-to-snip-dna-could-revolutionize-gene-therapy/>> pp. 1-3 (3 pages).

(56)

References Cited

OTHER PUBLICATIONS

- Sanjana, N. E., et al., A Transcription Activator-Like Effector (TALE) Toolbox for Genome Engineering. *Nat Protoc.* 2012, 7(1):171-192. Supplementary Materials (28 pages).
- Sapranaukas, R., et al., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Research.* 2011, 39(21):9275-9282 and Supplementary Materials. (15 pages).
- Schmidt, A., et al., Sensing of viral nucleic acids by RIG-I: from translocation to translation. *European Journal of Cell Biology.* 2012, 91(1):78-85.
- Second Declaration of Dr. Boch dated Apr. 26, 2019 as filed in Opposition to European Patent No. 2825654 (20 pages).
- Segal, D., J., Genome Engineering: Bacteria herald a new era of gene editing. *eLife.* 2013, 2:e00563.
- Shalem, O., et al., Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science.* 2013, 343(6166):84-87.
- Shan, Q., et al., Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol.* 2013, 31(8):686-688 and Supplemental Materials.
- Shen, B., et al., Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. *Cell Res.* 2013, 23:720-723.
- Singapore Written Opinion dated Mar. 18, 2016 in relation to Singapore Patent Application No. 11201503059X.
- Sinkunas, T., et al., Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO Journal.* 2011, 30:1335-1342. (8 pages).
- Strecker, J., et al., Engineering of CRISPR-Cas12b for human genome editing. *Nat Commun.* 2019, 10(1):212.
- Sternberg, S. H., et al., DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Biophysical Journal.* 2014, 106(2):62-67. (17 pages).
- Sternberg, S. H., et al., Mechanism of substrate selection by a highly specific CRISPR endonuclease. *RNA* 2012, 18:661-672. (12 pages).
- Streptococcus*—bacterial genus—Microbiology Dictionary, Hardy Diagnostics, available at https://catalog.hardydiagnostics.com/cp_prod/Content/hugo/Strepococcus.htm (last accessed Oct. 25, 2020), (4 pages).
- Sun, C., et al., Functional reconstruction of human eukaryotic translation initiation factor 3 (eIF3). *PNAS.* 2011, 108(51):20473-20478. (6 pages).
- Sung, Y. H., et al., Mouse genetics: Catalogue and scissors. *BMB Rep.* 2012, 45(12):686-692.
- Sung, Y. H., et al., Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. *Genome Res.* 2014, 24:125-131. (7 pages).
- Svitashev, S., et al., Genome editing in maize directed by CRISPR-Cas9 ribonucleoprotein complexes. *Nature Communications.* 2016, 7:13274. (7 pages).
- Szczepankowska, A., Role of CRISPR/cas system in the development of bacteriophage resistance. *Advances in Virus Research.* 2012, 82:289-338. (50 pages).
- Terns, M. P. and Terns, R. M., CRISPR-Based Adaptive Immune Systems. *Current Opinion in Microbiology.* 2011, 14(3):321-327. (13 pages).
- Tesson, L., et al., Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol.* 2011, 29(8):695-696 and Supplementary Materials. (15 pages).
- Thurman, R. E., et al., The accessible chromatin landscape of the human genome. *Nature.* 2012, 489:75-82. (8 pages).
- Tzur, Y. B., et al., Heritable Custom Genomic Modifications in *Caenorhabditis elegans* via a CRISPR-Cas9 System. *Genetics.* 2013, 195(3):1181-1185. (14 pages).
- Úrnov, F. D., et al., Genome editing with engineered zinc finger nucleases. *Nat Rev Genet.* 2010, 11(9):636-646.
- Van Der Oost, J., New Tool for Genome Surgery. *Science.* 2013, 339(6121):768-770.
- Vasu, K., et al., Endonuclease Active Site Plasticity Allows DNA Cleavage with Diverse Alkaline Earth and Transition Metal Ions. *ACS Chemical Biology.* 2011, 6(9):934-942. (9 pages).
- Villion, M., and Moineau, S., The double-edged sword of CRISPR-Cas9 systems. *Cell Res.* 2013, 23:15-17. (4 pages).
- Walsh, R. M., and Hochedlinger, K., A variant CRISPR-Cas9 system adds versatility to genome engineering. *PNAS.* 2013, 110(39):15514-15515.
- Wang, H., et al., One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell.* 2013, 153(4):910-918. (17 pages).
- Wang, J., et al., Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Res.* 2012, 22(7):1316-1326. (11 pages).
- Wang, T., et al., Genetic Screens in Human Cells Using the CRISPR/Cas9 System. *Science.* 2013, 343(6166):80-84.
- Handel, E.-M., et al., Versatile and Efficient Genome Editing in Human Cells by Combining Zinc-Finger Nucleases with Adeno-Associated Viral Vectors. *Human Gene Therapy.* 2011, 23(3):321-329. (9 pages).
- Haurwitz, R. E., et al., Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science.* 2010, 329(5997):1355-1358. (4 pages).
- PCT International Search Report dated Jan. 27, 2014, in relation to International Application No. PCT/KR2013/009488 filed Oct. 23, 2013, pp. 1-5.
- Hocine, S., et al., RNA processing and export. *Cold Spring Harb Perspect Biol.* 2010, 2:a000752.
- Holt, N., et al., Zinc finger nuclease-mediated CCR5 knockout hematopoietic stem cell transplantation controls HIV-1 in vivo. *Nat Biotechnol.* 2010, 28(8):839-847. (26 pages).
- Horvath, P., and Barrangou, R., RNA-guided genome editing a la carte. *Cell Res.* 2013, 23:733-734.
- Horvath, P., and Barrangou, R., CRISPR/Cas, the immune system of bacteria and archaea. *Science.* 2010, 327(5962):167-170. (4 pages).
- Hoshijima, K., et al., Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Developmental Cell.* 2019, 51(5):645-657.e4. (35 pages).
- Hruscha, A., et al., Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development.* 2013, 140(24):4982-4987. Supplementary Information (12 pages).
- Hsu, P. D., et al., DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol.* 2013, 31(9):827-832.
- Hu, P., et al., Comparison of various nuclear localization signal-fused Cas9 proteins and Cas9 mRNA for genome editing in zebrafish. *G3 Genes/Genomes/Genetics.* 2018, 8(3):823-831. (9 pages).
- Huang, P., et al., Heritable gene targeting in zebrafish using customized TALENs. *Nature Biotechnology.* 2011, 29(8):699-700. (2 pages).
- Hwang, W. Y., et al., Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol.* 2013, 31:227-229 and Supplementary Materials.
- Hwang, W. Y., et al., Heritable and Precise Zebrafish Genome Editing Using a CRISPR-Cas System. *PLoS One.* 2013, 8(7):e68708 and Supplementary Materials. (14 pages).
- Jani, B. and Fuchs R., In Vitro Transcription and Capping of *Gaussia* Luciferase mRNA Followed by HeLa Cell Transfection. *J. Vis. Exp.* 2012, 61:3702. (9 pages).
- Japanese Office Action dated Nov. 24, 2015 for Japanese Patent Application No. 2015-538033, pp. 1-10 (with English Translation).
- Jensen, N. M., et al., An update on targeted gene repair in mammalian cells: methods and mechanisms. *J. Biomed. Sci.* 2011;18:10. (14 pages).
- Jeyarajan, S., et al., 2010, Plasmid DNA delivery into MDA-MB-453 cells mediated by recombinant Her-NLS fusion protein. *International Journal of Nanomedicine.* 2010, 5:725-733. (9 pages).
- Jiang, F., et al., 2011, Structural basis of RNA recognition and activation by innate immune receptor RIG-I. *Nature.* 2011, 479:423-429.
- Jiang, W., et al., Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis*, tobacco, sorghum and rice. *Nucleic Acids Research.* 2013, 41(20):e188.

(56)

References Cited

OTHER PUBLICATIONS

- Jiang, W., et al., RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* 2013, 31:233-239.
- Jiang, W., et al., CRISPR-assisted editing of bacterial genomes. *Nat Biotechnol.* 2013, 31(3):233-239.
- Jinek, M., et al., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* 2012, 337(6096):816-821 and Supplementary Materials. (7 pages).
- Jinek, M., et al., Methods and Compositions for RNA-Directed Site-Specific DNA Modification. U.S. Appl. No. 61/652,086, filed May 25, 2012.
- Jinek, M., et al., RNA-programmed genome editing in human cells. *eLife.* 2013, 2:e00471.
- Jore, M. M., et al., Structural basis for CRISPR RNA-guided DNA recognition by Casade. *Nat. Struct. Mol. Biol.* 2011, 18(5):529-536. (9 pages).
- Joung, J. K. and Sander J. D., TALENs: A widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* 2013, 14:49-55. (7 pages).
- Kang, H. C. and Bae, Y. H., Co-delivery of small interfering RNA and plasmid DNA using a polymeric vector incorporating endosomolytic oligomeric sulfonamide. *Biomaterials.* 2011, 32(21):4914-4924. (11 pages).
- Karginov, F. V. and Hannon, G. J., The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Molecular Cell.* 2010, 37(1):7-19. (13 pages).
- Karvelis, T., et al., crRNA and tracrRNA guide Cas9-mediated DNA interference in *Streptococcus thermophilus*. *RNA Biology,* 2013, 10(5):841-851.
- Katic, I. and Grobhans, H., Targeted Heritable Mutation and Gene Conversion by Cas9-CRISPR in *Caenorhabditis elegans*. *Genetics.* 2013, 195(3):1173-1176. (4 pages).
- Kim, E., et al., Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res.* 2012, 22:1327-1333.
- Kim, H., et al., Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat Methods.* 2011, 8:941-943.
- Kim, S., et al., Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* 2014, 24:1012-1019. (9 pages).
- Kim, S., et al., 2018, CRISPR RNAs trigger innate immune responses in human cells. *Genome Res.* 2018, 28:367-373. (8 pages).
- Kim, T. K., and Eberwine, J. H., Mammalian cell transfection: the present and the future. *Anal Bioanal Chem.* 2010, 397:3173-3178. (6 pages).
- Kolomeisky, A. B., Physics of protein-DNA interactions: mechanisms of facilitated target search. *Phys. Chem. Chem. Phys.* 2011, 13:2088-2095. (8 pages).
- Kondo, S., et al., Highly Improved Gene Targeting by Germline-Specific Cas9 Expression in *Drosophila*. *Genetics.* 2013, 195(3):715-721.
- Koo, T., et al., Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Molecules and Cells.* 2015, 38(6):475-481. (7 pages).
- Kouranova, E., et al., CRISPRs for optimal targeting: delivery of CRISPR components as DNA, RNA, and protein into cultured cells and single-cell embryos. *Human Gene Therapy.* 2016, 27(6):464-475. (12 pages).
- Lambowitz, A. M. and Zimmerly, S., Group II Introns: Mobile Ribozymes that Invade DNA. *Cold Spring Harb Perspect Biol.* 2011, 3:a003616. (19 pages).
- Lane, J., et al., Targeting RHoC by way of ribozyme transgene in human breast cancer cells and its impact on cancer invasion. *World J Oncol.* 2010, 1(1):7-13. (7 pages) Lane, J., et al., Targeting RHoC by way of ribozyme transgene in human breast cancer cells and its impact on cancer invasion. *World J Oncol.* 2010, 1(1):7-13. (7 pages).
- Larson, D. R., et al., Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science.* 2011, 332(6028):475-478.
- Larsen, H. O., et al., Nonviral transfection of leukemic primary cells and cells lines by siRNA—a direct comparison between Nucleofection and Accell delivery. *Experimental Hematology.* 2011, 39:1081-1089. (9 pages).
- Larson, M. H. et al., CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc.* 2013, 8:2180-2196.
- Ledford, H., Targeted gene editing enters clinic. *Nature.* 2011, 471:16. (1 page).
- Lee, H. Y., et al., Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res.* 2010, 20:81-89.
- Lee, H. J., et al., Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res.* 2012, 22:539-548.
- Lee, J-S., et al., RNA-guided genome editing in *Drosophila* with the purified Cas9 protein. *G3 Genes/Genomes/Genetics.* 2014, 4(7):1291-1295. (5 pages).
- Notice of Allowance and Fees Due dated Feb. 4, 2015, U.S. Appl. No. 14/226,274, filed Mar. 26, 2014, pp. 1-22.
- Wee, L., et al., Argonaute divides its RNA Guide into domains with distinct functions and RNA-binding properties. *Cell.* 2012, 151(5):1055-1067. (23 pages).
- Weeks, A., et al., Structural and Biochemical Studies of a Fluoroacetyl-CoA-Specific Thioesterase Reveal a Molecular Basis for Fluorine Selectivity. *Biochemistry.* 2010, 49(43):9269-9279 and Supplementary Materials. (21 pages).
- Welch, M., et al., Chapter 3: Designing Genes for Successful Protein Expression. Eds: Voigt, C., *Methods in Enzymology.* Academic Press. 2011, 498:43-66. (24 pages).
- Wente, S. R. and Rout, M. P., The Nuclear Pore Complex and Nuclear Transport. *Cold Spring Harb Perspect Biol.* 2010, 2:a000562. (21 pages).
- Whisnant, A. W., et al., In depth analysis of the interaction of HIV-1 with cellular microRNA biogenesis and effector mechanisms. *mBio.* 2013, 4:10.1128/mbio.00193-13.
- White, T. B. and Lambowitz, A. M., The Retrohoming of Linear Group II Intron RNAs in *Drosophila melanogaster* Occurs by Both DNA Ligase 4-Dependent and -Independent Mechanisms. *PLoS Genetics.* 2012, 8(2):e1002534. (16 pages).
- Wiedenheft, B., et al., RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *PNAS.* 2011, 108(25):10092-10097. (7 pages).
- Wiedenheft, B., et al., Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature.* 2011, 477:486-489. (5 pages).
- Wiedenheft, B., et al., RNA-Guided Genetic Silencing Systems in Bacteria and Archaea. *Nature.* 2012, 482:331-338.
- Wieland, M., et al., Engineering of ribozyme-based riboswitches for mammalian cells. *Methods.* 2012, 56(3):351-357. (7 pages).
- Wilen, C. B., et al., Engineering HIV-Resistant Human CD4+ T Cells with CXCR4-Specific Zinc-Finger Nucleases. *PLoS Pathogens.* 2011, 7(4):e1002020. (15 pages).
- Wilusz, J. E., et al., tRNAs marked with CCACCA are targeted for degradation. *Science.* 2011, 334(6057):817-821.
- Woo, J. W., et al., DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology.* 2015, 33:1162-1164. (4 pages).
- Wood, A. J., et al., Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science.* 2011, 333 (6040):307 and Supplementary Materials. (27 pages).
- Wu, S. et al., Establishment of a PEG-mediated protoplast transformation system based on DNA and CRISPR/Cas9 ribonucleoprotein complexes for banana. *BMC Plant Biol.* 2020, 20:425. (10 pages).
- Wu, Z., et al., Effect of Genome Size on AAV Vector Packaging. *Molecular Therapy,* 2010, 18(1):80-86. (7 pages).
- Xu, K., The Next Generation Biotechnology for Apple Improvement and Beyond: The CRISPR/Cas9 Story. *New York Fruit Quarterly.* 2013, 21(4):19-22. (4 pages).
- Yamano, S., et al., Comparison of Transfection Efficiency of Nonviral Gene Transfer Reagents. *Mol Biotechnol.* 2010, 46:287-300. (14 pages).

(56)

References Cited

OTHER PUBLICATIONS

- Yang, H., et al., One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell*. 2013, 154(6):1370-1379 and Supplementary Materials.
- Yarris, 2012, "Programmable DNA scissors found for bacterial immune system", downloaded from <http://newscenter.lbl.gov/2012/06/28/programmabledna.scissors/> (4 pages).
- Yi, Y., et al., Current Advances in Retroviral Gene Therapy. *Current Gene Therapy*. 2011, 11(3):218-228.
- Yu, Z., et al., Highly Efficient Genome Modifications Mediated by CRISPR/Cas9 in *Drosophila*. *Genetics*. 2013, 195(1):289-291.
- Zaret, K. S., and Carroll, J. S., Pioneer transcription factors: establishing competence for gene expression. *Genes & Dev*. 2011, 25:2227-2241.
- Zhang, F., et al., Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol*. 2011, 29(2):149-153 and Supplementary Materials. (20 pages).
- Zhang, Y., et al., Processing-Independent CRISPR RNAs Limit Natural Transformation in *Neisseria meningitidis*. *Molecular Cell*. 2013, 50(4):488-503. (16 pages).
- Zhang, Y. et al., A highly efficient rice green tissue protoplast system for transient gene expression and studying light/chloroplast-related processes. *Plant Methods* 2011, 7(1):30. (14 pages).
- Zou, J., et al., Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease-mediated safe harbor targeting. *Blood*. 2011, 117(21):5561-5572. (13 pages).
- Zuris, J. A., et al., Efficient Delivery of Genome-Editing Proteins In Vitro and In Vivo. *Nat. Biotechnol*. 2015, 33(1):73-80. (26 pages).
- Angel, M. and Yanik, M. F., Innate Immune Suppression Enables Frequent Transfection with RNA Encoding Reprogramming Proteins. *PLoS One*. 2010, 5(7):e11756. (7 pages).
- Anonymous, Product Information Thermo Scientific TurboFect Transfection Reagent Pub. No. MAN0013147 Rev. Date Aug. 24, 2018. Thermo Scientific, (2018) (2 pages).
- Anonymous, PCT Third Party Observation dated Feb. 19, 2015, in relation to International Application No. PCT/KR2013/009488 filed on Oct. 23, 2013, pp. 1-5.
- Anonymous, Third-Party Observation dated Jul. 18, 2014, in relation to International Application No. PCT/US2013/033106 filed Mar. 20, 2013 (7 pages).
- Anonymous, Third-Party Observation dated Sep. 24, 2014, in relation to International Application No. PCT/US2013/032589 filed Mar. 15, 2013 (8 pages).
- Anonymous, Third-Party Observations dated Jan. 17, 2019, in relation to European Patent Application No. 18158147.1 (Publication No. EP 3372679 A1) (302 pages).
- Anonymous, Third-Party Submissions dated Dec. 22, 2014, in relation to U.S. Appl. No. 14/104,977, filed Dec. 12, 2013, pp. 1-56.
- Anonymous, Third-Party Submission Under 37 C.F.R. §1.290 filed Apr. 6, 2016 in relation to U.S. Appl. No. 14/438,098.
- Artimo, P., et al., ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research*. 2012, 40:W597-W603. (7 pages).
- ATCC record ATCC 700294, 2019, "*Streptococcus pyogenes* Rosenbach (ATCC 700294)", retrieved from the internet on Apr. 18, 2019 at http://www.lgcstandards-atcc.org/products/all/700294.aspx?geo_country=GB#history (2 pages).
- Australian Patent Application No. 2013335451 prosecution history.
- Barrangou, R., RNA-mediated programmable DNA cleavage. *Nat Biotechnol*. 2012, 30(9):836-838. (3 pages).
- Bassett, A. R., et al., Highly Efficient Targeted Mutagenesis of *Drosophila* with the CRISPR/Cas9 System. *Cell Reports*. 2013, 4(1):220-228 and Supplementary Table.
- Baum, A., et al., Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *PNAS*. 2010, 107(37):16303-16308.
- Bedell, V. M., et al., In vivo genome editing using a high-efficiency TALEN system. *Nature*. 2012, 491:114-118. (7 pages).
- Bell, O., et al., Determinants and dynamics of genome accessibility. *Nature Reviews Genetics*. 2011, 12:554-564. (11 pages).
- Bhaya, D., et al., CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual Reviews of Genetics*. 2011;45:273-297. (27 pages).
- Bikard, D., et al., CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In Vivo Bacterial Infection. *Cell Host & Microbe*. 2012, 12(2):177-186.
- Bikard, D., et al., Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*. 2013, 41(15):7429-7437.
- Boch, J. and Bonas, U., *Xanthomonas AvrBs3* Family-Type III Effectors: Discovery and Function. *Annual Review of Phytopathology*. 2010, 48:419-436.
- Bogerd, H. P., et al., A Mammalian Herpesvirus uses Noncanonical Expression and Processing Mechanisms to Generate Viral MicroRNAs. *Molecular Cell*. 2010, 37(1):135-142.
- Briggs, A.W., et al., Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Research*. 2012, 40(15):e117. (10 pages).
- Brouns, S. J. J., A Swiss Army Knife of Immunity. *Science*. 2012, 337(6096):808-809.
- Brzostek-Racine, S., et al., The DNA Damage Response Induces IFN. *The Journal of Immunology*. 2011, 187(10):5336-5345.
- Cain, C., CRISPR genome editing. *SciBX Sci. Exch.* 1-3, 2013, doi:10.1038/scibx.2013.77.
- Carlson, D. F., et al., Targeting DNA with fingers and TALENs. *Molecular Therapy-Nucleic Acids*. 2012, 1:e3. (4 pages).
- Carlson, E. D., et al., Cell-free protein synthesis: Applications come of age. *Biotechnology Advances*. 2012, 30(5):1185-1194.
- Carroll, D., a Crispr Approach to Gene Targeting. *Molecular Therapy*. 2012, 20(9):1658-1660. (3 pages).
- Carroll, D., Genome Engineering with Zinc-Finger Nucleases. *Genetics*. 2011, 188(4):773-782.
- Carroll, D., Staying on target with CRISPR-Cas. *Nat. Biotechnol*. 2013, 31:807-809.
- Cermak, T., et al., Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*. 2011, 39(12):e82.
- Chang, K., et al., RNAi in Cultured Mammalian Cells Using Synthetic siRNAs. *Cold Spring Harb. Protoc*. 2012, 9:957-961. (9 pages).
- Chang, N., et al., Genome editing with RNA-guided Cas9 nuclease in Zebrafish embryos. *Cell Res*. 2013, 23(4):465-472 and Supplementary Materials. (8 pages).
- Chapelaine, P., et al., Meganucleases can restore the reading frame of a mutated dystrophin. *Gene Ther*. 2010, 17:846-858. (13 pages).
- Chen, B. et al., Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas system. *Cell*. 2013, 155(7):1479-1491.
- Chen, F., et al., High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nature Methods*. 2011, 8(9):753-755 with Supplementary Information. (31 pages).
- Chen, F., et al., 2012, Methods and Reagents for Modifying Genomes Using RNA-Guided Endonucleases. in relation to U.S. Appl. No. 61/734,256.
- Cho, S. W., et al., Targeted genome engineering in human cells with Cas9 RNA-guided endonucleases. *Nat Biotechnol*. 2013, 31:230-232, and supplementary materials.
- Cho, S. W., et al., Heritable gene knockout in *Caenorhabditis elegans* by direct injection of Cas9-sgRNA ribonucleoproteins. *Genetics*, 2013, 195(3):1177-1180. (5 pages).
- Christian, M., et al., Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. *Genetics*. 2010, 186(2):757-761 and 2S1-8S1.
- Chugh, A., et al., Cell-penetrating peptides: Nanocarrier for Macromolecule Delivery in Living Cells. *IUBMB Life*. 2010, 62(3):183-193. (11 pages).
- Chylinski, K., et al., The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biology*. 2013, 10(5):726-737.

(56)

References Cited

OTHER PUBLICATIONS

- Clark, K. J., et al., A TALE of Two Nucleases: Gene Targeting for the Masses. *Zebrafish*, 2011, 8(3):147-149.
- Close, D., et al., Expression of Non-Native Genes in a Surrogate Host Organism. 2012. Retrieved from the Internet at <<http://www.intechopen.com>> (33 pages).
- Close, D., et al., The evolution of the bacterial luciferase gene cassette (*lux*) as a real-time bioreporter. *Sensors*. 2012;12(1):732-752. (21 pages).
- Cohen, J., et al., The Emerging Race to Cure HIV Infections. *Science*. 2011, 332(6031):784-789.
- Collins, C. A., and Brown, E. J., Cytosol as battleground: ubiquitin as a weapon for both host and pathogen. *Trends in Cell Biology*. 2010, 20(4):205-213.
- Cong, L., et al., Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*. 2013, 339(6121):819-823 and Supplementary Materials.
- CNLS Mapper results of *S. pyogenes* Cas9. NLS mapper. URL: http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_ycgi (Apr. 25, 2016) (1 page).
- Cradick, T. J., et al., CRISPR/Cas 9 systems targeting β -globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Research*. 2013, 41(20):9584-9592.
- Arpaia, N. and Barton, G. M., Toll-like receptors: key players in antiviral immunity. *Current Opinion in Virology*. 2011, 1(6):447-454. (15 pages).
- Chen, X., et al., Fusion protein linkers: property, design and functionality. *Advanced Drug Delivery Reviews*. 2013, 65(10):1357-1369. (32 pages).
- Kanchiswamy, C. N., et al., Fine-tuning next-generation genome editing tools. *Trends in Biotechnology*. 2016;34(7):562-574. (28 pages).
- Kim, S., et al. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res*. 2014, 24(6):1012-1019 (with supplemental materials). (15 pages).
- Lin, S., et al., Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife*. 2014, 3:e04766 (with supplemental materials). (18 pages).
- Osakabe Y., et al. CRISPR-Cas9-mediated genome editing in apple and grapevine, *Nature Protocols* Dec. 2018;13 2844-2863 (20 pages).
- Oxford Dictionary of Chemistry (7th ed. 2016). Oxford University Press. p. 516. (3 pages).
- Vu, A. T., et al., Extracellular double-stranded RNA induces TSLP via an endosomal acidification- and NF- κ B-dependent pathway in human keratinocytes. *Journal of Investigative Dermatology*. 2011, 131(11):2205-2212.
- Yu, K., et al., Synthetic fusion protein design and applications. *Biotechnology Advances*. 2015, 33(1):155-164 (10 pages).
- Zuris, J. A., et al., Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol*. 2015, 33(1):73-80 (with supplemental materials). (58 pages).
- Kanchiswamy, C. N., DNA-free genome editing methods for targeted crop improvement. *Plant Cell Rep*. 2016, 35:1469-1474. (6 pages).
- Lee, S., et al., Optimization of CRISPR/Cas9 ribonucleoprotein delivery into cabbage protoplasts for efficient DNA-free gene editing. *Plant Biotechnol Rep*. 2024, 18:415-424. (10 pages).
- Malnoy, M., et al., DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins. *Frontiers in Plant Science*. 2016, 7:01904. (9 pages).
- Jiang, W., et al., A Versatile and Efficient Plant Protoplast Platform for Genome Editing. *Frontier in Genome Editing*. 2021, 22:3:719190 (105 pages).
- Cristea, S., et al., Dissection of Splicing Regulation at an Endogenous Locus by Zinc-Finger Nuclease-Mediated Gene Editing. *PLoS One*. 2011, 6(2):e16961. (5 pages).
- Davis, G. D., and Cui, X., Zinc Finger Nucleases for Genome Editing. *Genetic Engineering & Biotechnology News*. 2010, 30(13). (6 pages).
- De Vries, R., DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. *Biochimie*. 2010, 92(12):1715-1721. (7 pages).
- Declaration of Technical Expert Paul Simons, executed Dec. 22, 2015, in relation to U.S. Appl. No. 14/704,551 (77 pages).
- Deltcheva, E., et al., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011, 471(7340):602-607 and Supplementary Information (71 pages).
- Deng, D., et al., Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*. 2012, 335(6069):720-723.
- Dicarlo, J. E., et al., Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Research*. 2013, 41(7):4336-4343.
- Dickinson, D. J., et al., Engineering the *Caenorhabditis elegans* genome using Cas9-triggered homologous recombination. *Nat. Methods*. 2013, 10:1028-1034.
- Ding, Q., et al., Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell*. 2013, 12(4):393-394. (2 pages).
- Documents related to the prosecution of European Patent Application No. EP13849670.8 (Publication No. EP 2 912 175 A1).
- Documents related to the prosecution of European Patent Application No. EP20130793997 (Publication No. EP2800811).
- Documents related to the prosecution of European Patent Application No. EP20130824232 (Publication No. EP2764103).
- Documents related to the prosecution of U.S. Appl. No. 14/685,510.
- Documents related to the prosecution of U.S. Appl. No. 13/842,859.
- Documents related to the prosecution of European Patent Application No. EP2771468 (Application No. EP 13818570.7).
- Documents related to the prosecution of European Patent Application No. EP2784162 (application No. EP 14170383.5).
- Documents related to the prosecution of U.S. Appl. No. 14/054,414.
- Documents related to the prosecution of U.S. Appl. No. 14/685,568.
- Documents relating to Patent Interference No. 106,048 (DK).
- Documents relating to Patent Interference No. 106,115 (DK).
- Documents relating to Patent Interference No. 106,126 (DK).
- Documents relating to Patent Interference No. 106,127 (DK) (169 pages).
- Doyon, J. B., et al., Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nature Cell Biology*. 2011, 13(3):331-337. (31 pages).
- Doyon, Y., et al., Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nature Methods*. 2011, 8:74-79. (38 pages).
- Ebina, H., et al., Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci Rep*. 2013, 3:2510. (7 pages).
- Edgar, R., and Qimron, U., The *Escherichia coli* CRISPR system protects from λ lysogenization, lysogens, and prophage induction. *J Bacteriol*. 2010, 192(23):6291-6294.
- Esvelt, K. M. et al., Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods*. 2013, 10:1116-1121.
- Fath, S., et al., Multiparameter RNA and Codon Optimization: A Standardized Tool to Assess and Enhance Autologous Mammalian Gene Expression. *PLoS One*. 2011, 6(3):e17596. (14 pages).
- Feng, Z., et al., Efficient genome editing in plants using a CRISPR/Cas system. *Cell Res*. 2013, 23(10):1229-1232.
- Fonfara, I., et al., Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Research*. 2014, 42(4):2577-2590. (14 pages).
- Fortier, S., et al., Genome-Wide Interrogation of Mammalian Stem Cell Fate Determinants by Nested Chromosome Deletions. *PLoS Genetics*. 2010, 6(12):e1001241. (14 pages).
- Francis, D. M. and Page, R., Strategies to Optimize Protein Expression in *E. coli*. *Current Protocols in Protein Science*. 2010, 61:5. 24.1-5.24.29. (29 pages).
- Friedland, A. E., et al., Heritable genome editing in *C. elegans* via a CRISPR-Cas 9 system. *Nat. Methods*. 2013, 10(8):741-743.

(56)

References Cited

OTHER PUBLICATIONS

- Fu, Y., et al., High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol.* 2013, 31(9):822-826.
- Fujii, W., et al., Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Research.* 2013, 41(20):e187.
- Gabriel, R., et al., An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol.* 2011, 29(9):816-823.
- Gagnon, J. A., et al., Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs. *PLoS One.* 2014, 9(5):e98186. Supplementary Information (25 pages).
- Gaj, T., et al., ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology.* 2013, 31(7):397-405. (9 pages).
- Gao, H., et al., Heritable targeting mutagenesis in maize using a designed endonuclease. *The Plant Journal.* 2010, 61(1):176-187. (13 pages).
- Garneau, J. E., et al., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* 2010, 468:67-71.
- Gasiunas, G., et al., Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *PNAS.* 2012, 109(39):E2579-E2586.
- Genomic Cruise Missiles, *Science*, 2012;338:1526-1527. (12 pages).
- Gilbert, L. A. et al., CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. *Cell.* 2013, 154(2):442-451. (10 pages).
- Golic, K. G., RNA-Guided Nucleases: A New Era for Engineering the Genomes of Model and Nonmodel Organisms. *Genetics.* 2013, 195(2):303-308. (6 pages).
- Gonzalez, B., et al, Modular system for the construction of zinc-finger libraries and proteins. *Nat Protoc.* 2010, 5(4):791-810.
- Goren, M., et al., The bacterial CRISPR/Cas system as analog of the mammalian adaptive immune system. *RNA Biology.* 2012, 9(5):549-554.
- Gottwein, E., et al., Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host & Microbe.* 2011, 10(5):515-526.
- Gratz, S. J., et al., Genome engineering of *Drosophila* with the CRISPR RNA-guided Cas9 nuclease. *Genetics*, 2013, 194(4):1029-1035 and supplementary materials. (17 pages).
- Groisman, E. A., et al., Bacterial Mg²⁺ Homeostasis, Transport, and Virulence. *Annual Reviews of Genetics.* 2013, 47:625-646. (22 pages).
- Guschin, D. Y., et al., A Rapid and General Assay for Monitoring Endogenous Gene Modification. In: Mackay, J., Segal, D. (eds.) *Engineered Zinc Finger Proteins. Methods in Molecular Biology.* 2010, vol. 649. Humana Press, Totowa, NJ.
- Li, D., et al., Heritable gene targeting in the mouse and rat using a CRISPR-Cas system. *Nat Biotechnol.* 2013, 31(8):681-683 and Supplementary Materials.
- Li, J-F., et al., Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat Biotechnol.* 2013, 31(8):688-691 and Supplemental Materials.
- Li, T., et al., Modularly Assembled Designer TAL Effector Nucleases for Targeted Gene Knockout and Gene Replacement in Eukaryotes. *Nucleic Acids Research.* 2011, 39(14): 6315-6325.
- Li, T., et al., TAL nucleases (TALNs): hybrid protein composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Research.* 2011, 39(1):359-372. (14 pages).
- Li, W., et al., Simultaneous generation and germline transmission of multiple gene mutations in rat using CRISPR-Cas systems. *Nat Biotechnol.* 2013, 31(8):684-686 and Supplementary Materials.
- Lieber, M. R., The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End Joining Pathway. *Annual Review of Biochemistry.* 2010, 79:181-211. (34 pages).
- Lino, C. A., et al., Delivering CRISPR: a review of the challenges and approaches. *Drug Delivery.* 2018, 25(1):1234-1257.
- Liu, J., et al., Efficient and Specific Modifications of the *Drosophila* Genome by Means of an Easy TALEN Strategy. *Journal of Genetics and Genomics.* 2012, 39(5):209-215. (7 pages).
- Liu, P-Q., et al., Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnology and Bioengineering.* 2010, 106(1):97-105.
- Liu, Y., et al., Polyethylene glycol (PEG)-mediated transformation of the fused egfp-hph gene into *Pleurotus ostreatus*. *African Journal of Biotechnology.* 2012, 11(19):4345-4353. (9 pages).
- Liu, Y-C., et al., Efficient Polyethylene Glycol (PEG) Mediated Transformation of the Moss *Physcomitrella patens*. *J. Vis. Exp.* 2011, 50:2560. (4 pages).
- Lo, T-W., et al., Precise and Heritable Genome Editing in Evolutionarily Diverse Nematodes Using TALENs and CRISPR/Cas9 to Engineer Insertions and Deletions. *Genetics.* 2013, 195(2):331-348. (26 pages).
- Lu, C. and Li, P., Preparation of Short RNA by In Vitro Transcription. In: Conn, G. (eds) *Recombinant and In Vitro RNA Synthesis. Methods in Molecular Biology*, 2013, vol. 941. Humana Totowa, NJ. (12 pages).
- Ma, S. et al., Highly efficient and specific genome editing in silkworm using custom TALENs. *PLoS One.* 2012, 7(9):e45035. (7 pages).
- Madrugal, P., and Pawel, K., Current bioinformatic approaches to identify DNase I hypersensitive sites and genomic footprints from DNase-seq data. *Frontiers in Genetics.* 2012, 3(230):1-3. (3 pages).
- Maeder, M. L. et al., CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods.* 2013, 10:977-979.
- Magnani, L., et al., Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends in Genetics.* 2011, 27(11):465-474. (10 pages).
- Mahfouz, M. M., et al., De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. *PNAS.* 2011, 108(6):2623-2628.
- Mak, A. N-S., et al., The Crystal Structure of TAL Effector PthXo1 Bound to Its DNA Target. *Science.* 2012, 335(6069):716-719. (5 pages).
- Makarova, K. S., et al., Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol.* 2011, 9:467-477. (11 pages).
- Mali, P., et al., Cas9 as a Versatile Tool for Engineering Biology. *Nat Methods.* 2013, 10(10):957-963.
- Mali, P., et al., RNA-Guided Human Genome Engineering via Cas9. *Science.* 2013, 339(6121):823-826 and Supplementary Materials. (40 pages).
- Mali, P., et al., CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol.* 2013, 31:833-838.
- Marchfelder, A., et al., Small RNAs for defence and regulation in archaea. *Extremophiles.* 2012, 16:685-696. (12 pages).
- Maresca, M., et al., Obligate ligation-gated recombination (ObliGaRe): Custom-designed nuclease-mediated targeted integration through nonhomologous end joining. *Genome Res.* 2013, 23:539-546.
- Marfori, M., et al., Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research.* 2011, 1813(9):1562-1577. (16 pages).
- Marraffini, L. A. and Sontheimer, E. J., Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature.* 2010, 463(7280):568-571. (13 pages).
- Mashimo, T., et al., Generation of Knockout Rats with X-Linked Severe Combined Immunodeficiency (X-SCID) Using Zinc-Finger Nucleases. *PLoS One* 2010, 5(1):e8870. (7 pages).
- Mazzara, G. P., et al., Maturation events leading to transfer RNA and ribosomal RNA. *Cell Biology.* 2012, 3:439-545.
- Miao, J., et al., Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* 2013, 23(10):1233-1236.
- Mika, J. T., and Poolman, B., Macromolecule diffusion and confinement in prokaryotic cells. *Current Opinion in Biotechnology.* 2011, 22(1):117-126. (10 pages).
- Miller, J. C., et al., A Tale nuclease architecture for efficient genome editing. *Nat Biotechnol.* 2011, 29(2):143-148 and online methods.

(56)

References Cited

OTHER PUBLICATIONS

Mougiakos, I., et al., Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nat Commun.* 2017, 8:1647. (11 pages).

Mussolino, C., et al., A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research.* 2011, 39(21):9283-9293.

Mussolino, C., and Cathomen, T., RNA guides genome engineering. *Nat. Biotechnol.* 2013, 31:208-209.

Mussolino, C., and Cathomen, T., Tale nucleases: tailored genome engineering made easy. *Current Opinion in Biotechnology.* 2012, 23(5):644-650. (7 pages).

Musunuru, K., Genome editing of human pluripotent stem cells to generate human cellular disease models. *Dis Model Mech.* 2013, 6(4):896-904. (17 pages).

Nakayama, T., et al., Simple and efficient CRISPR/Cas9-mediated targeted mutagenesis in *Xenopus tropicalis*. *Genesis.* 2013, 51(12):835-843. (15 pages).

Nature Biotechnology Journal webpage, vol. 30 issue 9, Sep. 10, 2012, available at <https://www.nature.com/nbt/volumes/30/issues/9> (last accessed on Mar. 22, 2021), (15 pages).

NCBI record for NC-002737.2, 2019, "*Streptococcus pyogenes* M1 GAS, complete sequence", retrieved from the internet at < https://www.ncbi.nlm.nih.gov/nuccore/NC_002737.2 > (1 page).

Nekrasov, V., et al., Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat Biotechnol.* 2013;31(8):691-693 and Supplemental Materials.

Ng, Y. S., et al., Chemical transfection of dye-conjugated microRNA precursors for microRNA functional analysis of M2 macrophages. *Journal of Cellular Biochemistry.* 2012, 113(5):1714-1723. (10 pages).

Niinaka, Y., et al., Silencing of autocrine motility factor induces mesenchymal-to-epithelial transition and suppression of osteosarcoma pulmonary metastasis. *Cancer Res.* 2010, 70(22):9483-9493. (11 pages).

Noland, C. L., et al., siRNA Repositioning for Guide Strand Selection by Human Dicer Complexes. *Molecular Cell.* 2011, 43(1):110-121. (12 pages).

Ogawa, A., Rational design of artificial riboswitches based on ligand-dependent modulation of internal ribosome entry in wheat germ extract and their applications as label-free biosensors. *RNA.* 2011, 17:478-488.

Pandika, M., 2014, "Jennifer Doudna, CRISPR Code Killer", www.ozy.com/rising-stars-and-provocateurs/jennifer-doudna-crispr-code-killer/4690, Jan. 7, 2014, pp. 1-6.

Pardo, R., et al., pARIS-htt: an optimised expression platform to study huntingtin reveals functional domains required for vesicular trafficking. *Mol Brain.* 2010, 3:17. (17 pages).

Pattanayak, V., et al., Revealing off-target cleavage specificities of zinc finger nucleases by in vitro selection. *Nat Methods.* 2011, 8(9):765-770.

Pattanayak, V., et al., High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* 2013, 31:839-843.

Pennisi, E., Encode Project Writes Eulogy for Junk DNA. *Science.* 2012, 337(6099):1159-1161. (2 pages).

* cited by examiner

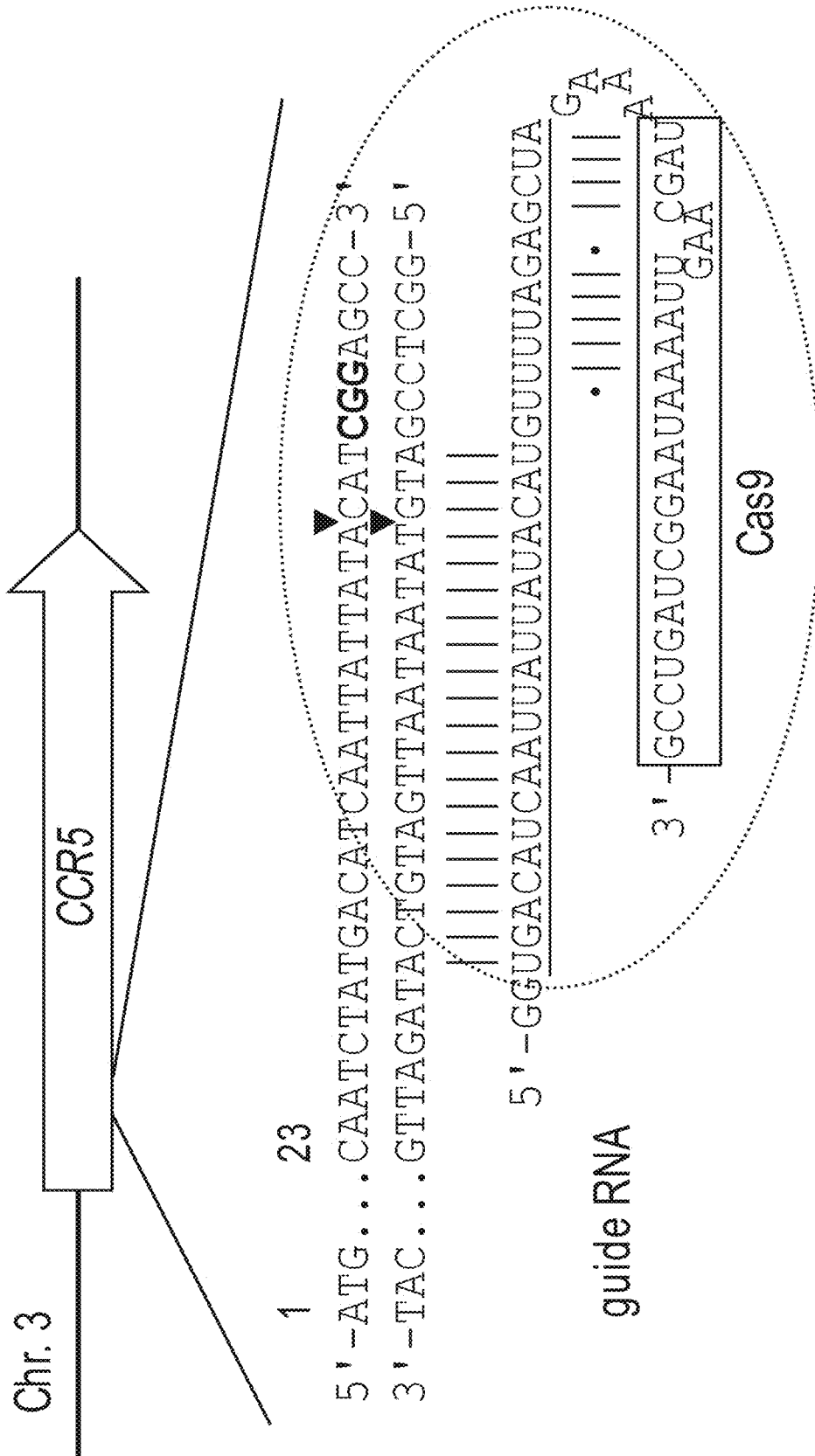
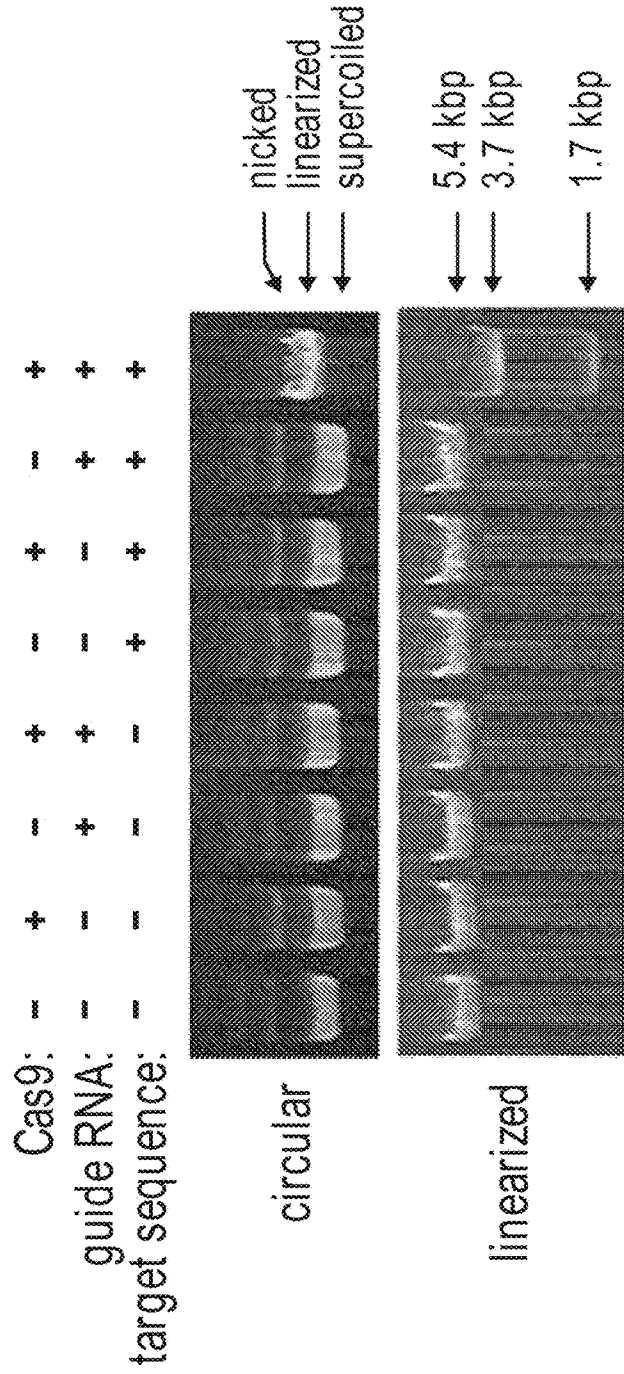
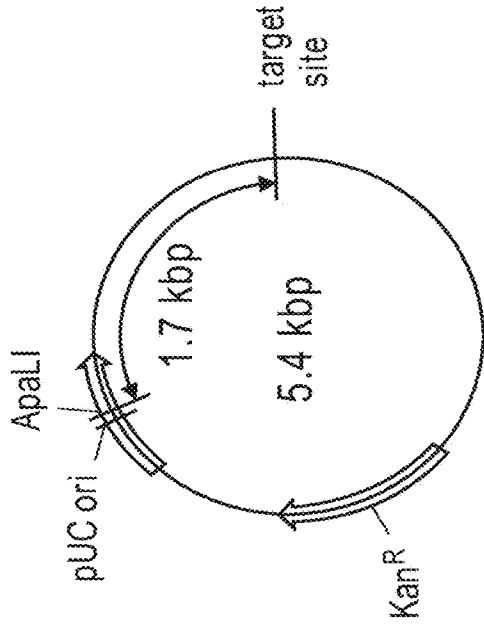


FIG. 1A

FIG. 1B



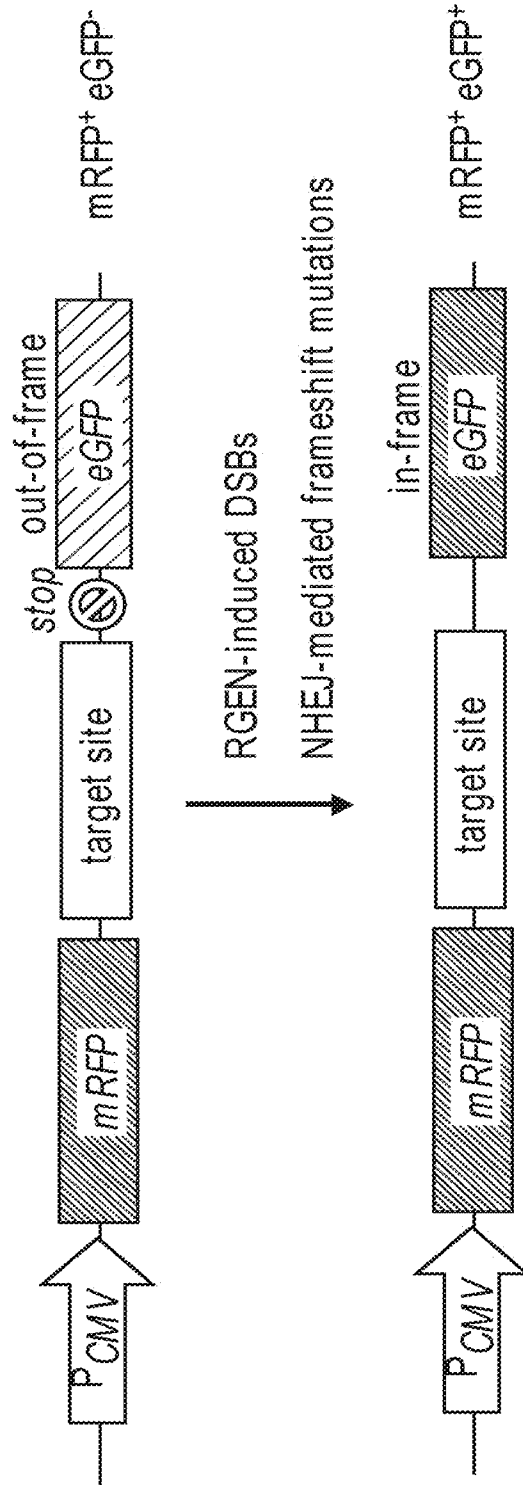


FIG. 2A

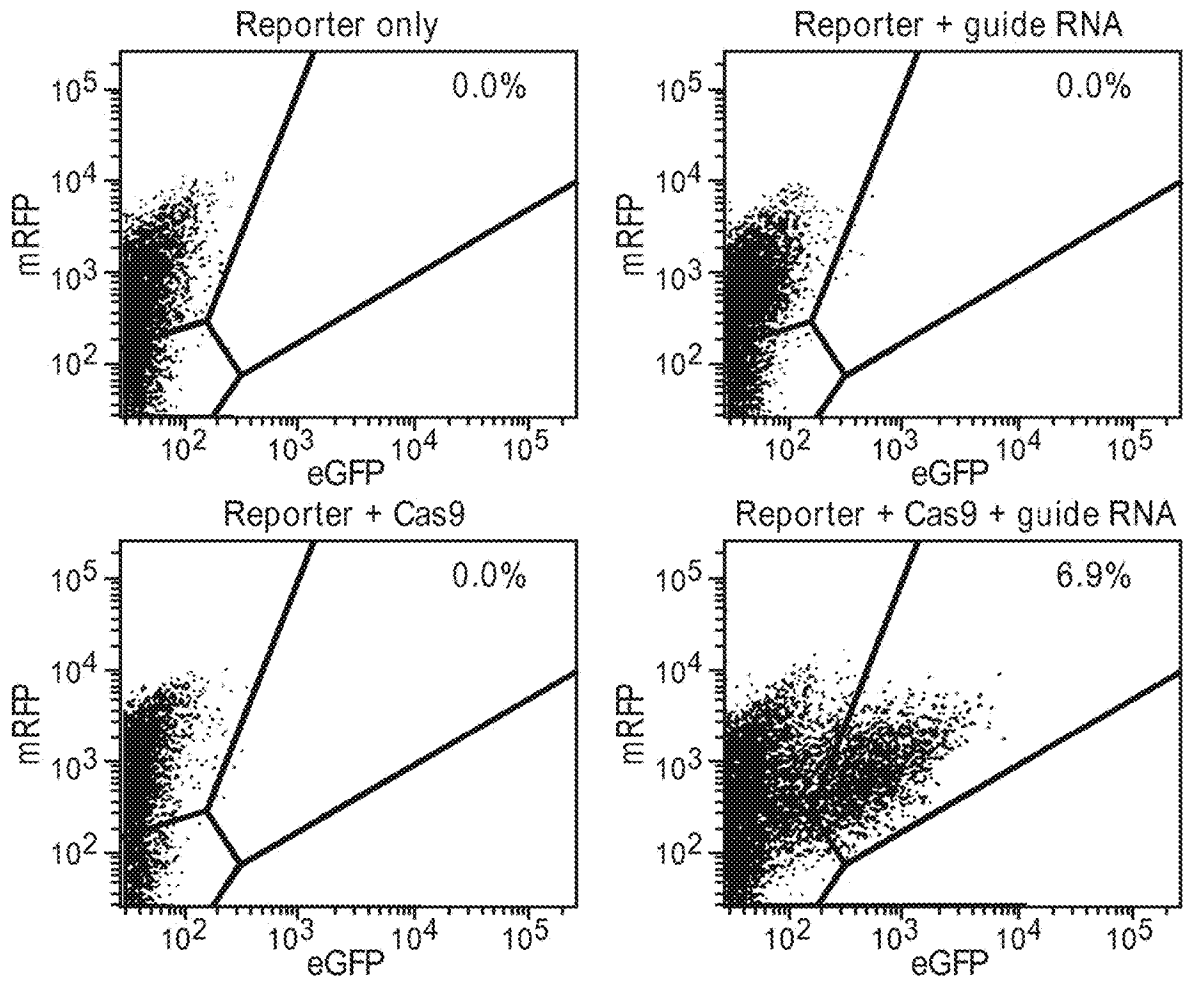


FIG. 2B

CCR5

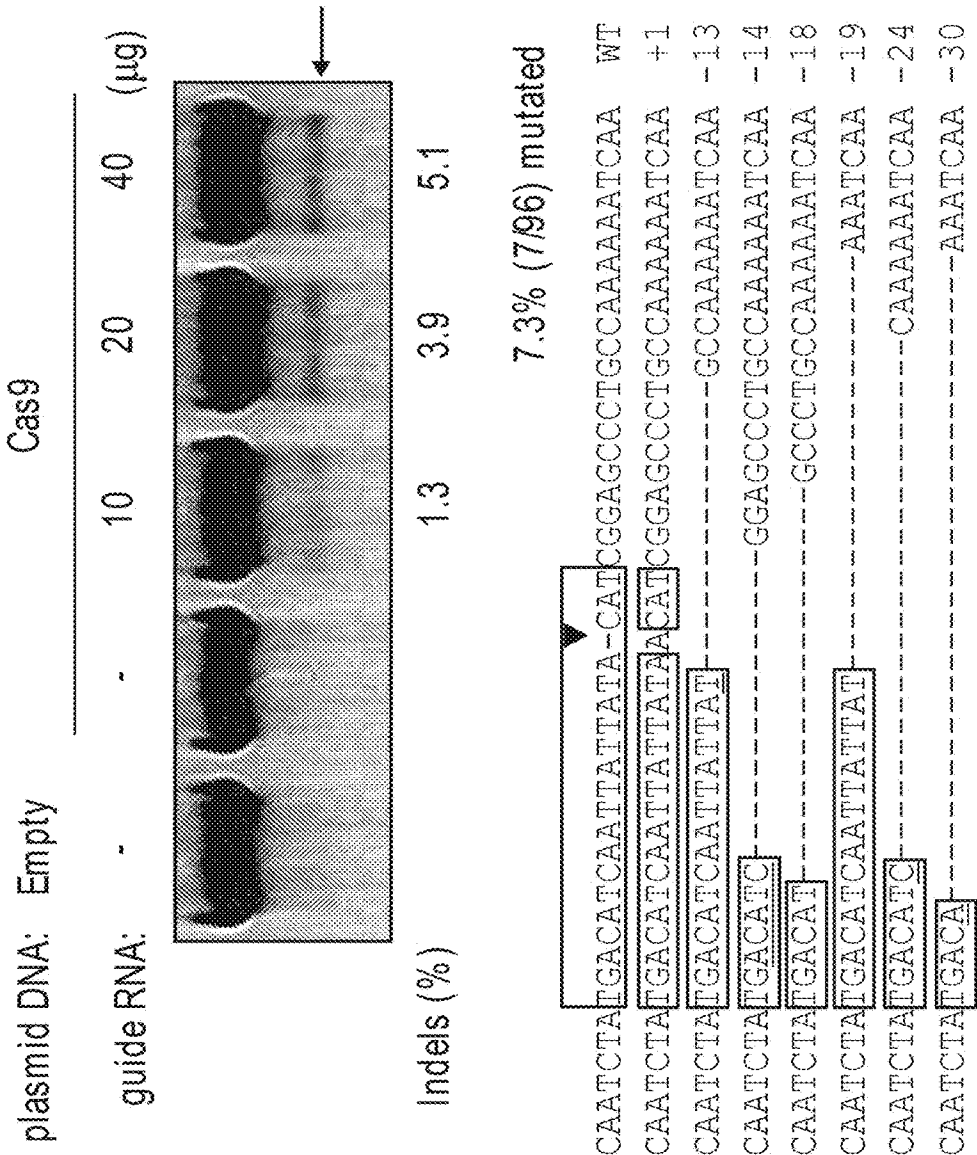
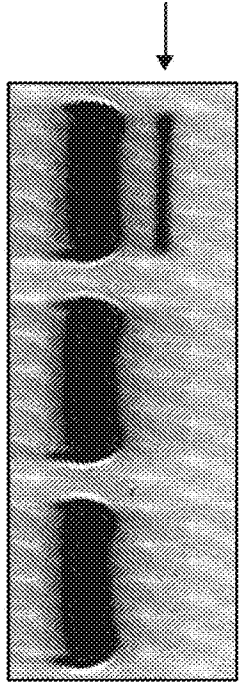


FIG. 3A

C4BPB

plasmid DNA: Empty Cas9

guide RNA: - +



Indels (%)

14

8.3% (4/48) mutated

TATGTGCAATGACCACTACATCCT--CAAAGGGCAGCAATCGGAGCCAG WT
 TATGTGCAATGACCACTACATCCTT-CAAAGGGCAGCAATCGGAGCCAG +1
 TATGTGCAATGACCACTACATCCTCTCAAAGGGCAGCAATCGGAGCCAG +2
 TATGTGCAATG AATG-----GAGCCAG -30
 TATGTGCAATGAC-----180 bp-----180

FIG. 3B

TGACATCAATTATTATACAT	▼	CGG	CCR5				
TGACATCAATTATTATA	g	AT	g	GA	ADCY5		
TGACATCA	c	TTATTAT	g	CAT	g	GG	KCNJ6
TGACAT	a	AATTATT	c	TACAT	g	GG	CNTNAP2
TGA	a	ATCAATTAT	c	ATA	g	ATCGG	Chr. 5 N/A

FIG. 4A

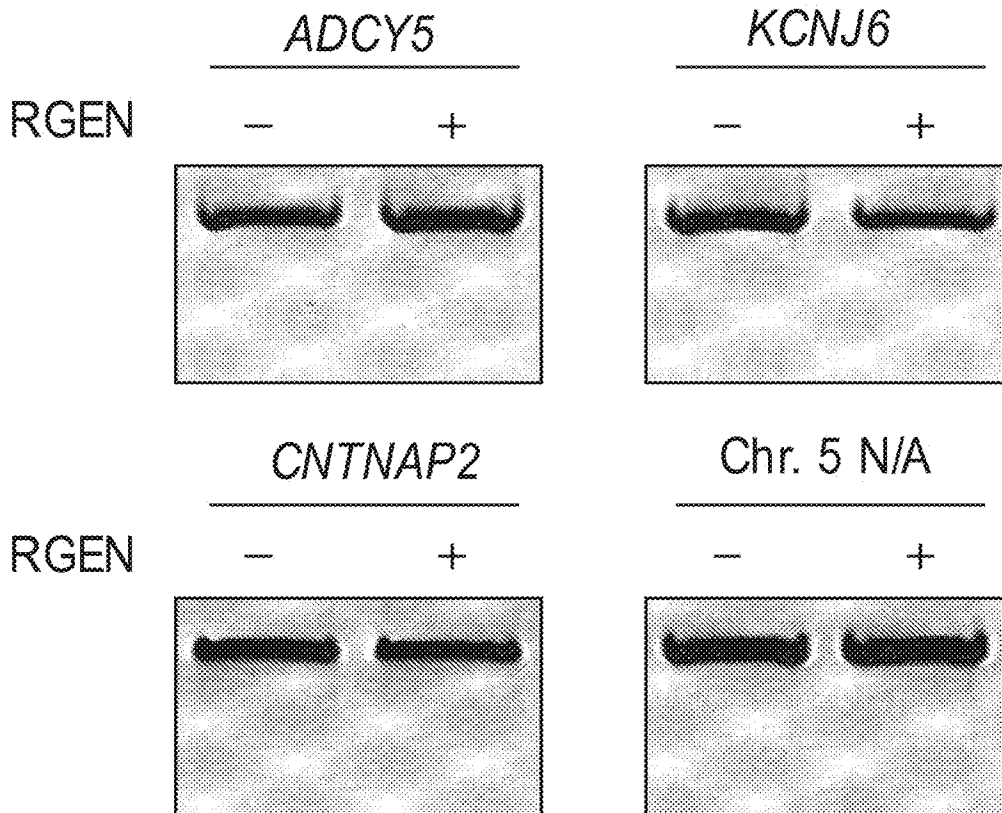


FIG. 4B

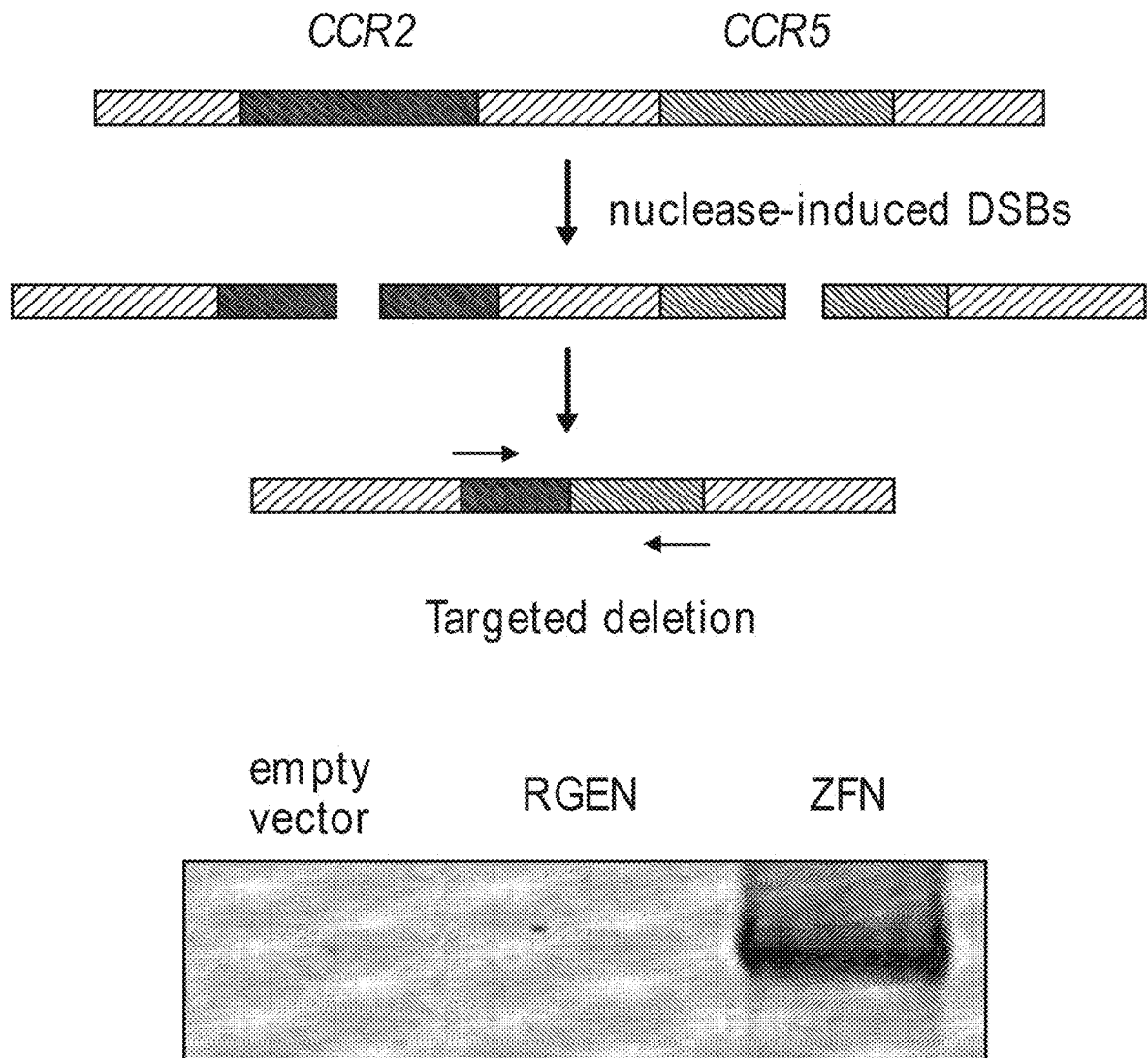


FIG. 4C

Foxn1

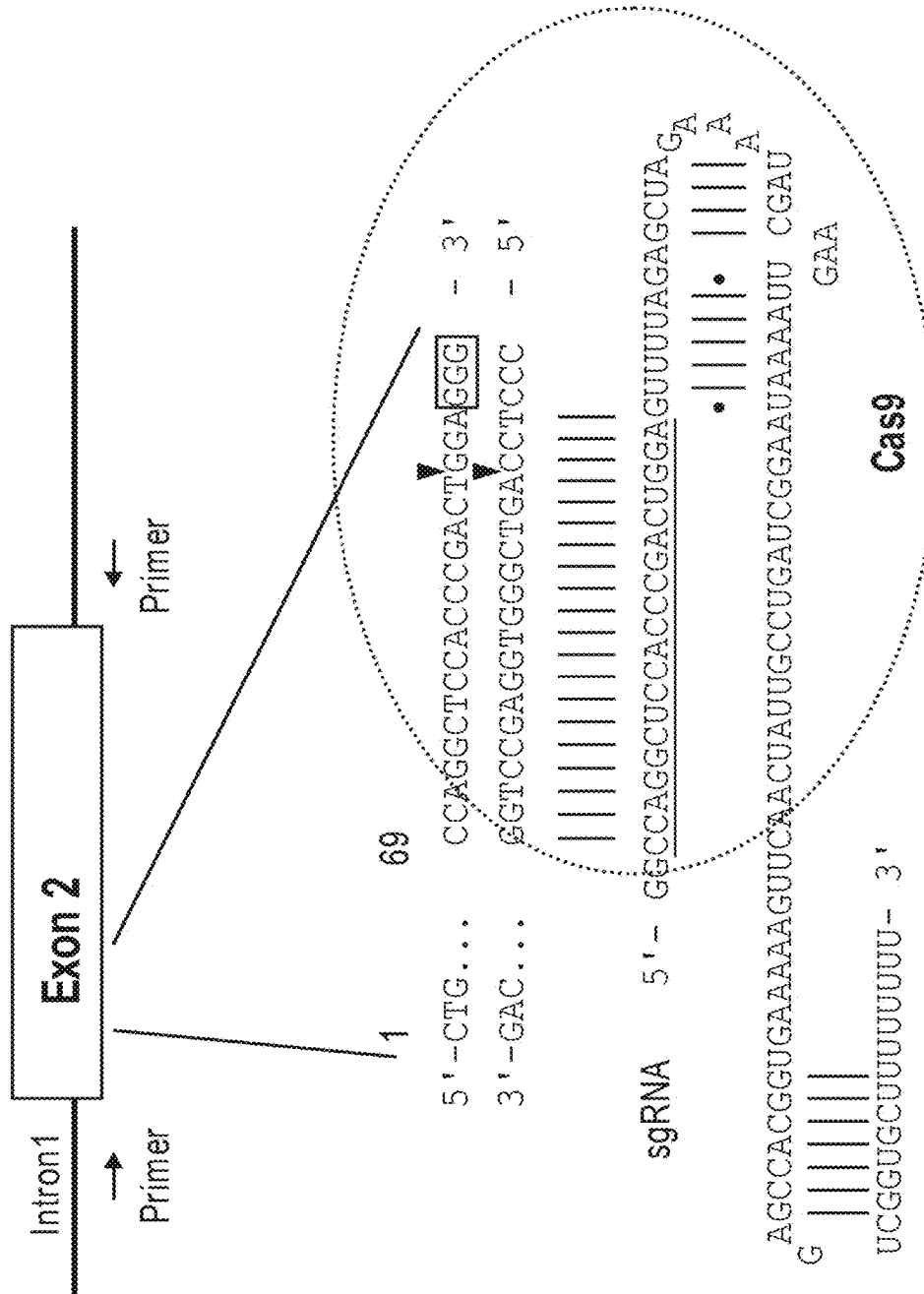


FIG. 5A

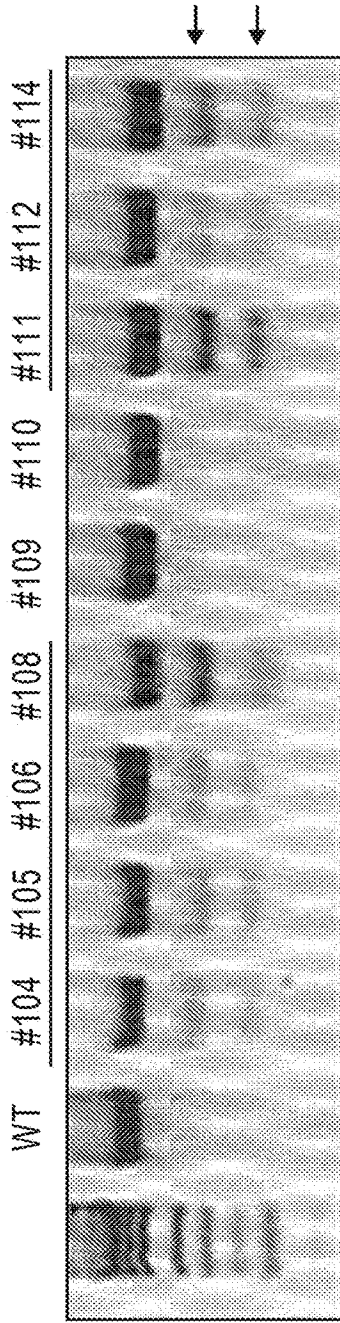


FIG. 5B

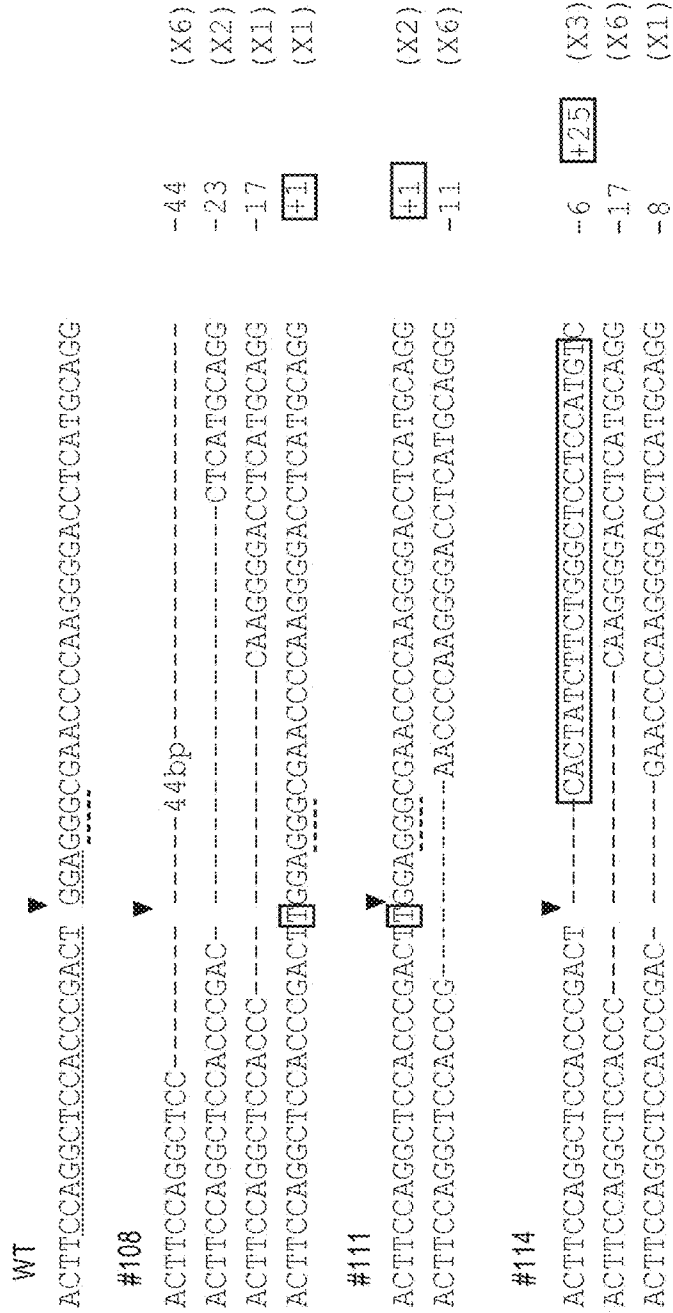


FIG. 5C

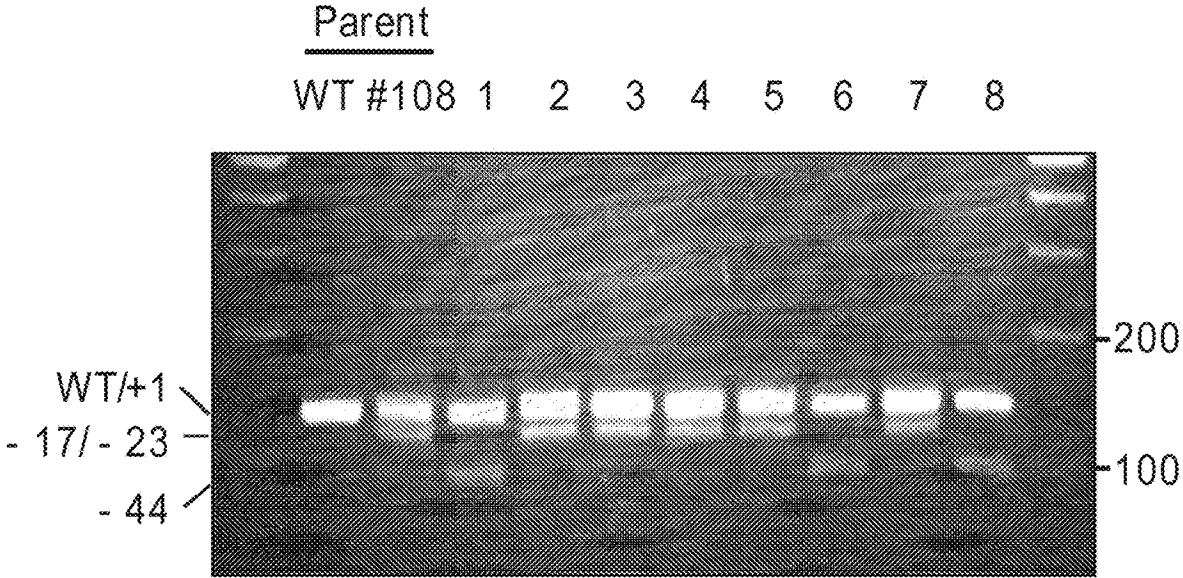


FIG. 5D

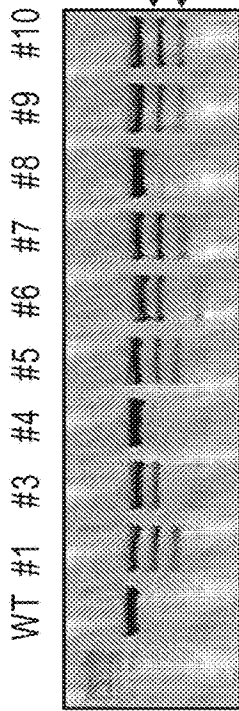


FIG. 6A

Cas9 mRNA (ng/μl)	sgRNA (ng/μl)	Tested embryos	Mutants (%)
10	1	27	9 (33)
10	10	49	28 (57)
10	100	45	41 (91)

FIG. 6B

sgRNA (ng/μl)	L	P	G	S	T	R	L	E	G	E	P	Q	G	D	L	M	Q	A	Mutants				
100	#1	A	CTT	CCA	GGC	TCC	ACC	CGA	CT	G	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC	WT		
	#3	A	CTT	CCA	GGC	TCC	ACC	CGA			--A	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC	Δ11		
	#2	A	CTT	CCA	GGC	TCC	ACC	CGA	C		--A	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GC	--	Δ11+Δ17		
	#5	A	CTT	CCA	GGC	TCC	ACC	CGA	C													Δ57	
	#11	A	CTT	CCA	GGC	TCC	ACC	CGA	C														Δ11
	#9	A	CTT	CCA	GGC	TCC	ACC	CGA	C														Δ17
	#11	A	CTT	CCA	GGC	TCC	ACC	CGA	CTG	GGG	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC		+1	
	#9	A	CTT	CCA	GGC	TCC	ACC	CGA	GGG	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC				Δ12
	#11	A	CTT	CCA	GGC	TCC	ACC	CGA	GGG	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC				Δ72
	#9	A	CTT	CCA	GGC	TCC	ACC	CGA	GGG	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC				Δ25
	#11	A	CTT	CCA	GGC	TCC	ACC	CGA	GGG	GAA	CCC	CAA	GGG	GAC	CTC	ATG	CAG	GCT	CC				Δ24

FIG. 6C

Pronucleus injection

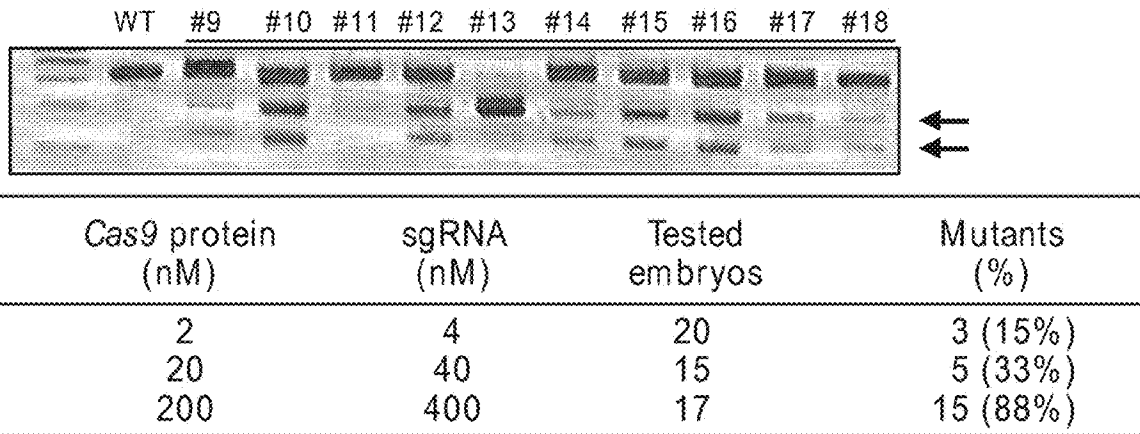


FIG. 7A

Intra-cytoplasmic injection

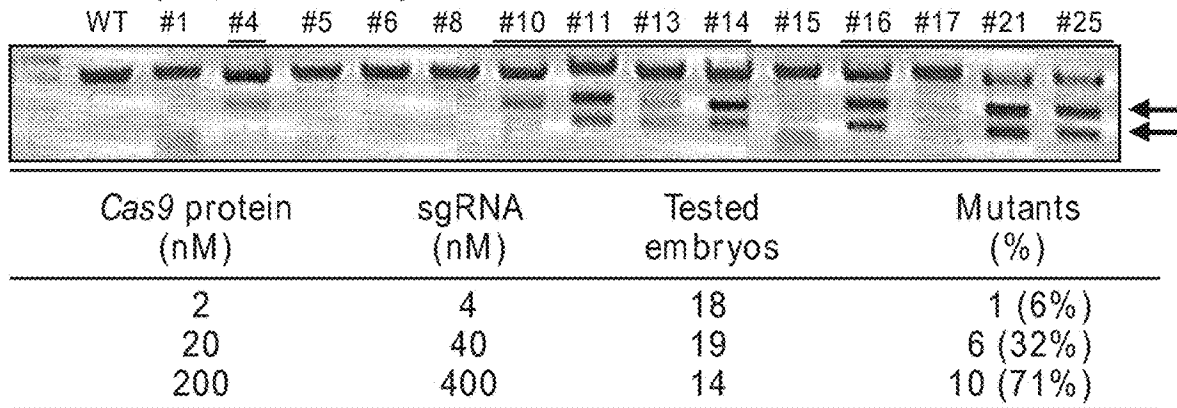


FIG. 7B

Sequence	Indels	Embryo no.
ACTTCCAGGCTCCACCCGACTGGAGGGCGGAACCCCAAGGGACCTCATGCAG	WT	
ACTTCCAGGCGAACCC-----AAGGGACCTCATGCAG	Δ18	2
ACTTCCAGGCTCCAC-----AAGGGACCTCATGCAG	Δ20	1
ACTTCCAGGCTCCACCC-----AAGGGACCTCATGCC	Δ19	1
ACTTCCAGGCTCCACCC-----CAAGGGACCTCATGCAG	Δ17	1
ACTTCCAGGCTCCACCCGA-----ACCCCAAGGGACCTCATGCAG	Δ11	3
ACTTCCAGGCTCCACCCGAA--GGAGGGCGAACCCCAAGGGACCTCATGCAG	Δ3+1	1
ACTTCCAGGCTCCACCCGACT--AGGGCGAACCCCAAGGGACCTCATGCAG	Δ2	1
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGACCTCATGCAG	+1	1
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGACCTCATGCAG	+1	10
ACTTCCAGGCTCCACCCGGA-----GGCGAACCCCAAGGGACCTCATGCAG	Δ6	1
ACTTCCAGGCTCCACCCGGA-----GGCGAACCCCAAGGGACCTCATGCAG	Δ5	2
ACTTCCAGGCTCCACCC-----TCATGCAG	Δ28	1
-----AGGGCGAACCCCAAGGGACCTCATGCAG	Δ126	1
Total		26

FIG. 7C

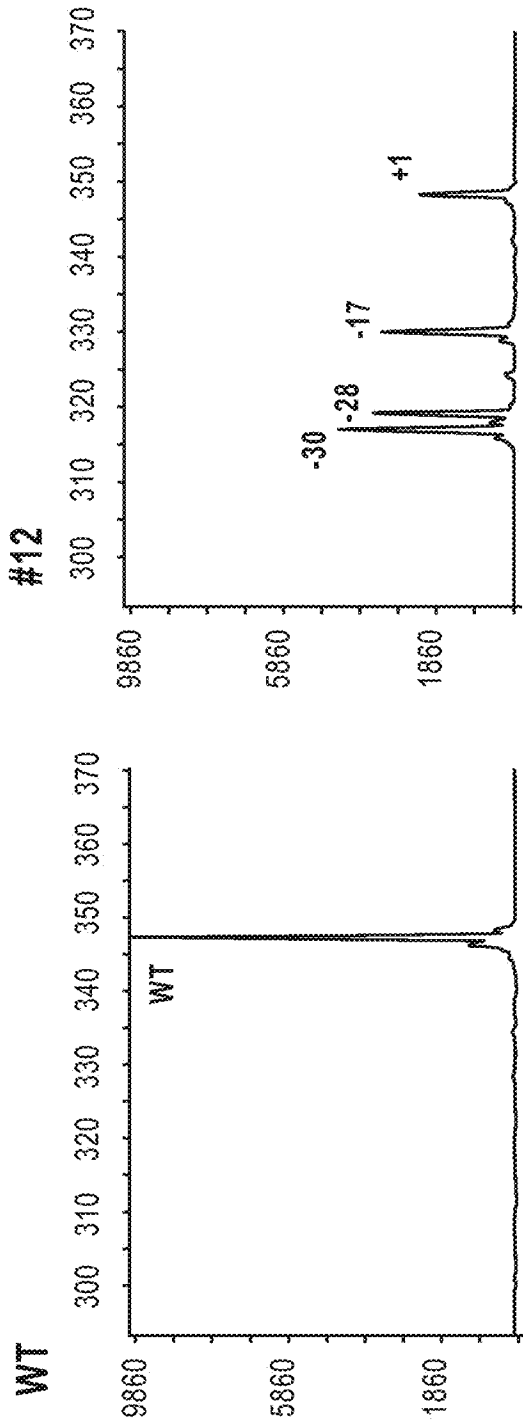


FIG. 8A

FIG. 8B

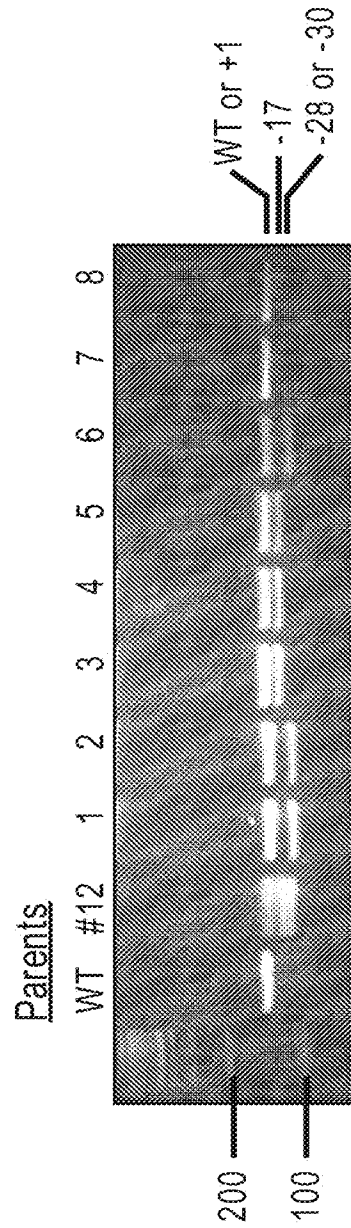


FIG. 8C

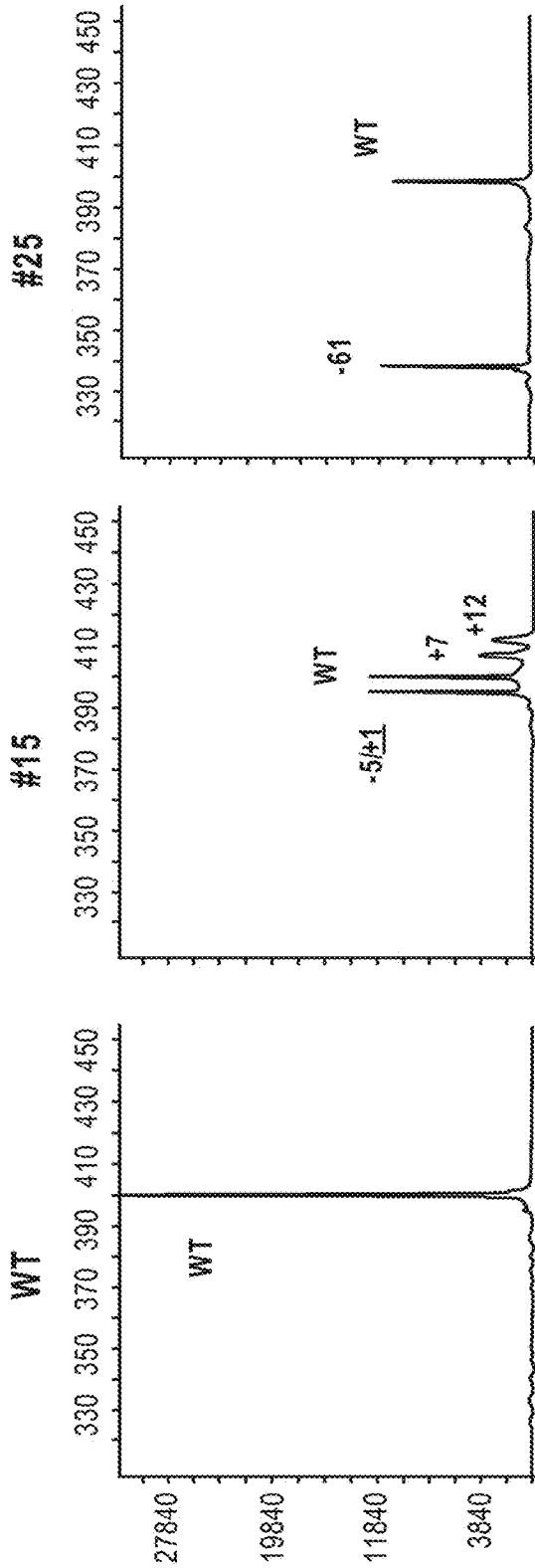


FIG. 9A

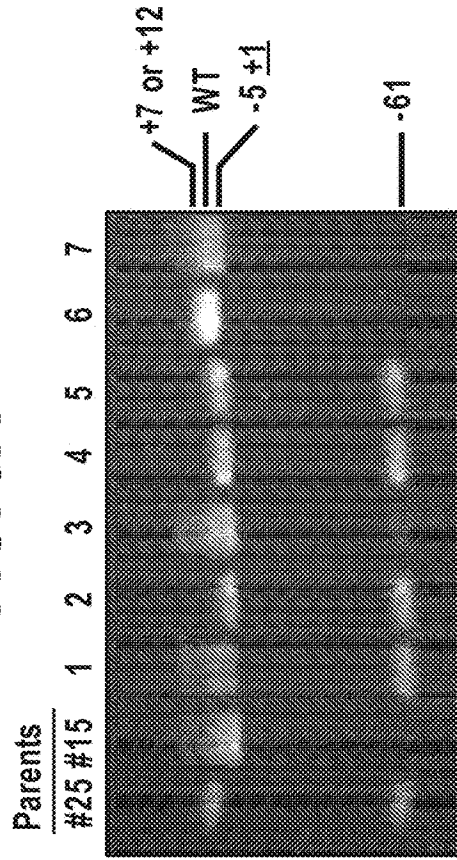
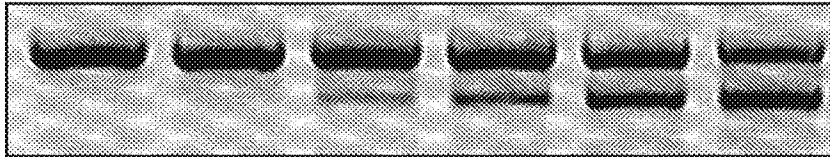


FIG. 9B

CCR5 #4

sgRNA	-	100 (29)	10(2.9)	30(8.8)	100(29)	50(15)	μg (μM)
Cas9 protein	225 (14)	-	22.5(1.4)	75(4.5)	225(14)	-	μg (μM)
Cas9 plasmid	-	-	-	-	-	5	μg



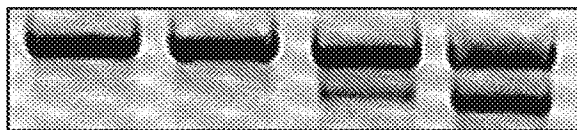
4.8 18 38 47 indel (%)

sgRNA	-	100 (29)				50(15)	μg (μM)
Cas9 protein	225 (14)	-	22.5(1.4)	75(4.5)	225(14)	-	μg (μM)
Cas9 plasmid	-	-	-	-	-	5	μg



12 30 33 43 indel (%)

crRNA	-	40(29)	20(15)	μg (μM)	
tracrRNA	-	80(29)	40(15)	μg (μM)	
Cas9 protein	225(14)	-	225(14)	-	μg (μM)
Cas9 plasmid	-	-	-	5	μg



9.4 42 indel (%)

FIG. 10A

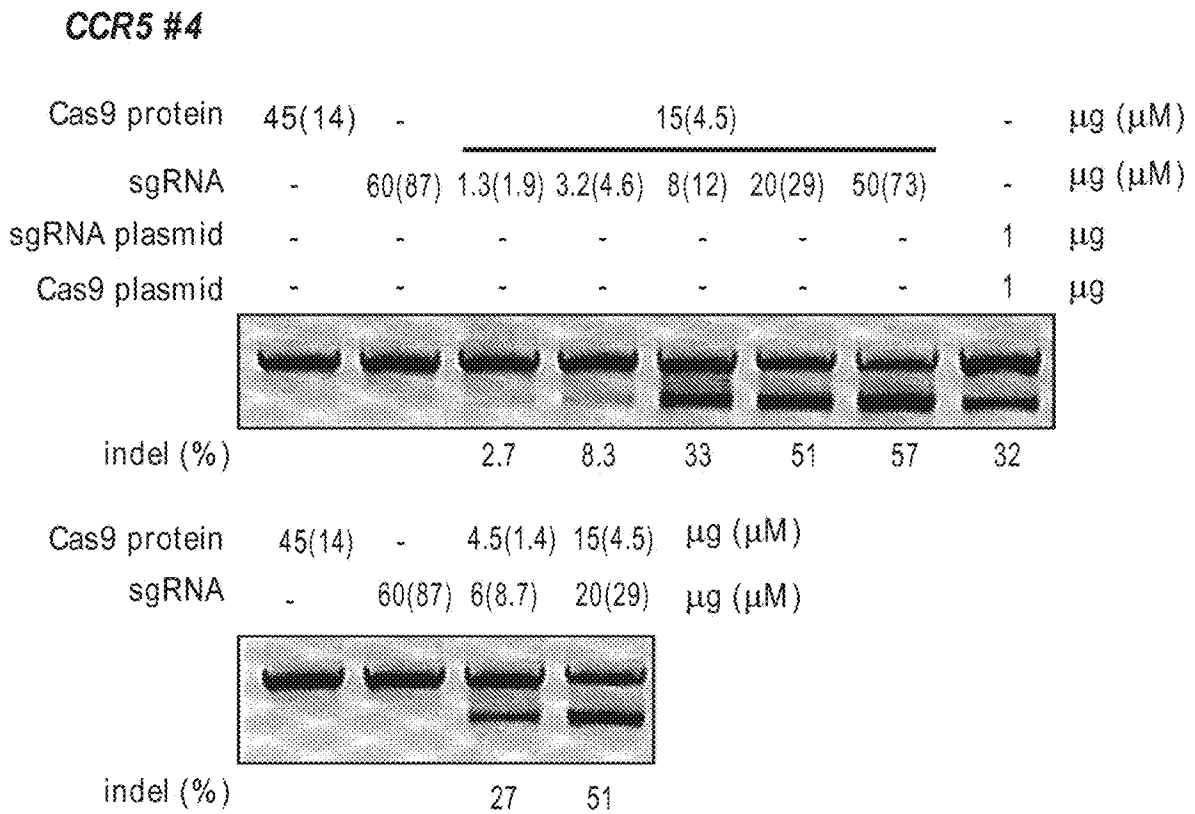


FIG. 10B

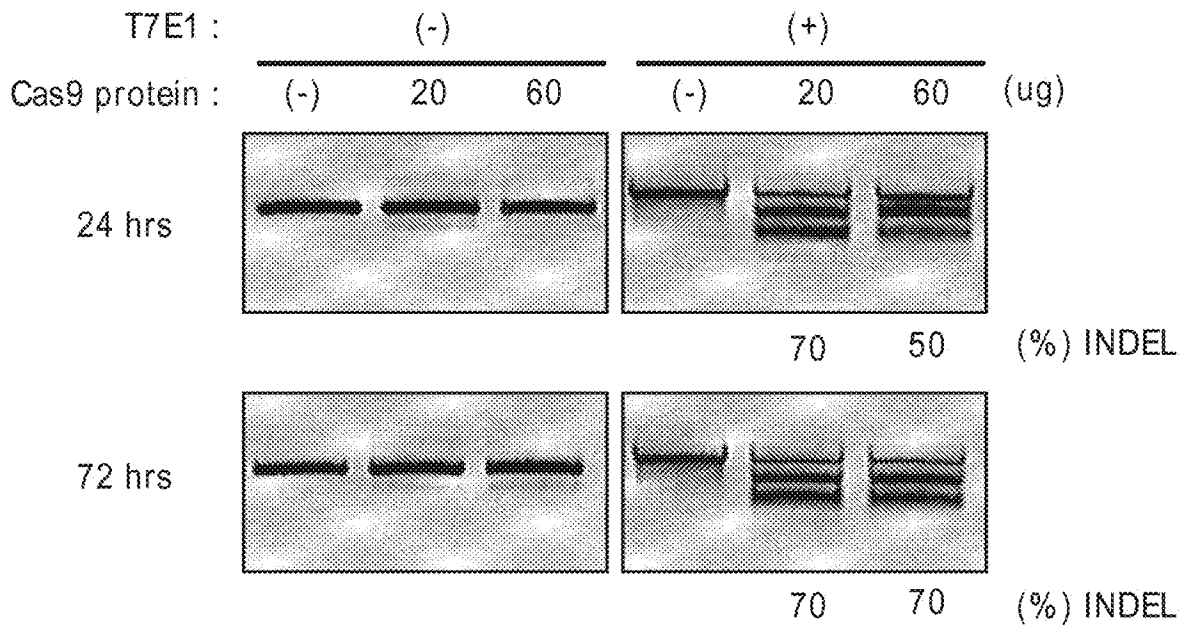


FIG. 11

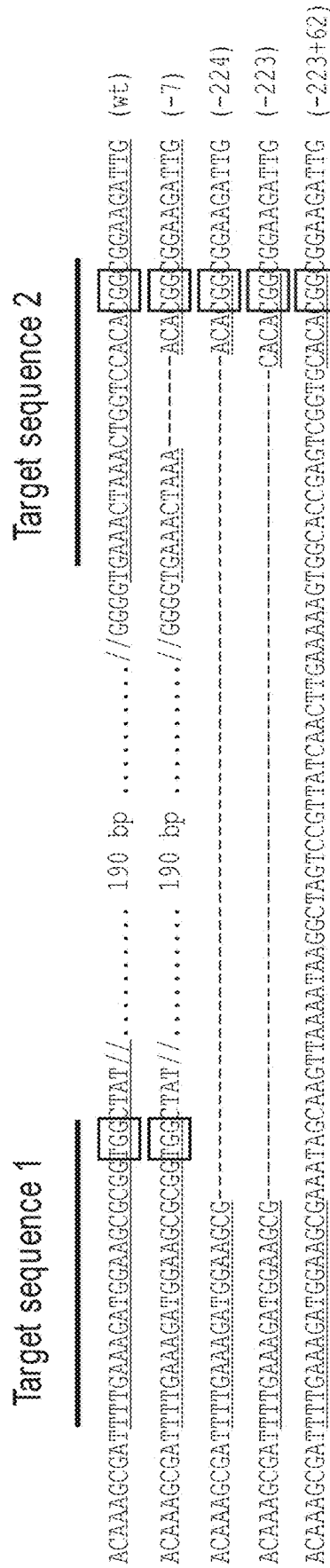


FIG. 12

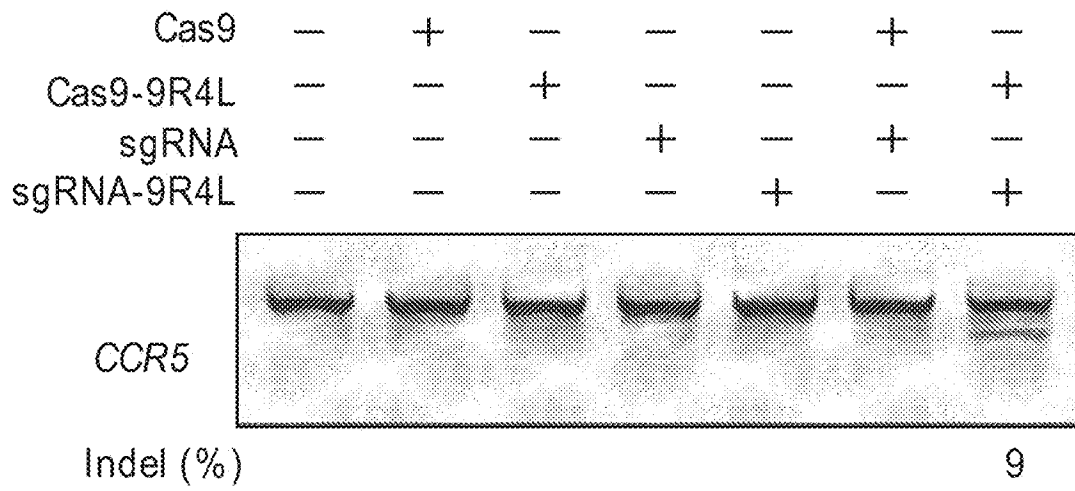


FIG. 13

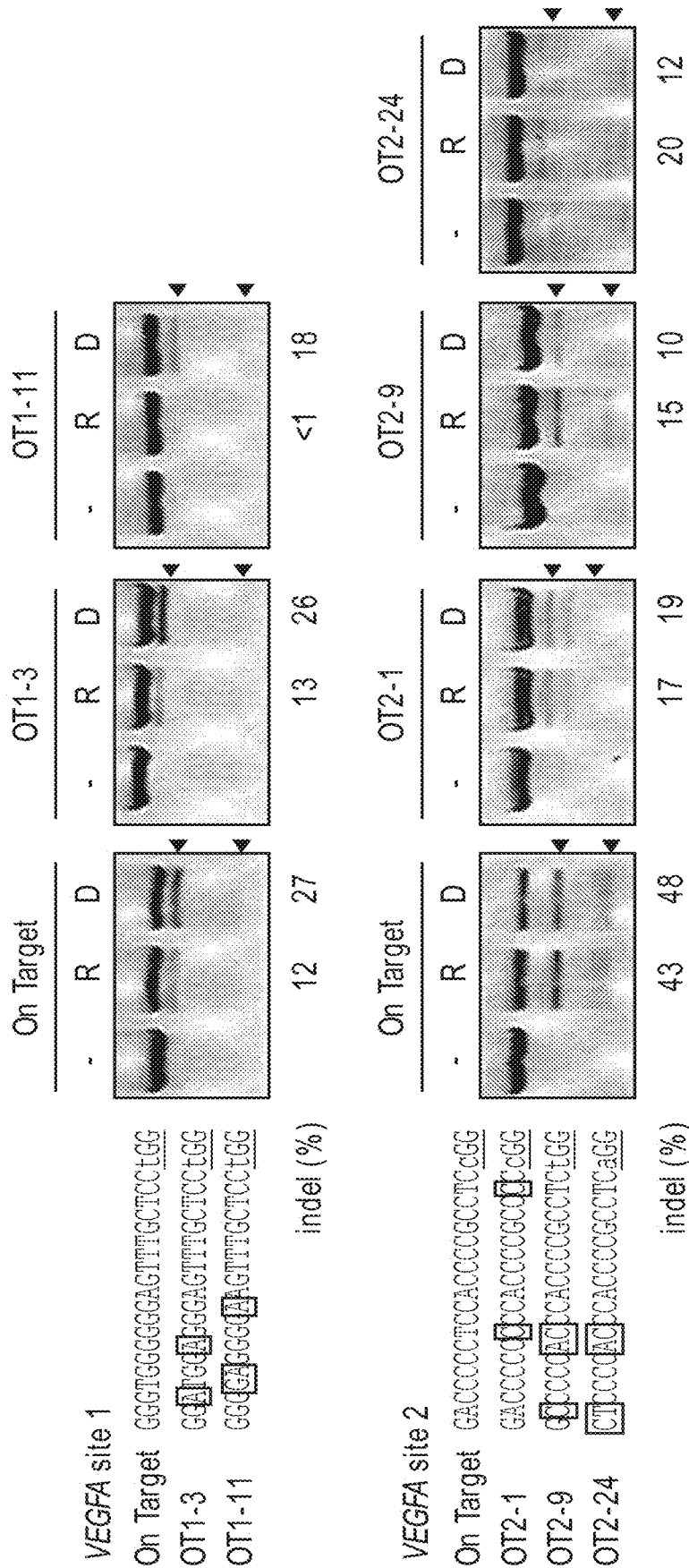


FIG. 14A

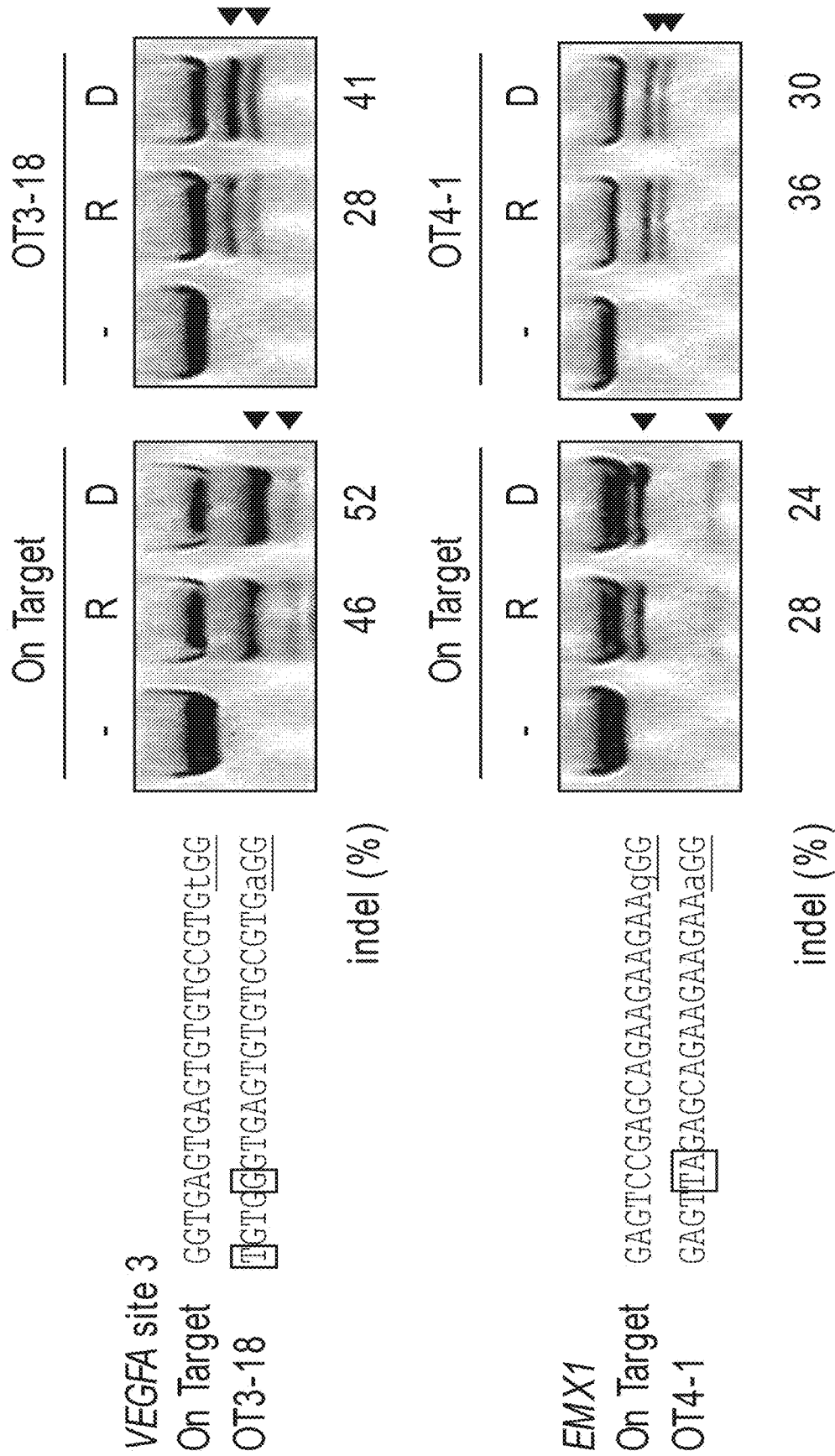


FIG. 14B

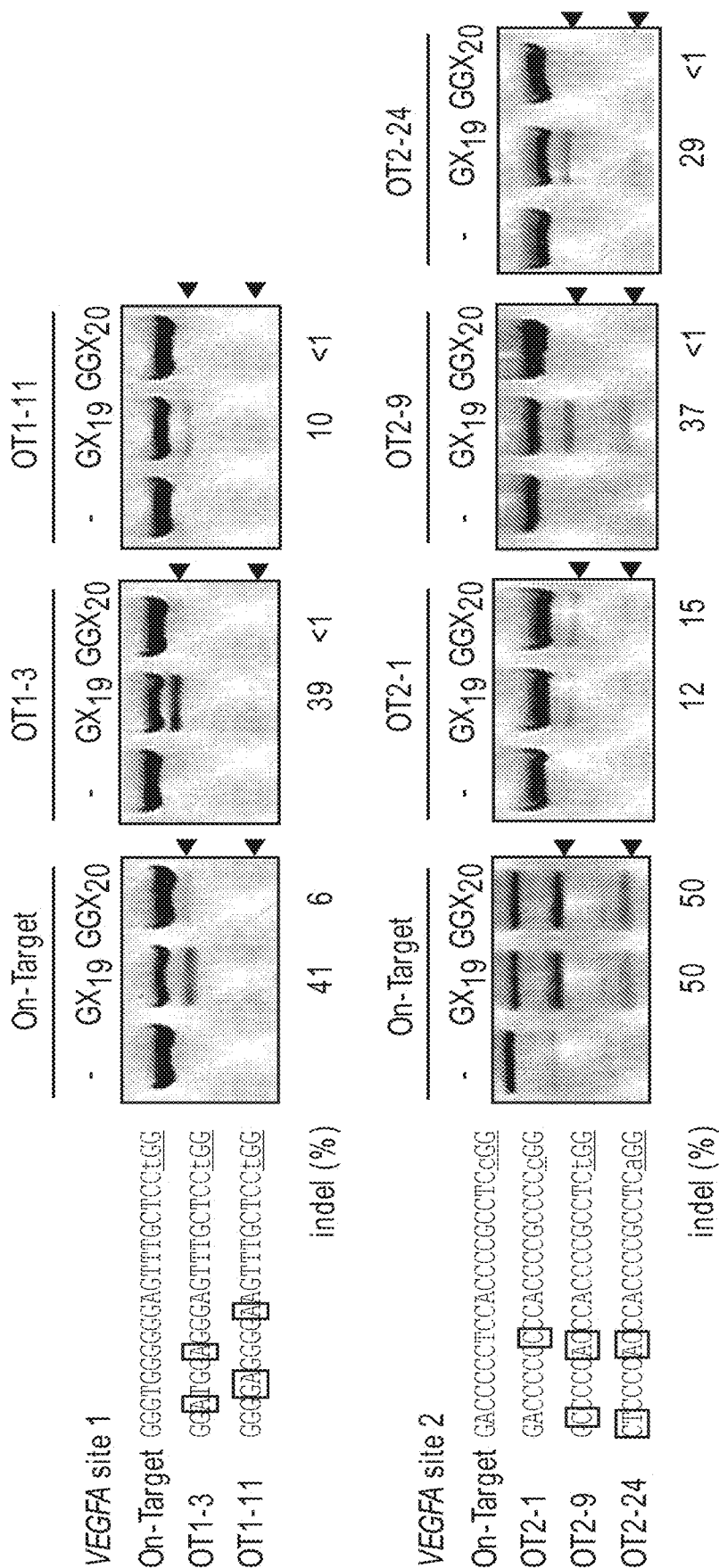


FIG. 15A

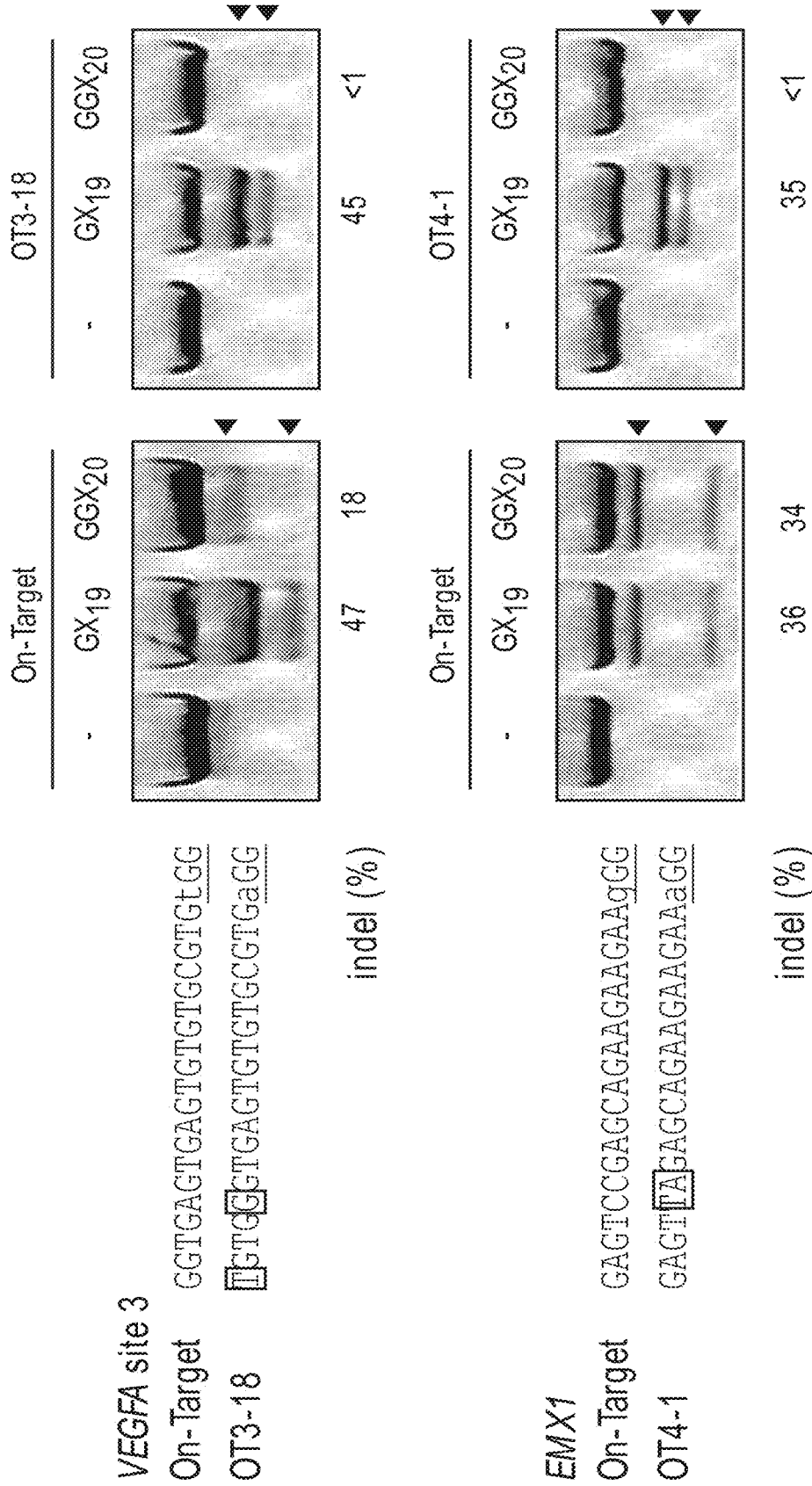


FIG. 15B

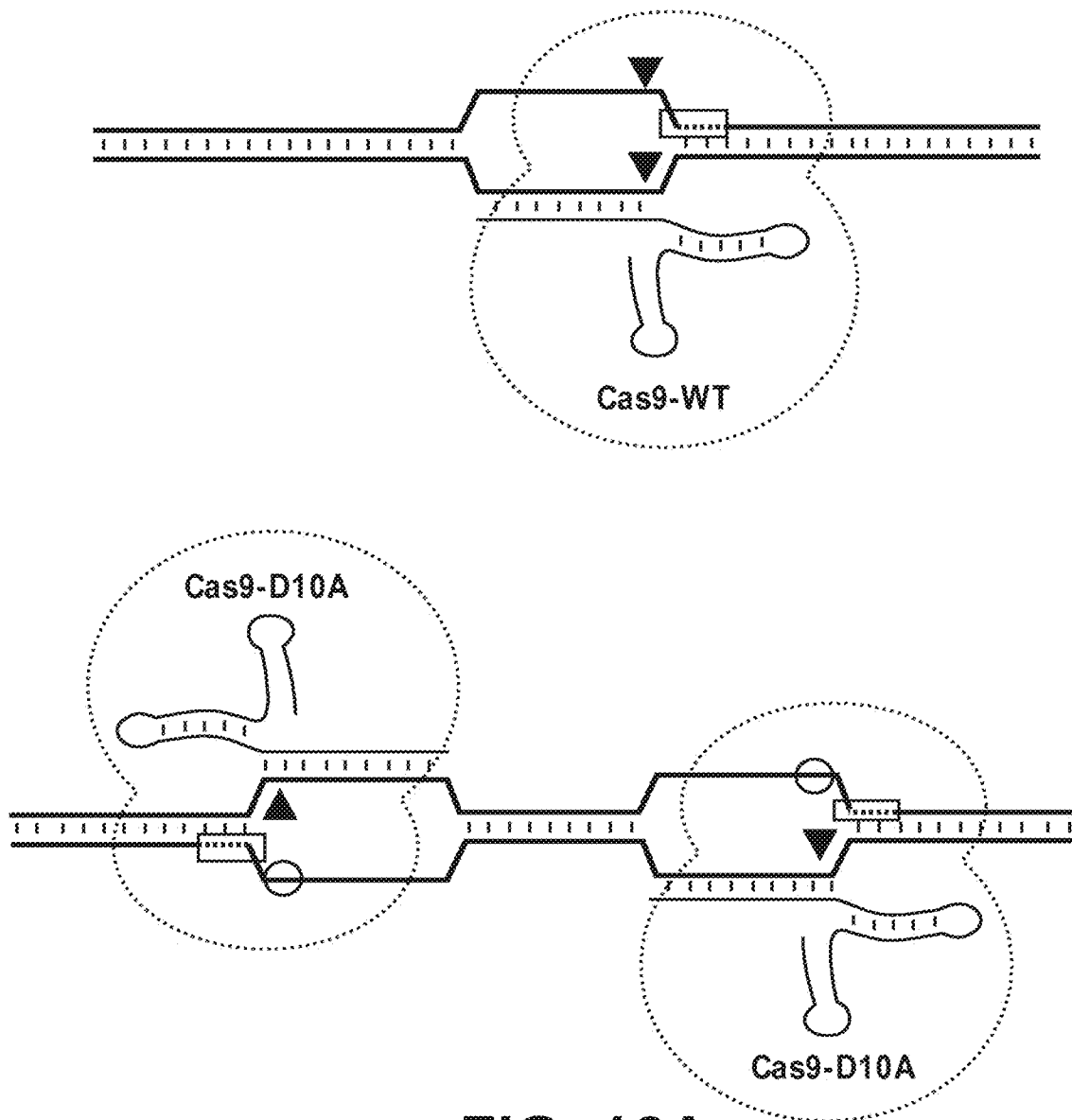


FIG. 16A

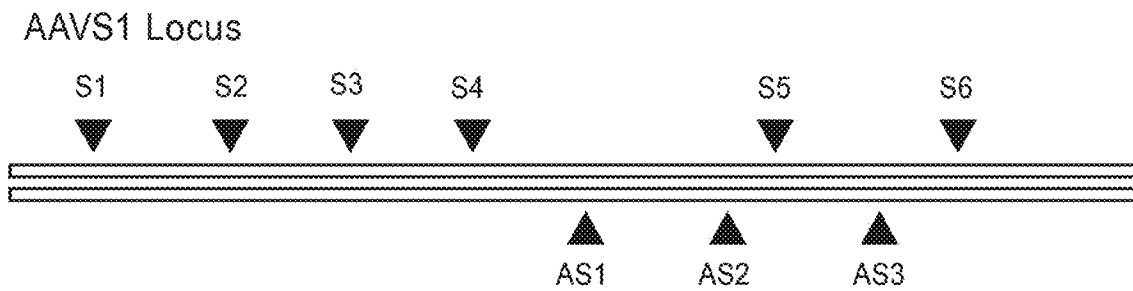


FIG. 16B

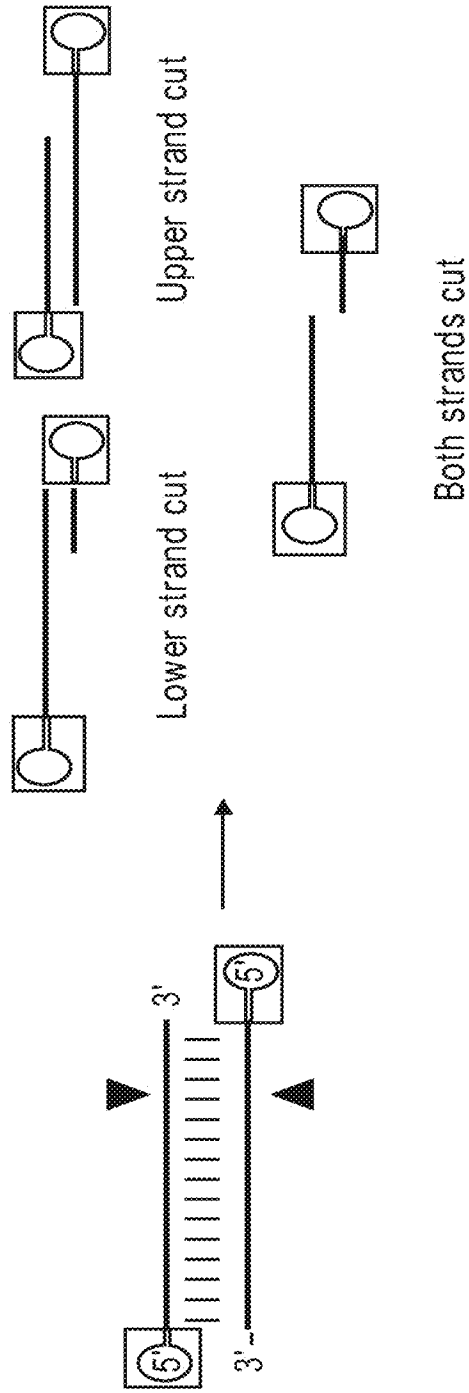


FIG. 16C

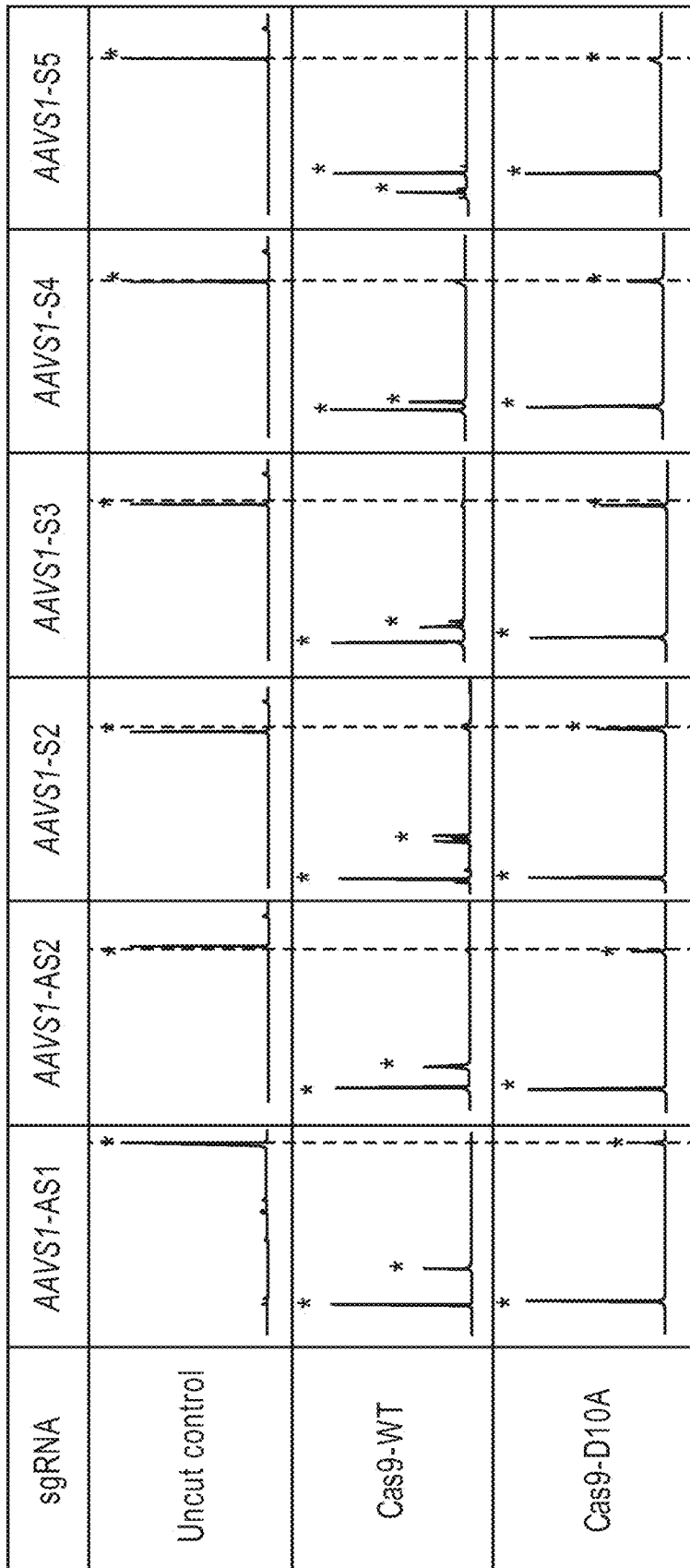


FIG. 16D

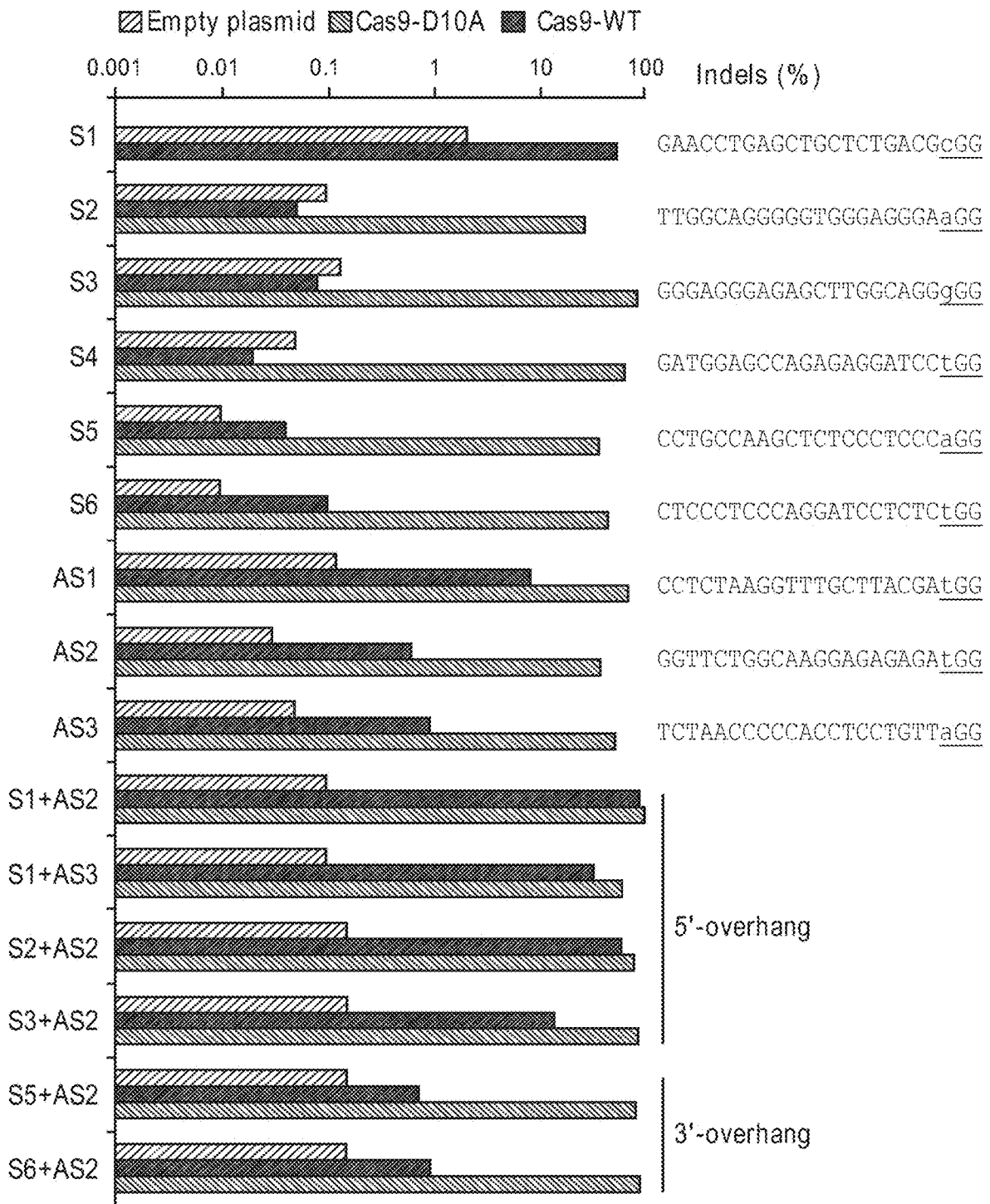


FIG. 17A

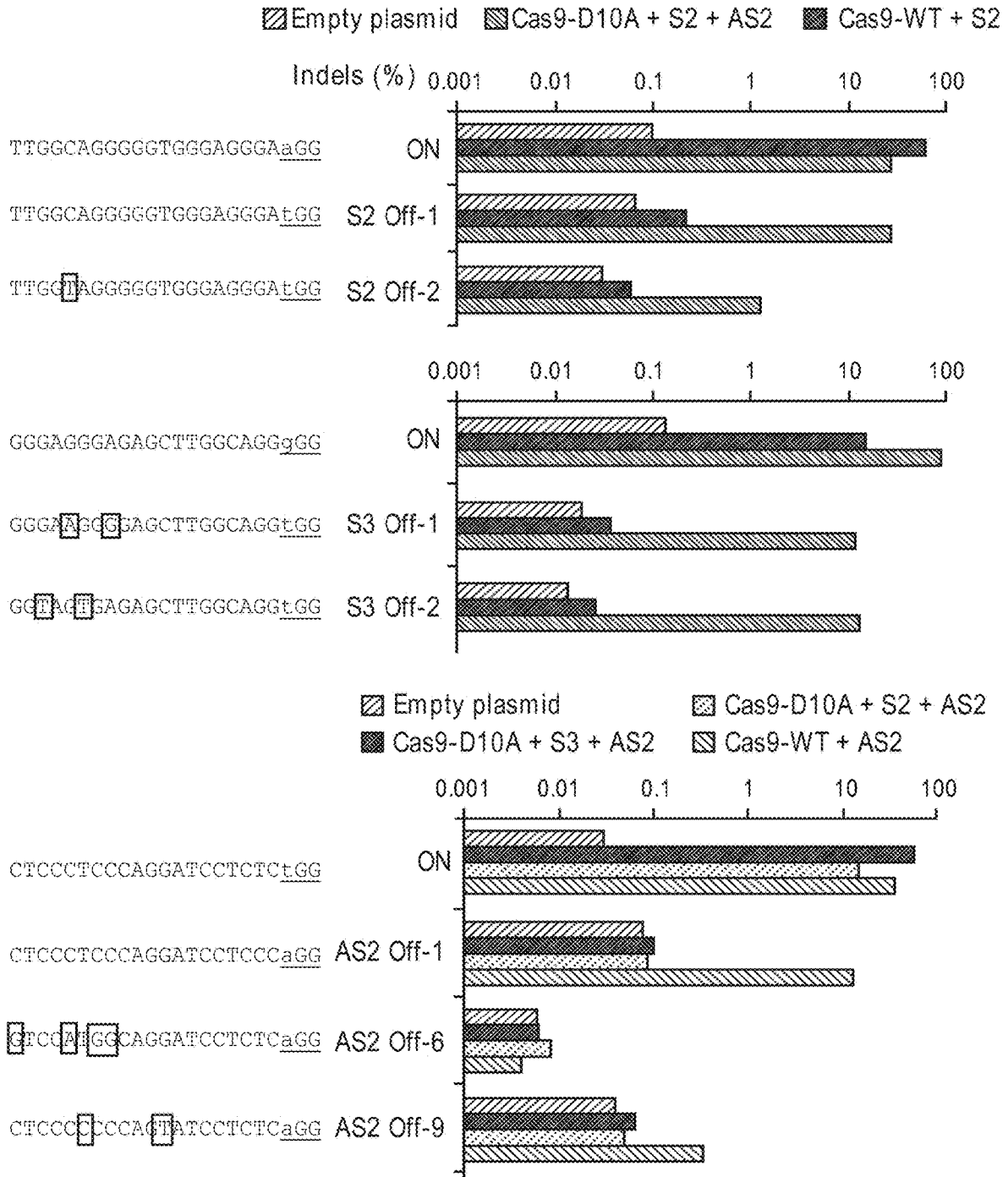


FIG. 17B

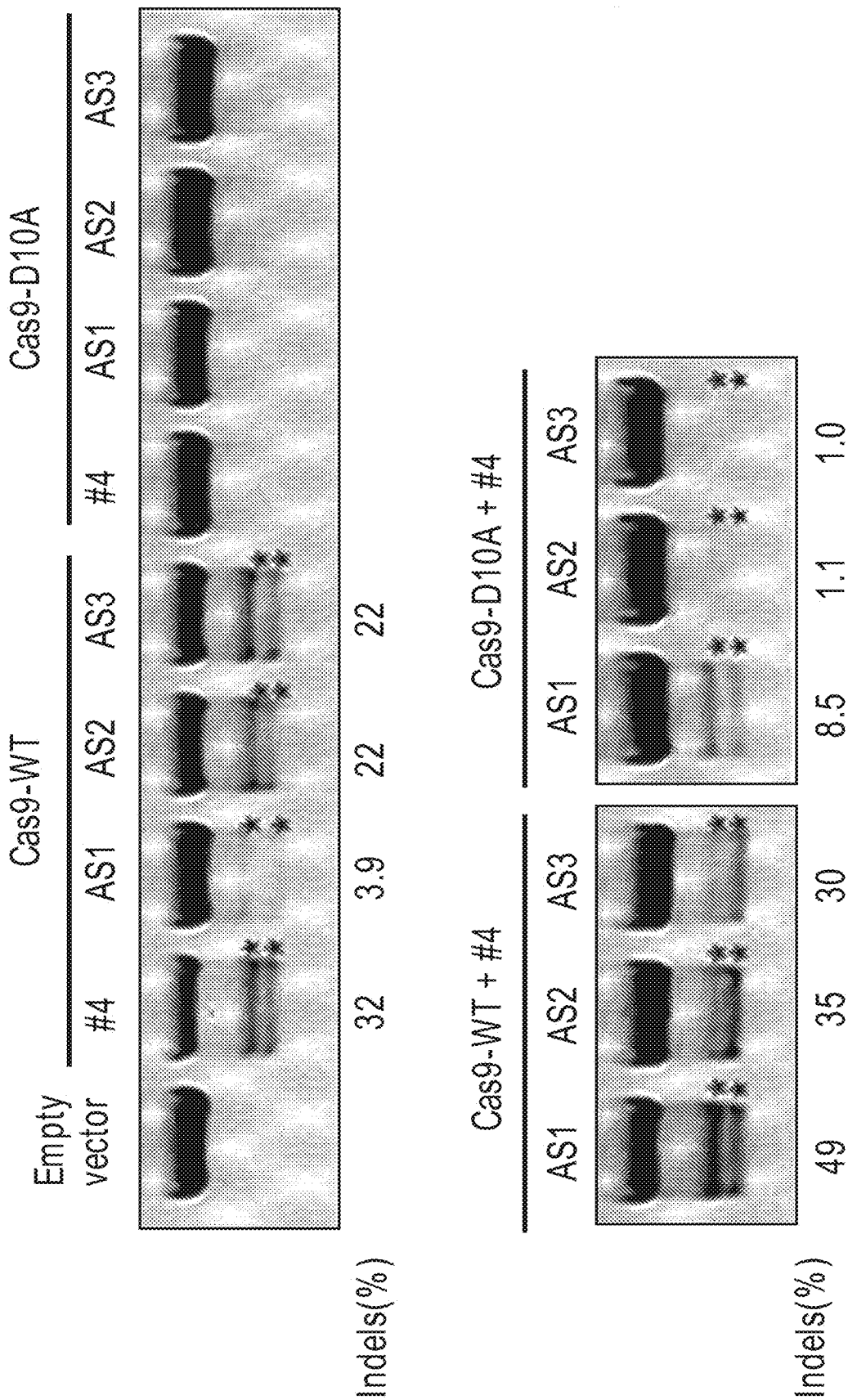


FIG. 18B

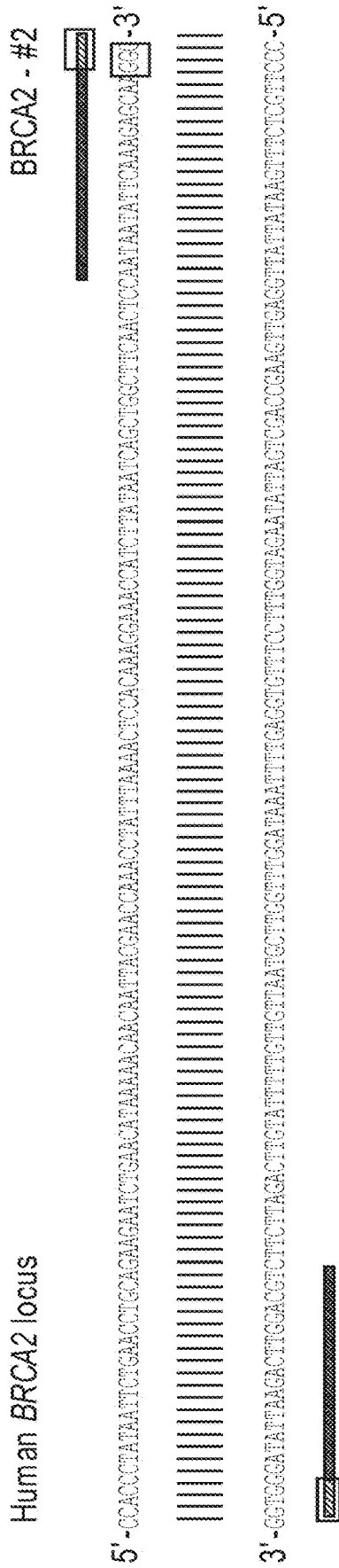


FIG. 18C

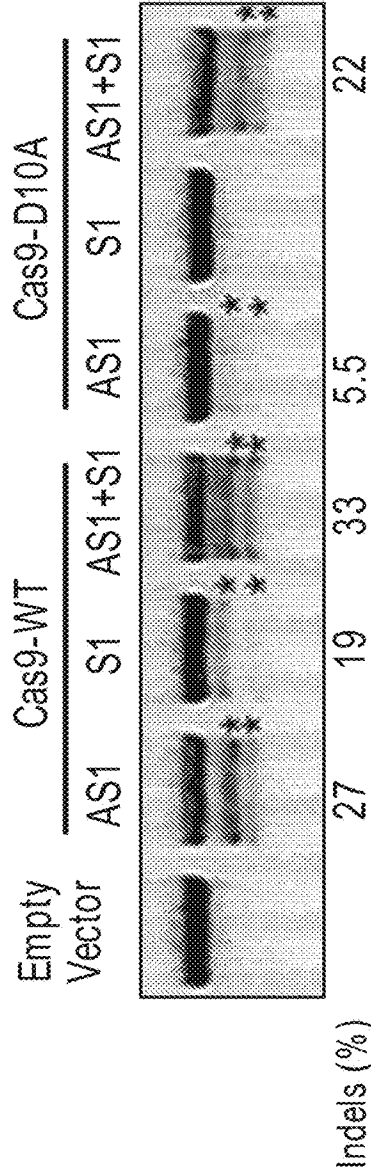


FIG. 18D

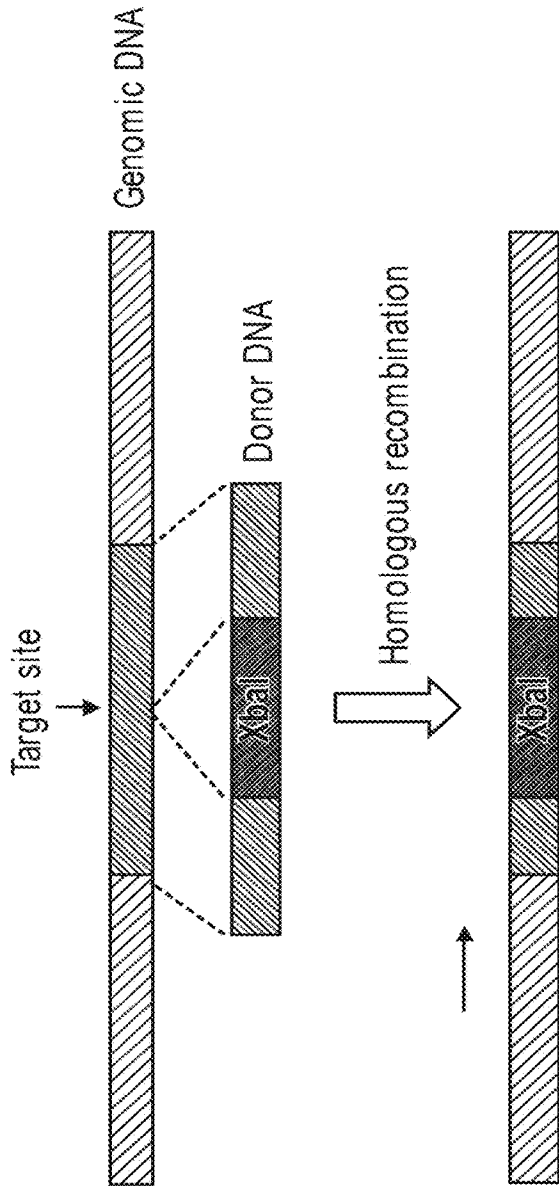


FIG. 19A

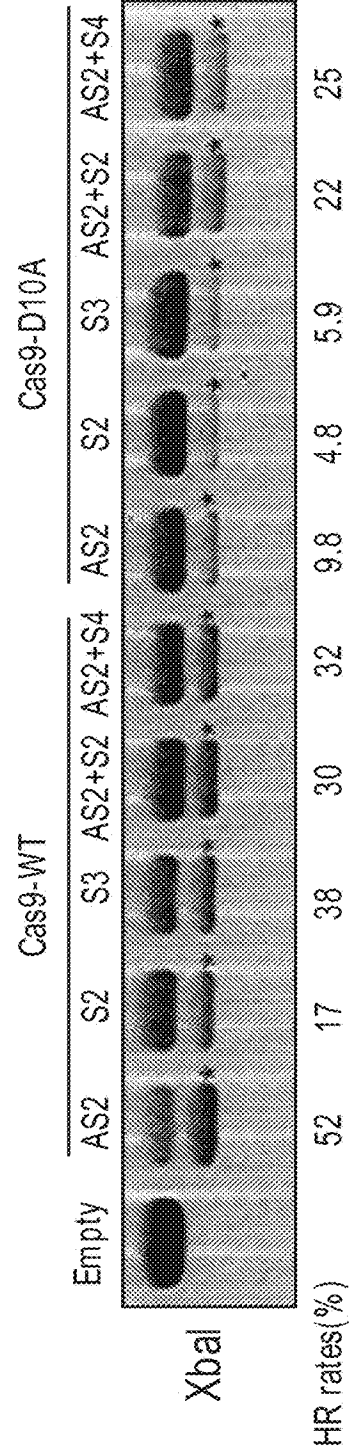


FIG. 19B

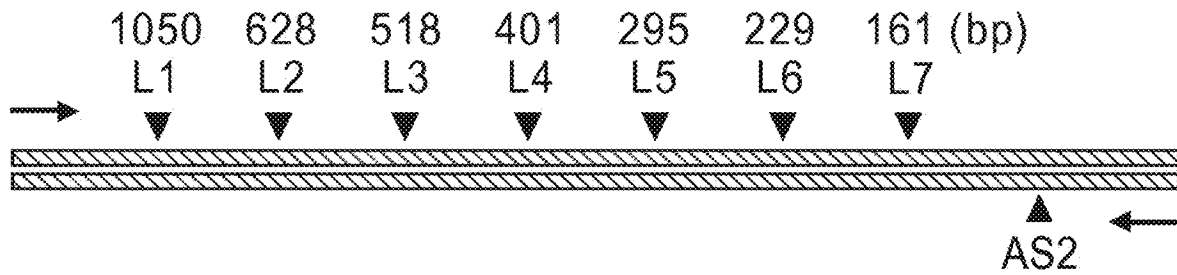


FIG. 20A

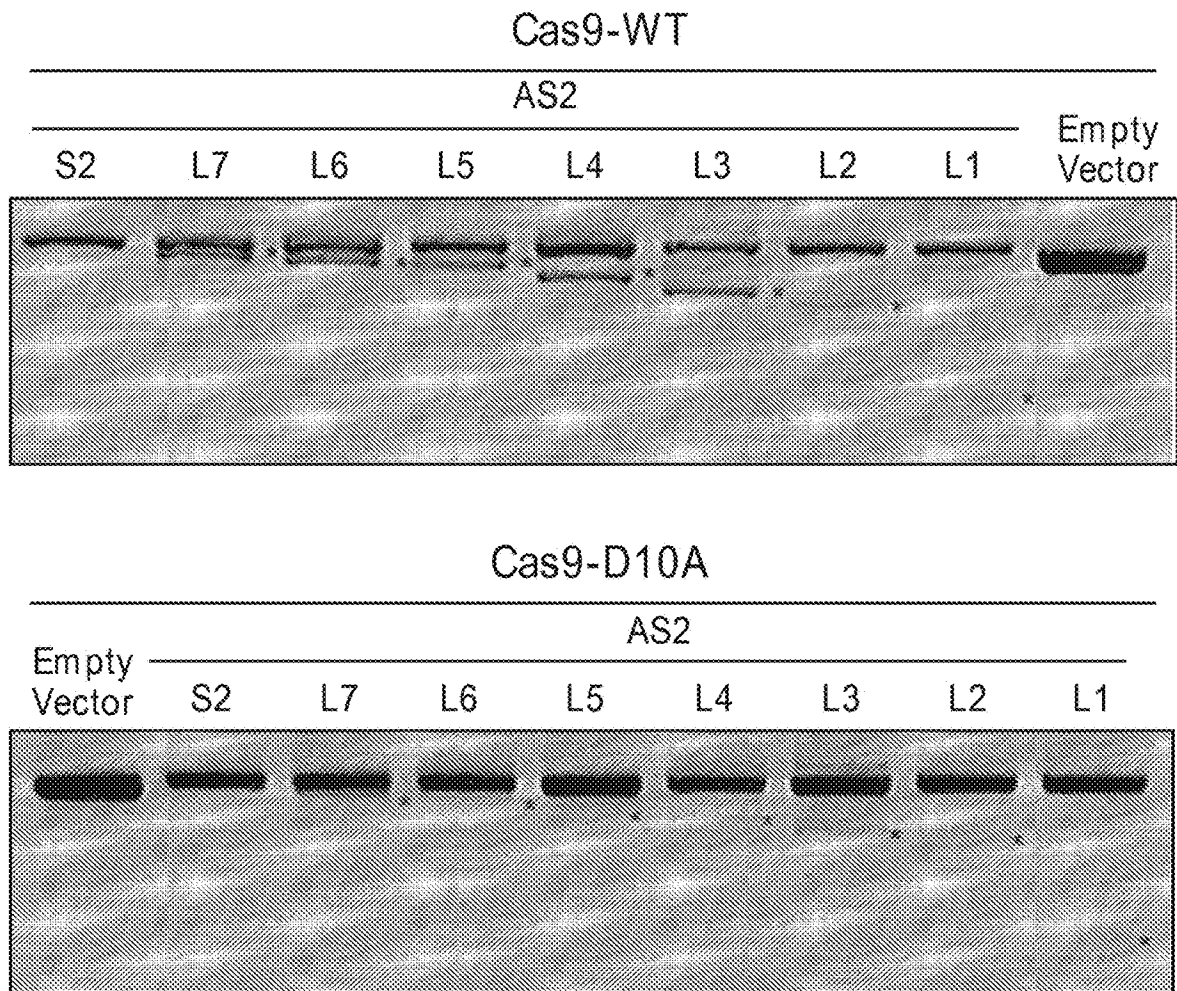


FIG. 20B

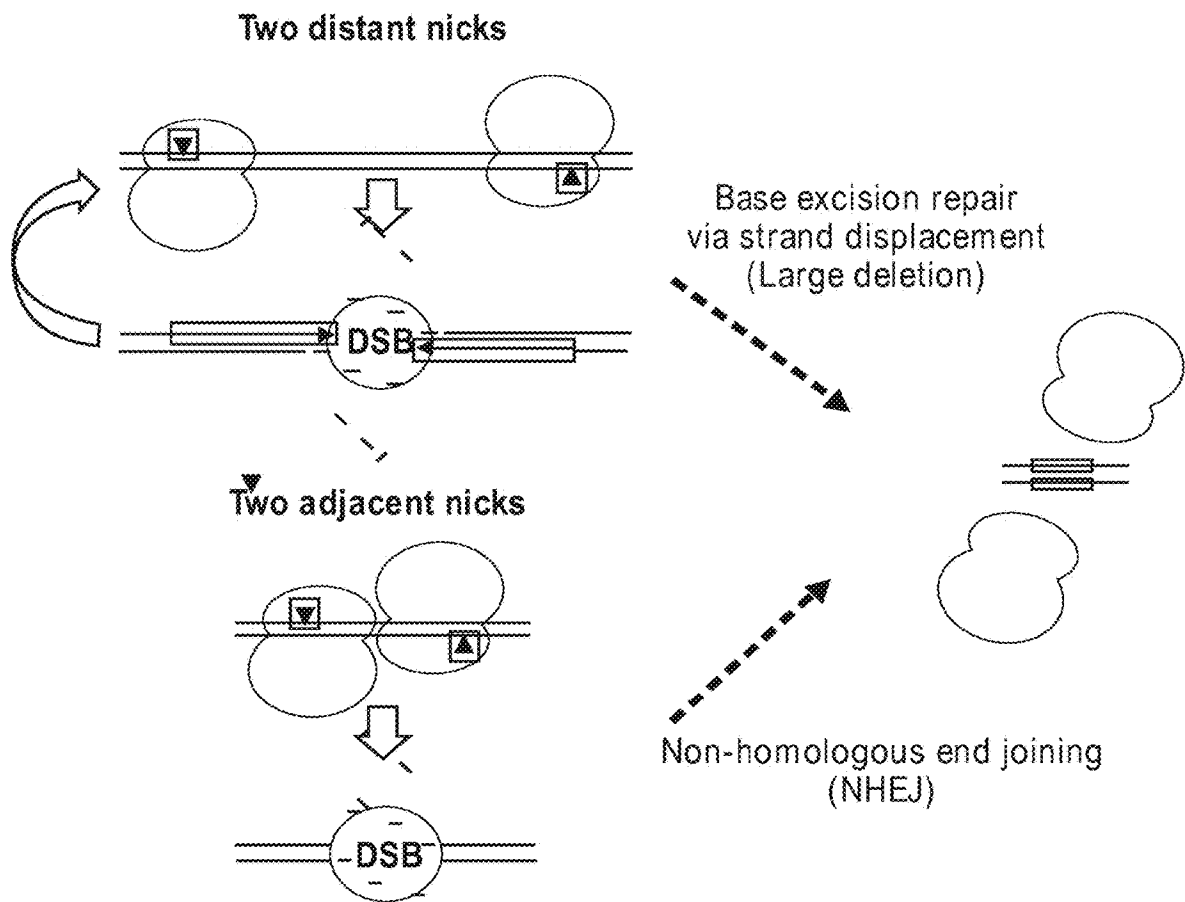


FIG. 20D

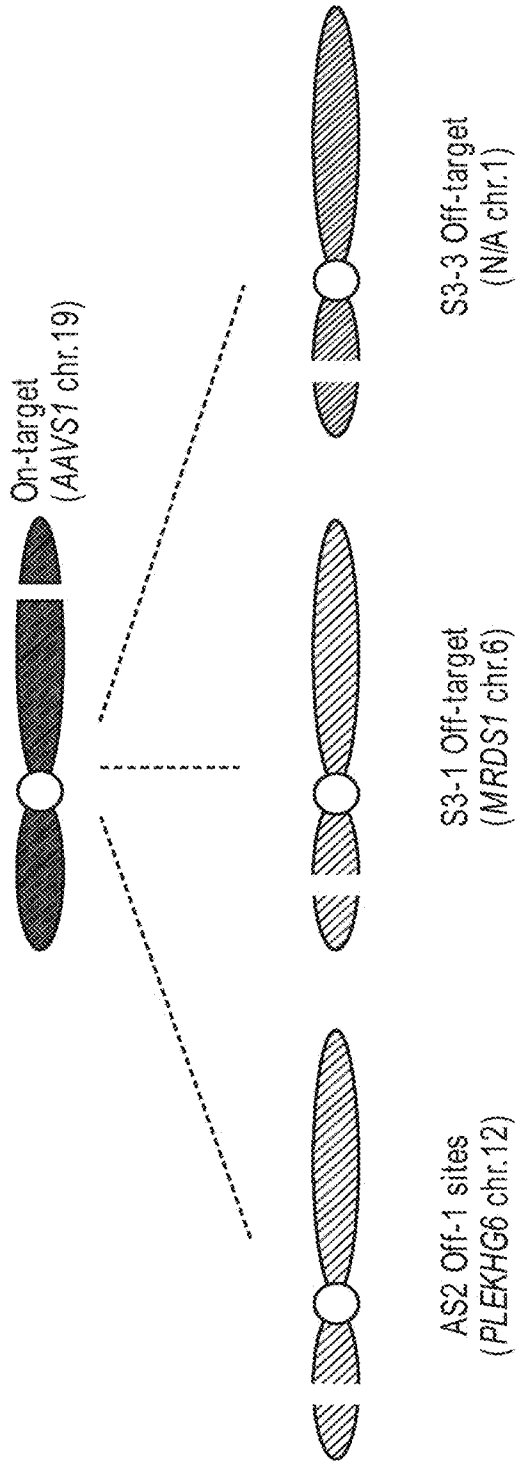


FIG. 21A

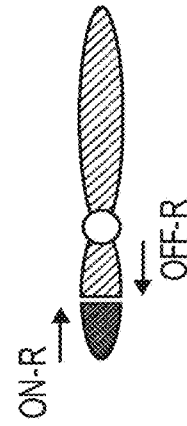
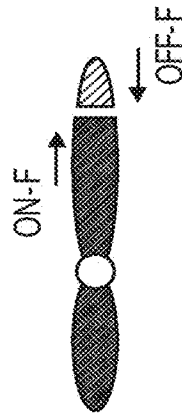


FIG. 21B

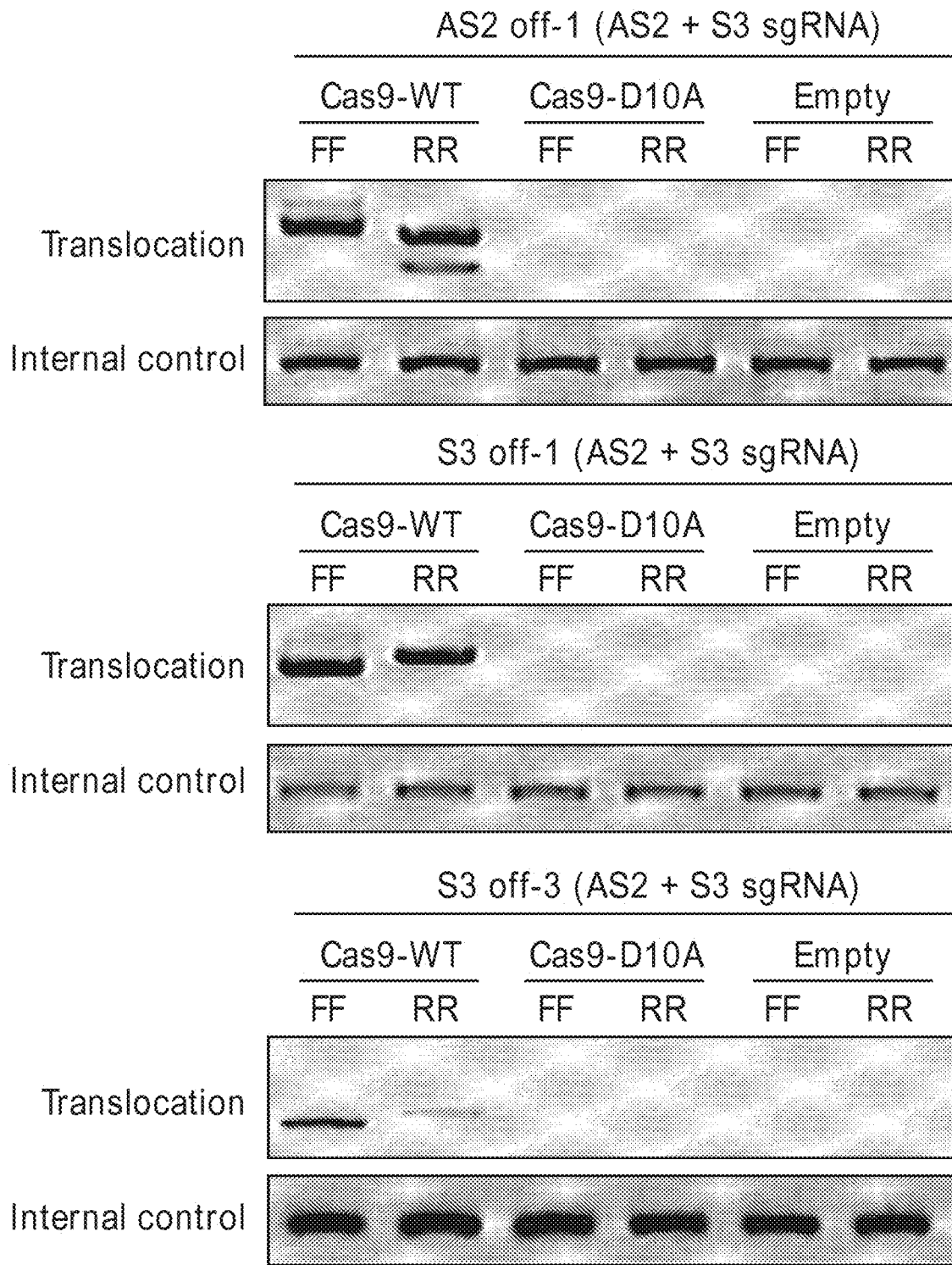


FIG. 21C

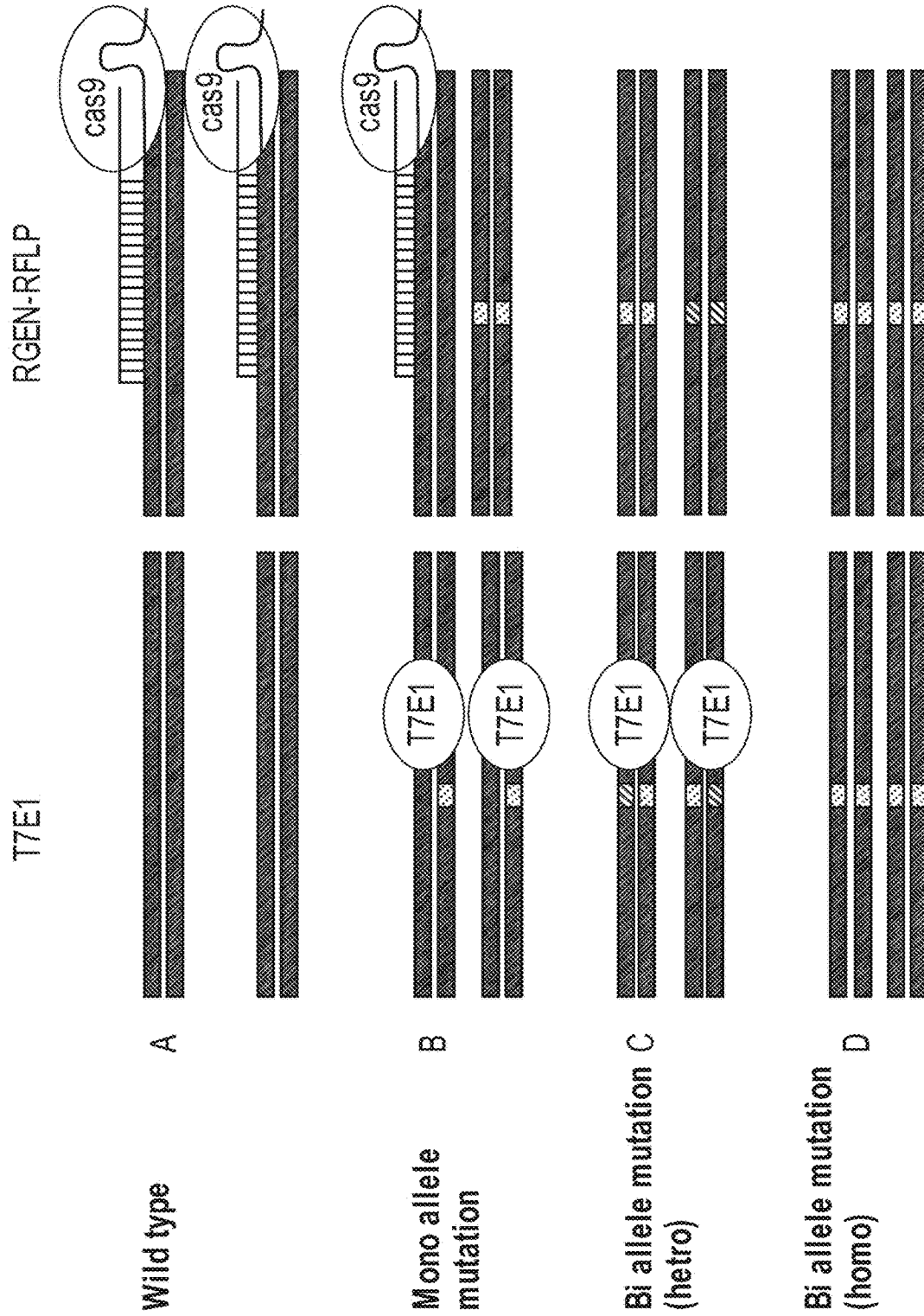


FIG. 22A

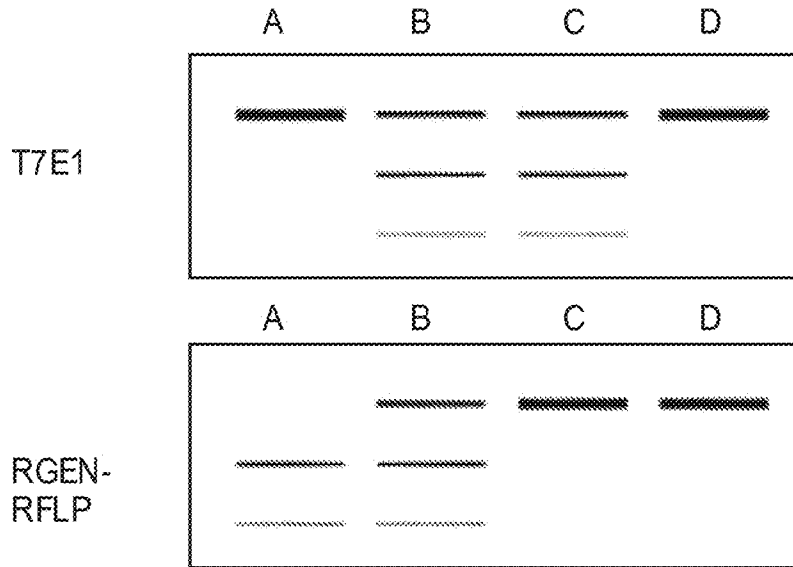


FIG. 22B

Plasmid target sequence

AATGACCACTACATCCT---	CAAGGG	WT	
AATGACCACTACATCCT	<u>T</u> ---	CAAGGG	I1
AATGACCACTACATCCT	<u>TT</u> ---	CAAGGG	I2
AATGACCACTACATCCT	<u>TTT</u>	CAAGGG	I3
AATGACCACTACATCCT----	AAGGG	D1	
AATGACCACTACATCCT-----	AGGG	D2	
AATGACCACTACATCCT-----	GGG	D3	

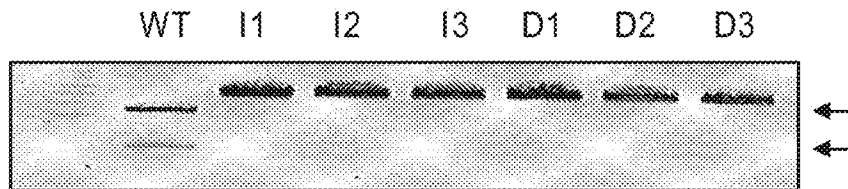


FIG. 23

#1 (+/-)

TATGTGC	AATGACCACTACATCCT	---	CAA	GGGCAGCAATCGGAG	WT
TATGTGC	AATGACCACTACATCCT	CCT	CAA	GGGCAGCAATCGGAG	+3

#2 (+/-)

TATGTGC	AATGACCACTACATCCT	CAA	GGGCAGCAATCGGAG	WT
TATGTGC	AATGACCACTACATC	-----	AATCGGAG	-12

#5 (+/-)

TATGTGC	AATGACCACTACATCCT	CAA	GGGCAGCAATCGGAG	WT
TATGTGC	AATGACCACTACAT	-----	CAGCAATCGGAG	-9

#6 (+/-)

TATGTGC	AATGACCACTACATCCT	CAA	GGGCAGCAATCGGAG	WT
TATGTGC	AATGACCACTACATCC	-----	AGCAATCGGAG	-8

#12 (-/-)

-----	-----	-----	CAGCAATCGG	-36	
TATGTGC	AATGACCACTACATCCT	-----	CAA	GGGCAGCAATCGG	+1
TATGTGC	AATGACCACTACATCCT	-----	CAA	GGGCAGCAATCGG	+1
TATGTGC	AATGACCACTACATCCT	/67bp/	CAA	GGGCAGCAATCGG	+67

#28 (-/-)

TATGTGC	AATGACCACTACATCCT	CAA	GGGCAGCAATCGG	+1
TATGTGC	AATGACCACTACAT	---	GGCAGCAATCGG	-7, +1
TATGTGC	AATGACCACTACAT	-----	-----	-94

FIG. 24A

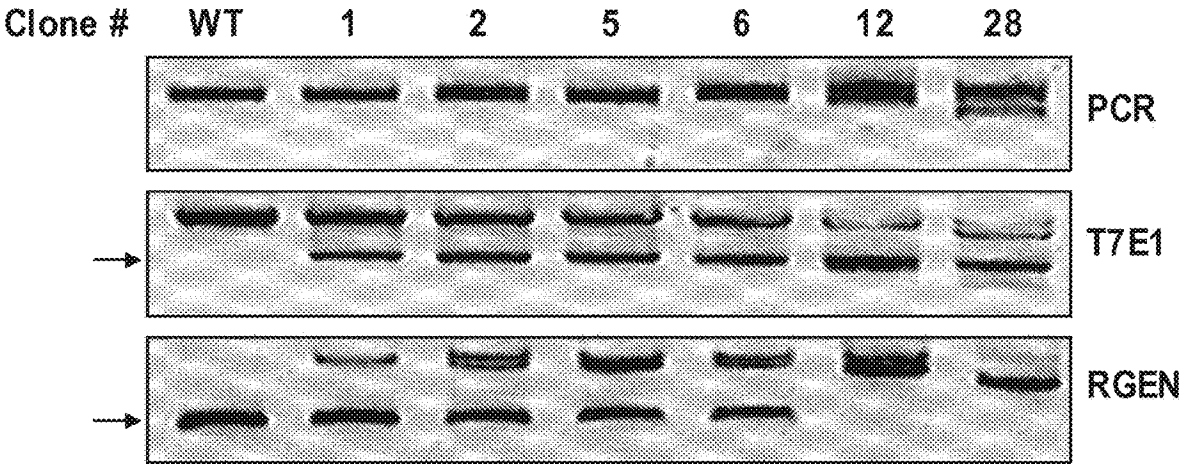


FIG. 24B

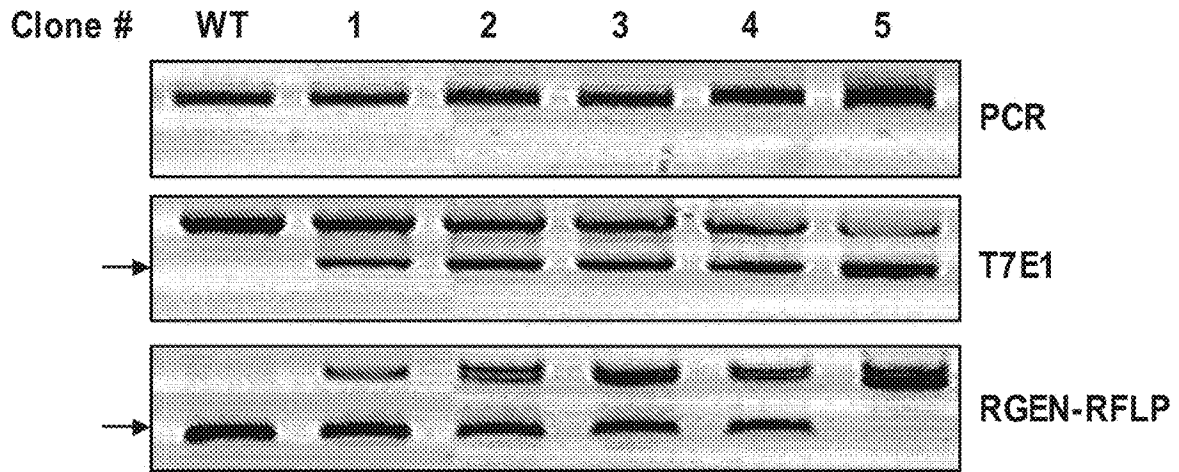


FIG. 25A

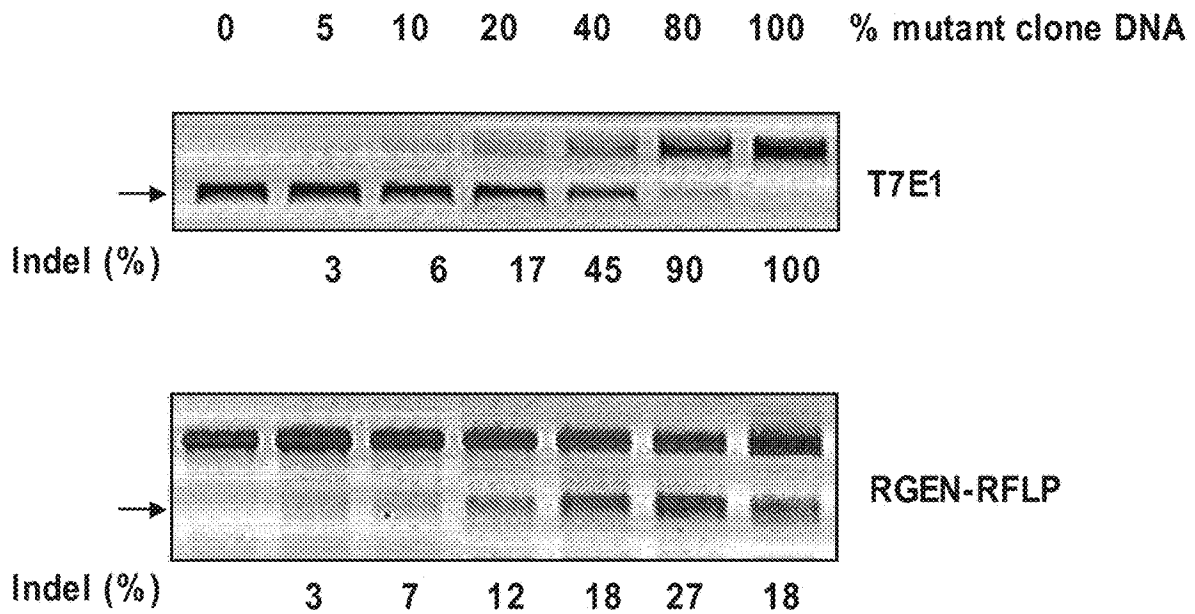


FIG. 25B

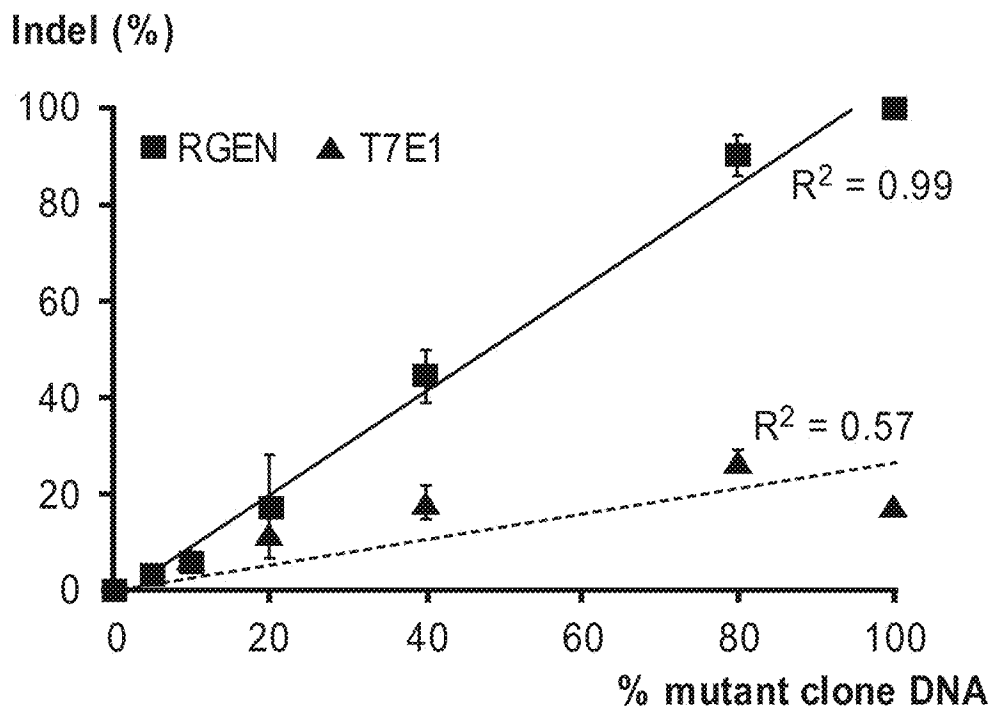
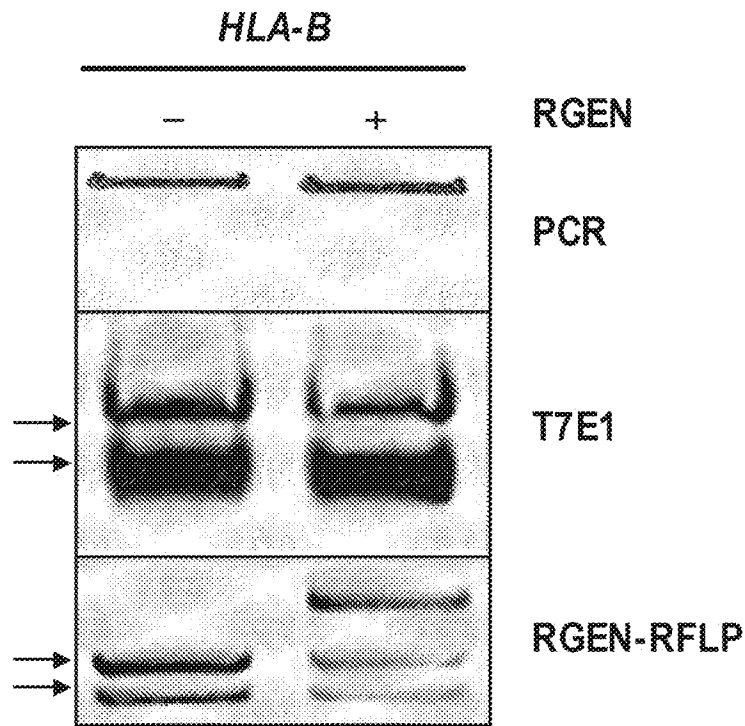


FIG. 25C

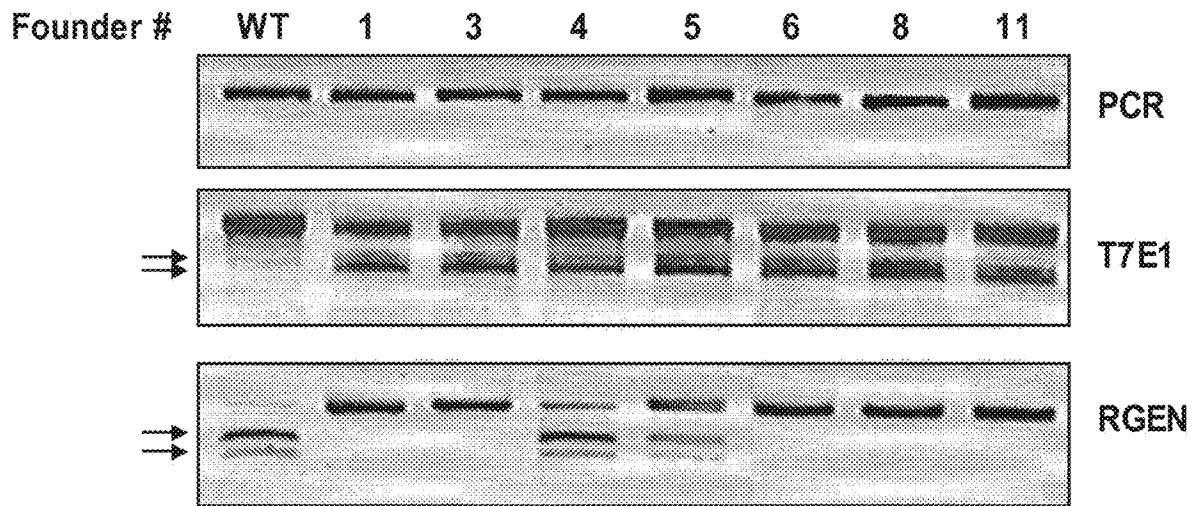


FIG. 26B

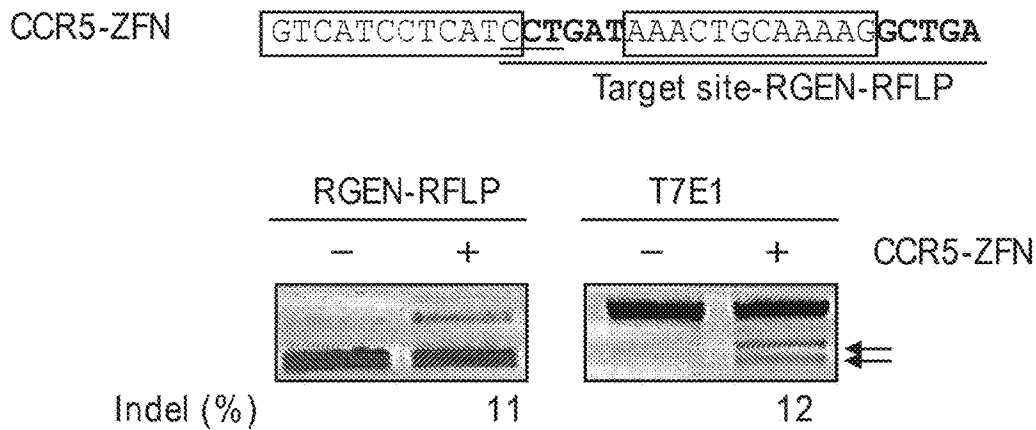


FIG. 27

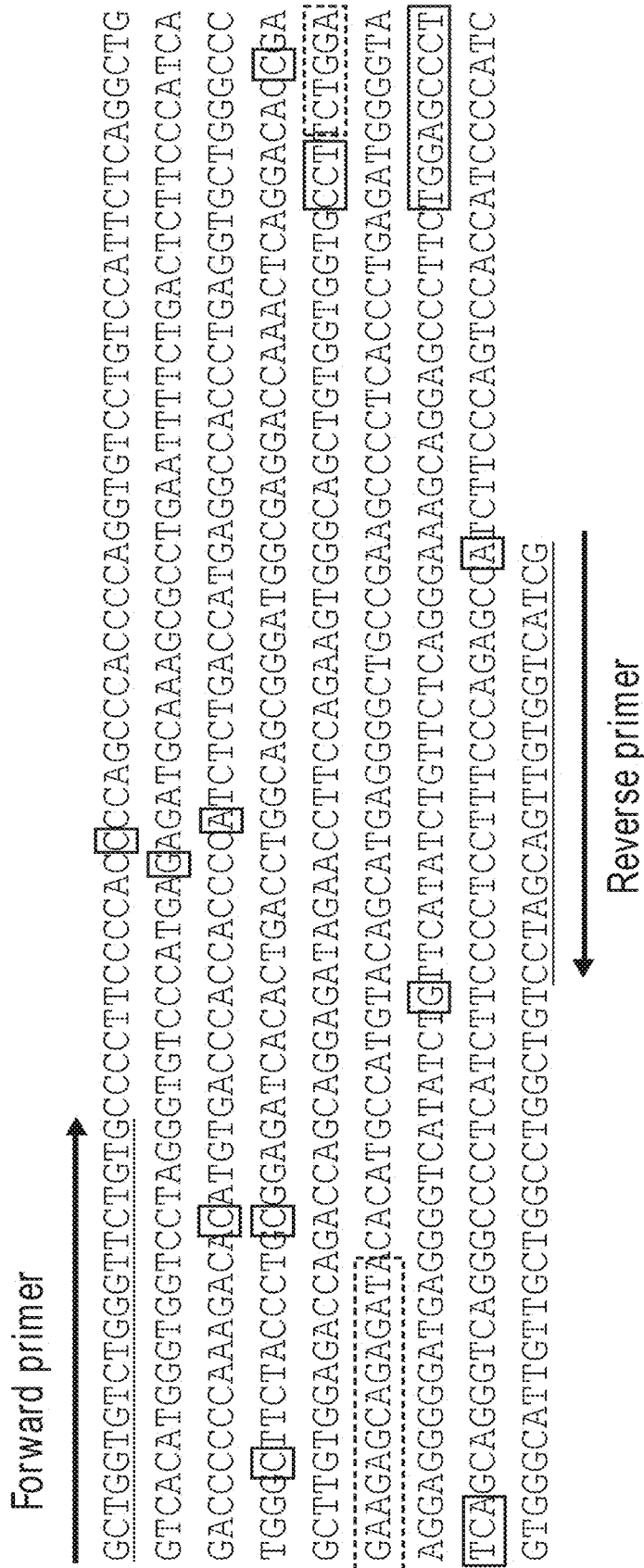
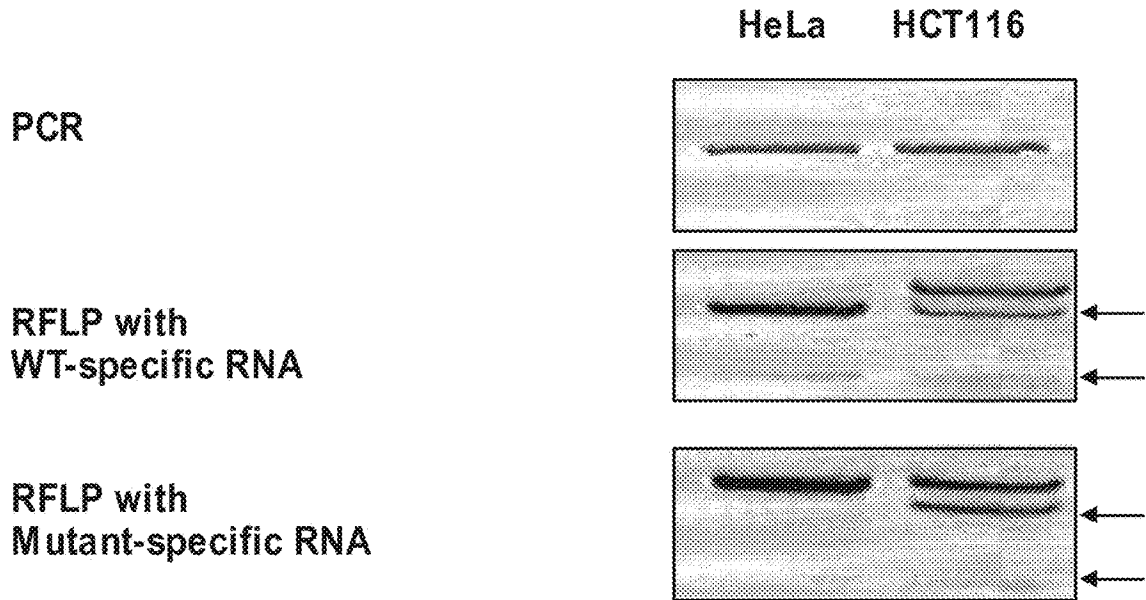


FIG. 28



HeLa

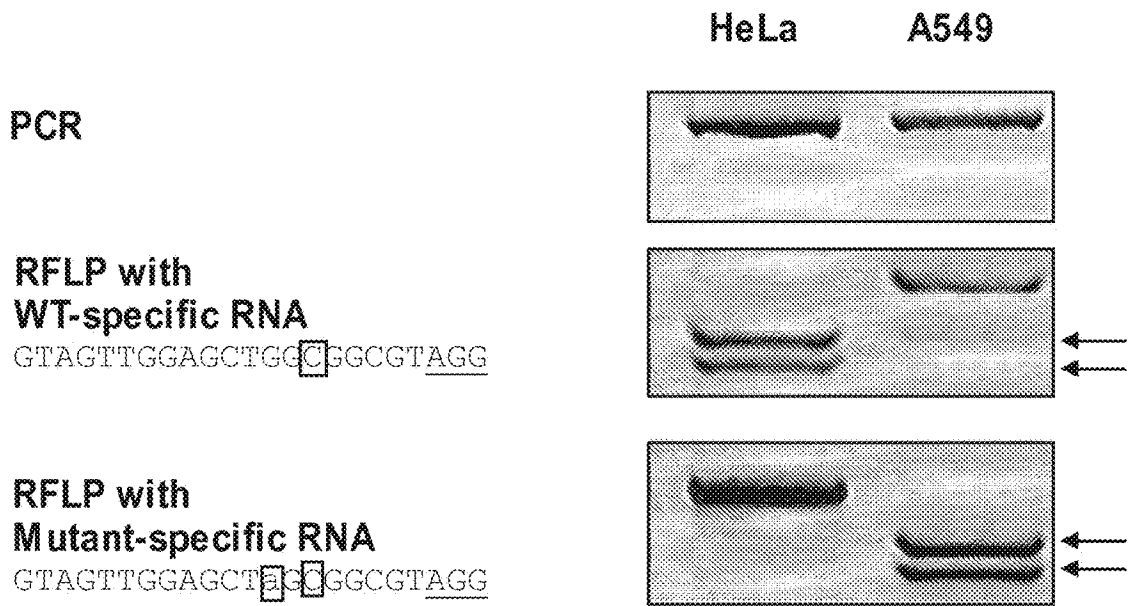
ACTACCACAGCTCCTTCTCTGAGTGG wild-type

HCT116

ACTACCACAGCTCCTTCTCTGAGTGG wild-type

ACTACCACAGCTCCT---CTGAGTGG c.133-135 del TCT

FIG. 29A



HeLa

GTAGTTGGAGCTGGTGGCGTAGG wild-type

A549

GTAGTTGGAGCTGGTGGCGTAGG c.34G>A

FIG. 29B

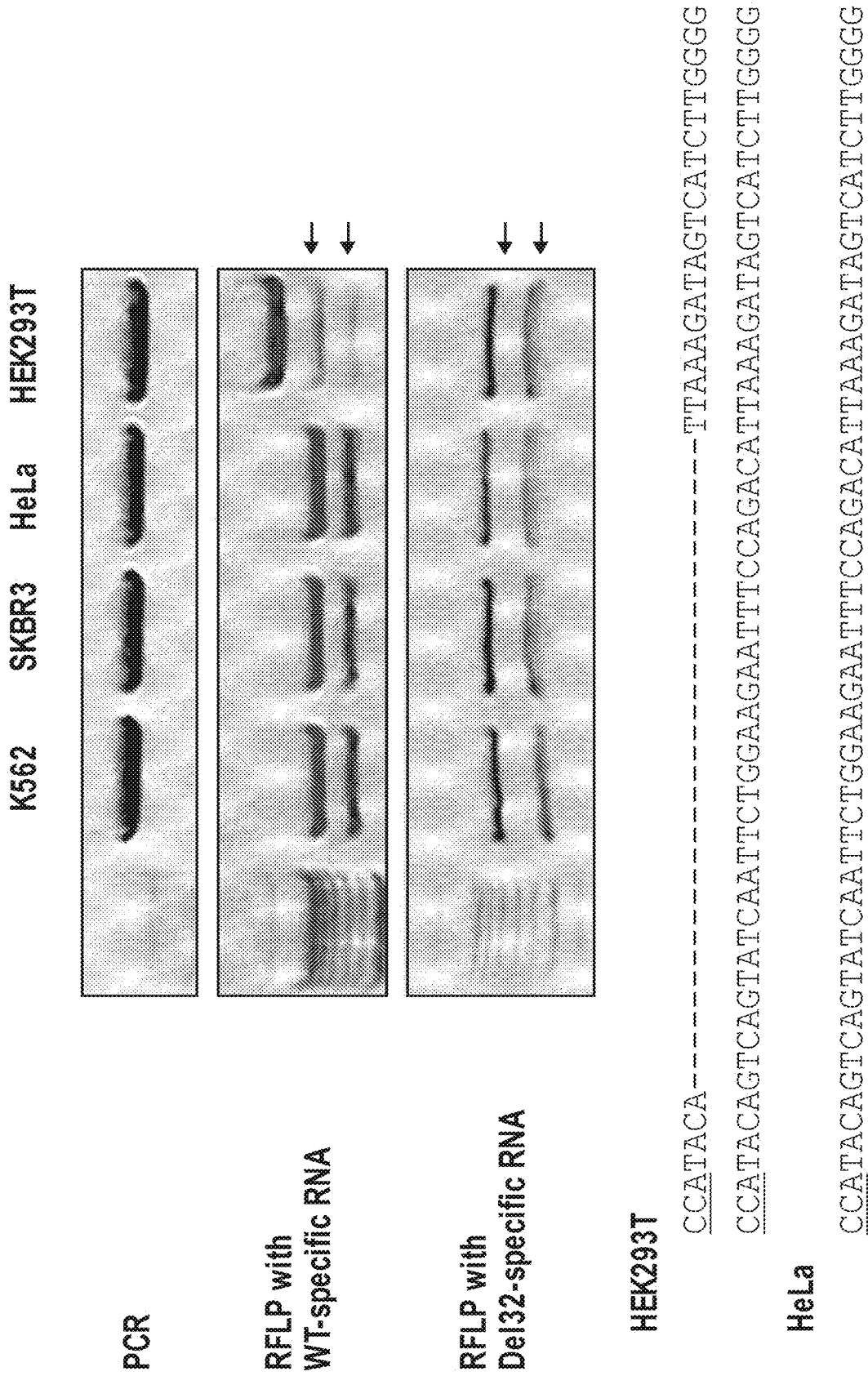


FIG. 30A

CCR5 WT
 ccatacagtcagtatcaatcttggaagaatttccagacattaaagatagtcattct

CCR5 WT target site

CCR5 Δ 32
 ccataca-----ttaagatagtcattct

CCR5 Δ 32 on-target site

CCR5 Δ 32 off-target site

FIG. 30B

WT Δ 32 Linearized plasmid

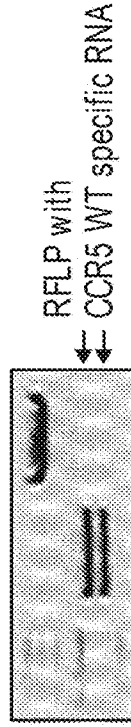
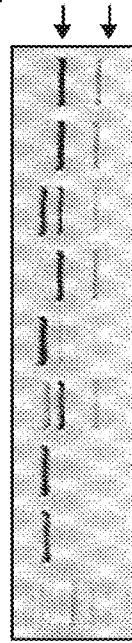


FIG. 30C

181 127 1/9
 on off on off on off on off
 Dilution (Cas9 0.1ug/crRNA 25ng / tracrRNA 60ng)
 Plasmid



On agatgactatctttaatgctcgg
 Off agatgactatctttaatgctcgg

FIG. 30D

KRAS

HeLa

GTAGTTGGAGCTGGTGGCGTAGG Wild-type

A549

GTAGTTGGAGCTaGTGGCGTAGG c.34G>A

**RFLP with
WT-specific RNA**
GTAGTTGGAGCTGGTGGCGTAGG

**RFLP with
Mutant-specific RNA**
GTAGTTGGAGCTaGTGGCGTAGG

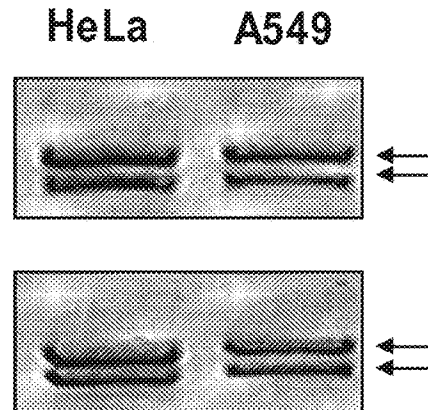


FIG. 31A

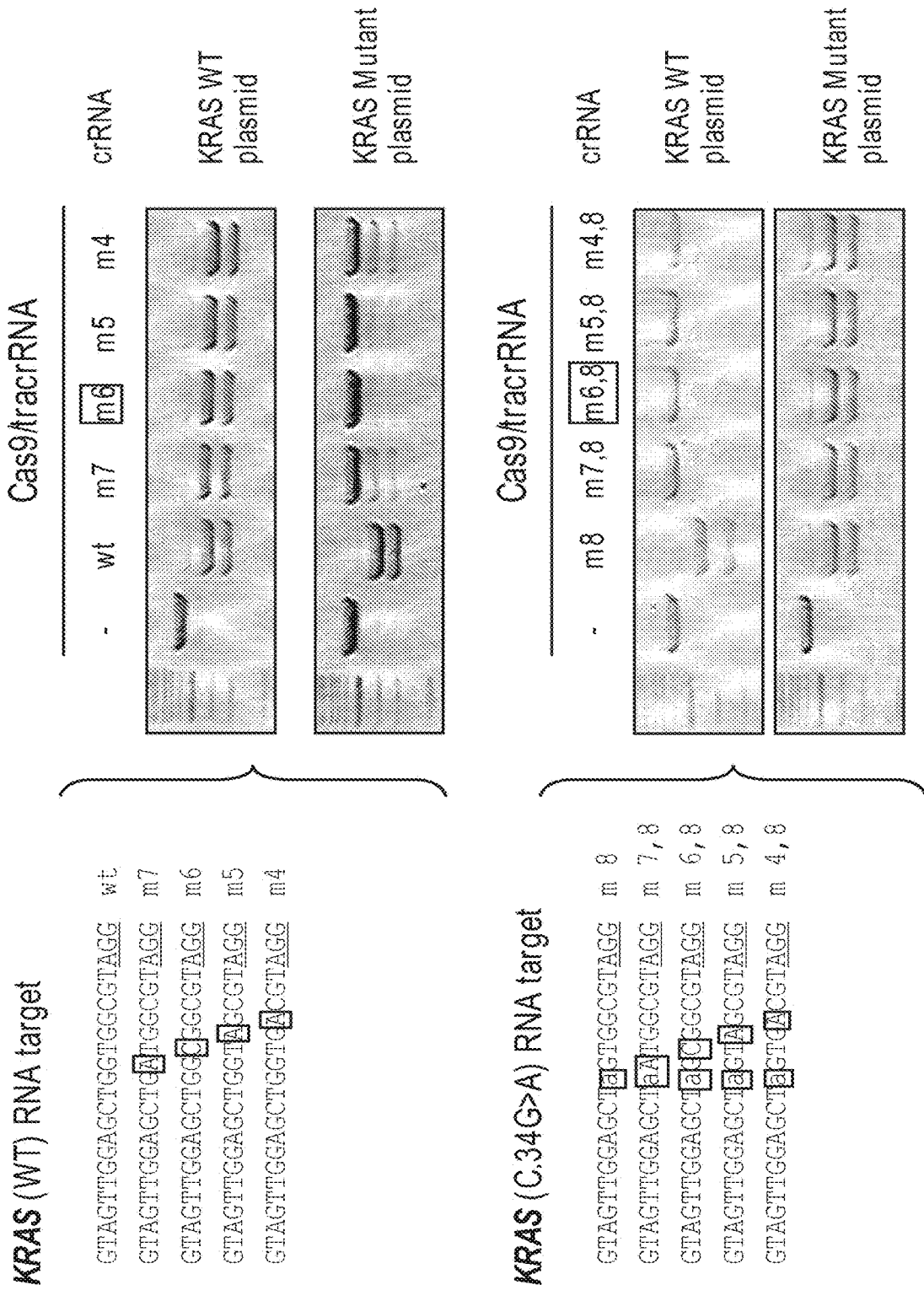


FIG. 31B

PIK3CA

HeLa

CAAATGAATGATGCACATCATGG Wild-type

HCT116

CAAATGAATGATGCACATCATGG Wild-type

CAAATGAATGATGCACAGTCATGG C.3140A>G

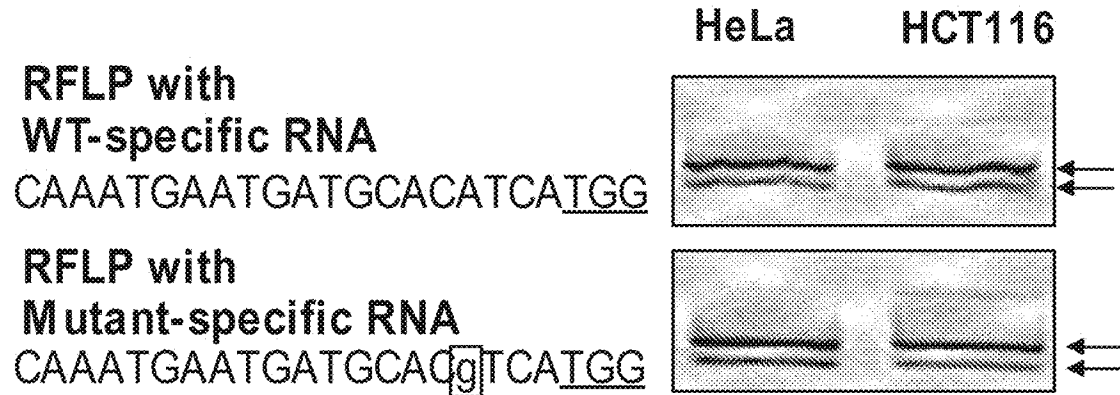


FIG. 32A

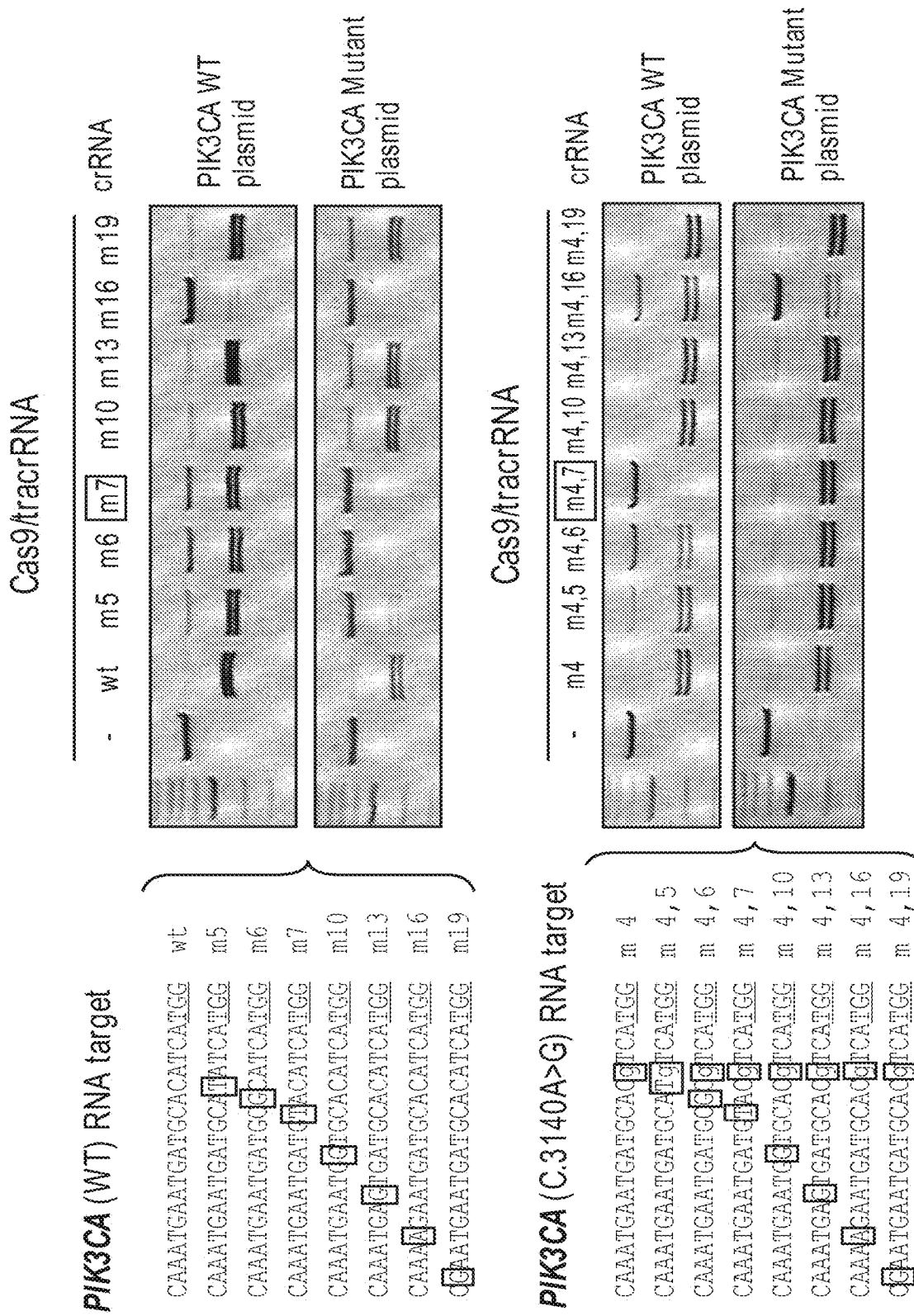


FIG. 32B

IDH1

HeLa

ATCATAGGTCGTCATGCTTATGG Wild-type

HT1080

ATCATAGGTCGTCATGCTTATGG Wild-typ

ATCATAGGTGTCATGCTTATGG c.394C>T

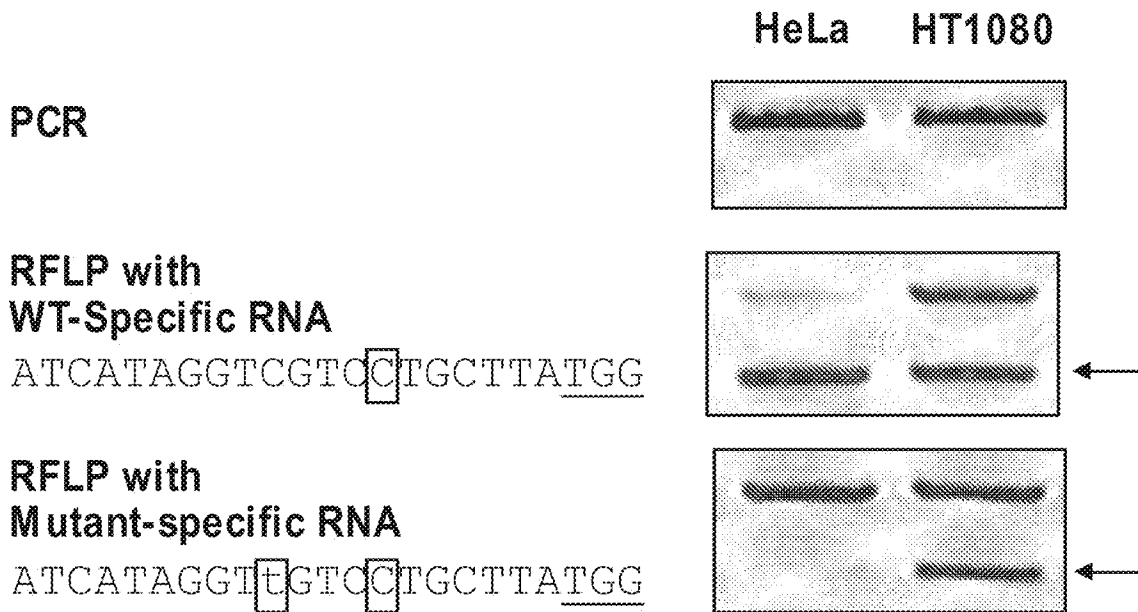


FIG. 33A

PIK3CA

HeLa

CAAATGAATGATGCACATCATGG Wild-type

HCT116

CAAATGAATGATGCACATCATGG Wild-type

CAAATGAATGATGCACgTCATGG C.3140A>G

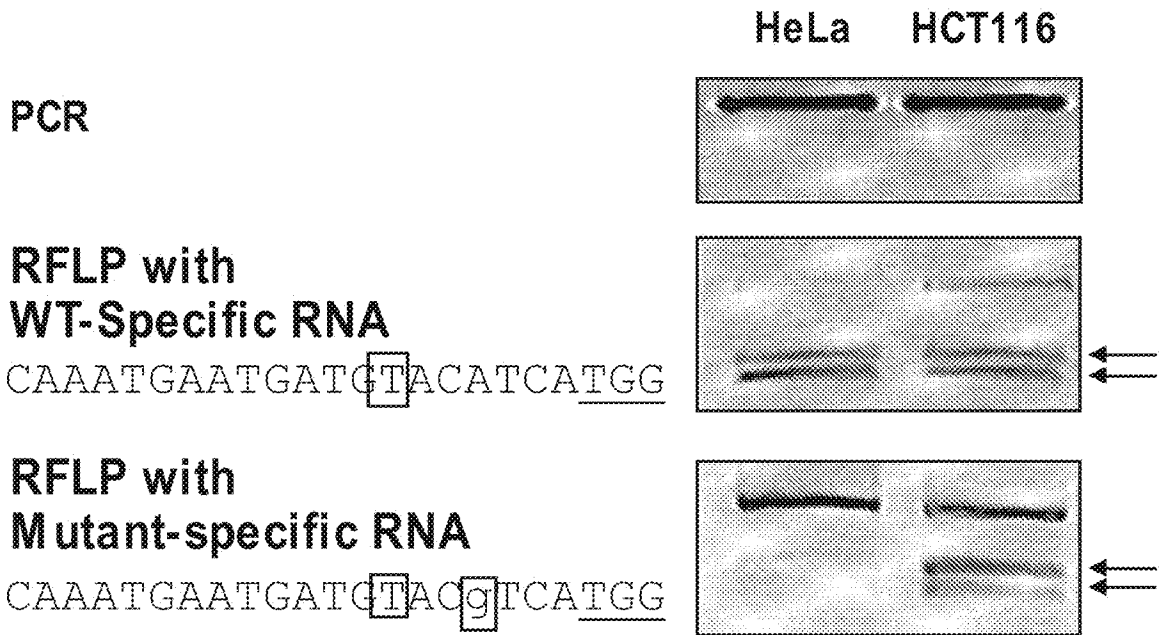


FIG. 33B

NRAS

HeLa

CTGGACAAGAAGAGTACAGTGCC Wild-type

HT1080

CTGGACAAGAAGAGTACAGTGCC Wild-type

CTGGAA**a**AAGAAGAGTACAGTGCC c.181C>A

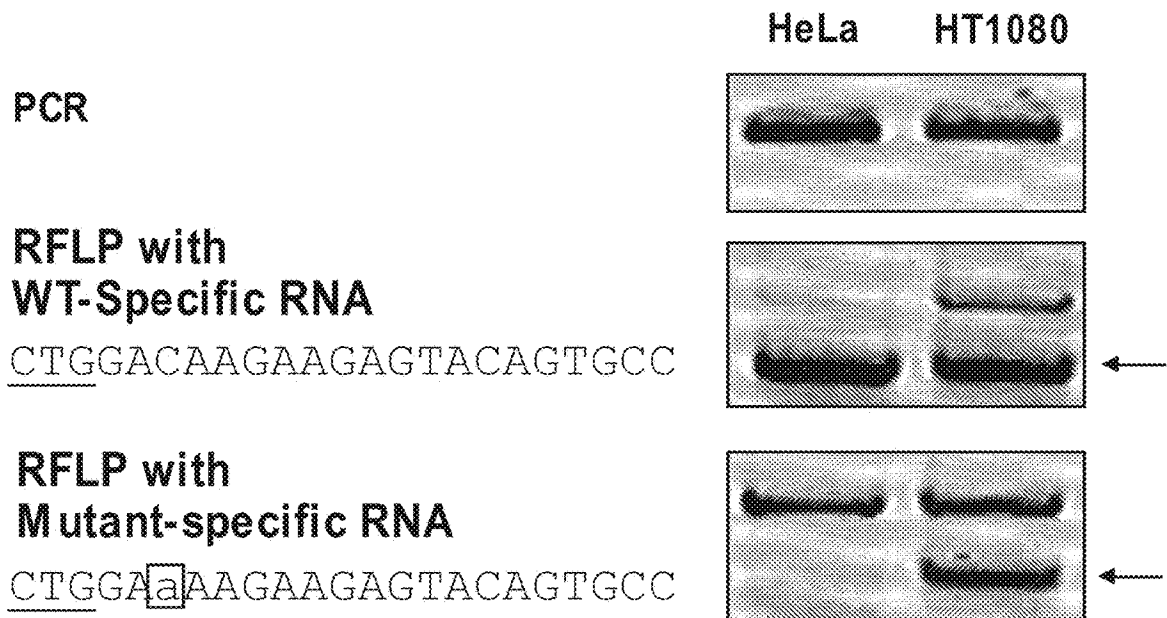


FIG. 33C

BRAF

HeLa

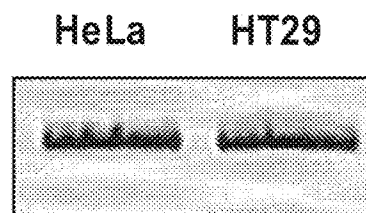
ACTCCATCGAGATTTCACTGTAG Wild-type

HT29

ACTCCATCGAGATTTCACTGTAG Wild-type

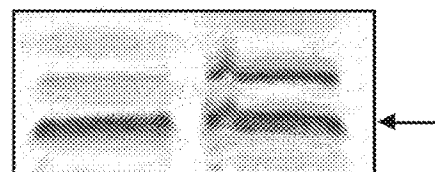
ACTCCATCGAGATTTCTCTGTAG (c.1799T>A)

PCR



**RFLP with
WT-Specific RNA**

ACTCCATCGAGATTTCACTGTAG



**RFLP with
Mutant-specific RNA**

ACTCCATCGAGATTTCTCTGTAG

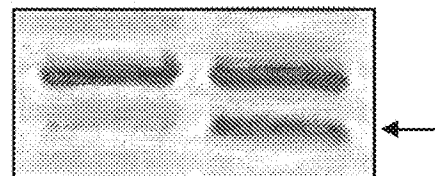


FIG. 33D

1

**COMPOSITIONS FOR INDUCING
MODIFICATIONS OF TARGET
ENDOGENOUS NUCLEIC ACID SEQUENCES
IN NUCLEUSES OF EUKARYOTIC CELLS**

CROSS REFERENCE TO RELATED
APPLICATIONS

The present application is a continuation application of U.S. application Ser. No. 18/932,745 filed Oct. 31, 2024, which is a continuation application of U.S. application Ser. No. 17/004,338 filed Aug. 27, 2020, which is a continuation application of U.S. application Ser. No. 14/685,568 filed Apr. 13, 2015, which is a continuation of PCT/KR2013/009488 filed Oct. 23, 2013, which claims priority to U.S. Provisional Application No. 61/837,481 filed on Jun. 20, 2013, U.S. Provisional Application No. 61/803,599 filed Mar. 20, 2013, and U.S. Provisional Application No. 61/717,324 filed Oct. 23, 2012, the entire contents of each aforementioned application are incorporated herein by reference.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in XML file format and is hereby incorporated by reference in its entirety. Said XML copy, created on Dec. 18, 2024, is named 00153_SL.xml and is 495,164 bytes in size.

TECHNICAL FIELD

The present invention relates to targeted genome editing in eukaryotic cells or organisms. More particularly, the present invention relates to a composition for cleaving a target DNA in eukaryotic cells or organisms comprising a guide RNA specific for the target DNA and Cas protein-encoding nucleic acid or Cas protein, and use thereof.

BACKGROUND ART

CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) are loci containing multiple short direct repeats that are found in the genomes of approximately 40% of sequenced bacteria and 90% of sequenced archaea. CRISPR functions as a prokaryotic immune system, in that it confers resistance to exogenous genetic elements such as plasmids and phages. The CRISPR system provides a form of acquired immunity. Short segments of foreign DNA, called spacers, are incorporated into the genome between CRISPR repeats, and serve as a memory of past exposures. CRISPR spacers are then used to recognize and silence exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

Cas9, an essential protein component in the Type II CRISPR/Cas system, forms an active endonuclease when complexed with two RNAs termed CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), thereby slicing foreign genetic elements in invading phages or plasmids to protect the host cells. crRNA is transcribed from the CRISPR element in the host genome, which was previously captured from such foreign invaders. Recently, Jinek et al. (1) demonstrated that a single-chain chimeric RNA produced by fusing an essential portion of crRNA and tracrRNA could replace the two RNAs in the Cas9/RNA complex to form a functional endonuclease.

CRISPR/Cas systems offer an advantage to zinc finger and transcription activator-like effector DNA-binding pro-

2

teins, as the site specificity in nucleotide binding CRISPR-Cas proteins is governed by a RNA molecule instead of the DNA-binding protein, which can be more challenging to design and synthesize.

5 However, until now, a genome editing method using the RNA-guided endonuclease (RGEN) based on CRISPR/Cas system has not been developed.

Meanwhile, Restriction fragment length polymorphism (RFLP) is one of the oldest, most convenient, and least expensive methods of genotyping that is still used widely in molecular biology and genetics but is often limited by the lack of appropriate sites recognized by restriction endonucleases.

10 Engineered nuclease-induced mutations are detected by various methods, which include mismatch-sensitive T7 endonuclease I (T7E1) or Surveyor nuclease assays, RFLP, capillary electrophoresis of fluorescent PCR products, Dideoxy sequencing, and deep sequencing. The T7E1 and Surveyor assays are widely used but are cumbersome. Furthermore, these enzymes tend to underestimate mutation frequencies because mutant sequences can form homoduplexes with each other and cannot distinguish homozygous bi-allelic mutant clones from wildtype cells. RFLP is free of these limitations and therefore is a method of choice. Indeed, RFLP was one of the first methods to detect engineered nuclease-mediated mutations in cells and animals. Unfortunately, however, RFLP is limited by the availability of appropriate restriction sites. It is possible that no restriction sites are available at the target site of interest.

DISCLOSURE OF INVENTION

Technical Problem

35 Until now, a genome editing and genotyping method using the RNA-guided endonuclease (RGEN) based on CRISPR/Cas system has not been developed.

Under these circumstances, the present inventors have made many efforts to develop a genome editing method based on CRISPR/Cas system and finally established a programmable RNA-guided endonuclease that cleave DNA in a targeted manner in eukaryotic cells and organisms.

40 In addition, the present inventors have made many efforts to develop a novel method of using RNA-guided endonucleases (RGENs) in RFLP analysis. They have used RGENs to genotype recurrent mutations found in cancer and those induced in cells and organisms by engineered nucleases including RGENs themselves, thereby completing the present invention.

Solution to Problem

55 It is an object of the present invention to provide a composition for cleaving target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is another object of the present invention to provide a composition for inducing targeted mutagenesis in eukaryotic cells or organisms, comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

60 It is still another object of the present invention to provide a kit for cleaving a target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

3

It is still another object of the present invention to provide a kit for inducing targeted mutagenesis in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for preparing a eukaryotic cell or organism comprising Cas protein and a guide RNA comprising a step of co-transfecting or serial-transfecting the eukaryotic cell or organism with a Cas protein-encoding nucleic acid or Cas protein, and a guide RNA or DNA that encodes the guide RNA.

It is still another object of the present invention to provide a eukaryotic cell or organism comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for cleaving a target DNA in eukaryotic cells or organisms comprising a step of transfecting the eukaryotic cells or organisms comprising a target DNA with a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for inducing targeted mutagenesis in a eukaryotic cell or organism comprising a step of treating a eukaryotic cell or organism with a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide an embryo, a genome-modified animal, or genome-modified plant comprising a genome edited by a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method of preparing a genome-modified animal comprising a step of introducing the composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein into an embryo of an animal; and a step of transferring the embryo into a oviduct of pseudopregnant foster mother to produce a genome-modified animal.

It is still another object of the present invention to provide a composition for genotyping mutations or variations in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence Cas protein.

It is still another object of the present invention to provide a method of using a RNA-guided endonuclease (RGEN) to genotype mutations induced by engineered nucleases in cells or naturally-occurring mutations or variations, wherein the RGEN comprises a guide RNA specific for target DNA and Cas protein.

It is still another object of the present invention to provide a kit for genotyping mutations induced by engineered nucleases in cells or naturally-occurring mutations or variations, comprising a RNA-guided endonuclease (RGEN), wherein the RGEN comprises a guide RNA specific for target DNA and Cas protein.

It is an object of the present invention to provide a composition for cleaving target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is another object of the present invention to provide a composition for inducing targeted mutagenesis in eukaryotic cells or organisms, comprising a guide RNA specific for

4

target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a kit for cleaving a target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a kit for inducing targeted mutagenesis in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for preparing a eukaryotic cell or organism comprising Cas protein and a guide RNA comprising a step of co-transfecting or serial-transfecting the eukaryotic cell or organism with a Cas protein-encoding nucleic acid or Cas protein, and a guide RNA or DNA that encodes the guide RNA.

It is still another object of the present invention to provide a eukaryotic cell or organism comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for cleaving a target DNA in eukaryotic cells or organisms comprising a step of transfecting the eukaryotic cells or organisms comprising a target DNA with a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method for inducing targeted mutagenesis in a eukaryotic cell or organism comprising a step of treating a eukaryotic cell or organism with a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide an embryo, a genome-modified animal, or genome-modified plant comprising a genome edited by a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

It is still another object of the present invention to provide a method of preparing a genome-modified animal comprising a step of introducing the composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein into an embryo of an animal; and a step of transferring the embryo into a oviduct of pseudopregnant foster mother to produce a genome-modified animal.

It is still another object of the present invention to provide a composition for genotyping mutations or variations in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence Cas protein.

It is still another object of the present invention to provide a composition for genotyping nucleic acid sequences in pathogenic microorganisms in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence and Cas protein.

It is still another object of the present invention to provide a kit for genotyping mutations or variations in an isolated biological sample, comprising the composition, specifically comprising a RNA-guided endonuclease (RGEN), wherein the RGEN comprises a guide RNA specific for target DNA and Cas protein.

It is still another object of the present invention to provide a method of genotyping mutations or variations in an

isolated biological sample, using the composition, specifically comprising a RNA-guided endonuclease (RGEN), wherein the RGEN comprises a guide RNA specific for target DNA and Cas protein.

Advantageous Effects of Invention

The present composition for cleaving a target DNA or inducing a targeted mutagenesis in eukaryotic cells or organisms, comprising a guide RNA specific for the target DNA and Cas protein-encoding nucleic acid or Cas protein, the kit comprising the composition, and the method for inducing targeted mutagenesis provide a new convenient genome editing tools. In addition, because custom RGENs can be designed to target any DNA sequence, almost any single nucleotide polymorphism or small insertion/deletion (indel) can be analyzed via RGEN-mediated RFLP, therefore, the composition and method of the present invention may be used in detection and cleaving naturally-occurring variations and mutations.

BRIEF DESCRIPTION OF DRAWINGS

FIGS. 1A and 1B show Cas9-catalyzed cleavage of plasmid DNA in vitro. FIG. 1A: Schematic representation of target DNA (SEQ ID NO: 112) and chimeric RNA sequences (SEQ ID NO: 113). Triangles indicate cleavage sites. The PAM sequence recognized by Cas9 is shown in bold. The sequences in the guide RNA (SEQ ID NO: 113) derived from crRNA and tracrRNA are shown in box and underlined, respectively. FIG. 1B: In vitro cleavage of plasmid DNA by Cas9. An intact circular plasmid or ApaLI-digested plasmid was incubated with Cas9 and guide RNA.

FIGS. 2A and 2B show Cas9-induced mutagenesis at an episomal target site. FIG. 2A: Schematic overview of cell-based assays using a RFP-GFP reporter. GFP is not expressed from this reporter because the GFP sequence is fused to the RFP sequence out-of-frame. The RFP-GFP fusion protein is expressed only when the target site between the two sequences is cleaved by a site-specific nuclease. FIG. 2B: Flow cytometry of cells transfected with Cas9. The percentage of cells that express the RFP-GFP fusion protein is indicated.

FIGS. 3A and 3B show RGEN-driven mutations at endogenous chromosomal sites. FIG. 3A: CCR5 locus. FIG. 3B: C4BPB locus. (Top) The T7E1 assay was used to detect RGEN-driven mutations. Arrows indicate the expected position of DNA bands cleaved by T7E1. Mutation frequencies (Indels (%)) were calculated by measuring the band intensities. (Bottom) DNA sequences of the wild-type (WT) CCR5 (SEQ ID NO: 114) and C4BPB (SEQ ID NO: 122) and mutant clones. DNA sequences of RGEN-induced mutations at the CCR5 locus: +1 (SEQ ID NO: 115), -13 (SEQ ID NO: 116), -14 (SEQ ID NO: 117), -18 (SEQ ID NO: 118), -19 (SEQ ID NO: 119), -24 (SEQ ID NO: 120), and -30 (SEQ ID NO: 121). DNA sequences of RGEN-induced mutations at the C4BPB locus: +1 (SEQ ID NO: 122), +2 (SEQ ID NO: 123), -30 (SEQ ID NO: 125), and -180 (SEQ ID NO: 126). The region of the target sequence complementary to the guide RNA is shown in box. The PAM sequence is shown in bold. Triangles indicate the cleavage site. Bases corresponding to microhomologies are underlined. The column on the right indicates the number of inserted or deleted bases.

FIGS. 4A, 4B, and 4C show that RGEN-driven off-target mutations are undetectable. FIG. 4A: On-target and potential off-target sequences. The human genome was searched in

silico for potential off-target sites. Four sites were identified, ADCY5 (SEQ ID NO: 128), KCNJ6 (SEQ ID NO: 129), CNTNAP2 (SEQ ID NO: 130), and Chr. 5 N/A (SEQ ID NO: 131), each of which carries 3-base mismatches with the CCR5 on-target (SEQ ID NO: 127). Mismatched bases are underlined. FIG. 4B: The T7E1 assay was used to investigate whether these sites were mutated in cells transfected with the Cas9/RNA complex. No mutations were detected at these sites. N/A (not applicable), an intergenic site. FIG. 4C: Cas9 did not induce off-target-associated chromosomal deletions. The CCR5-specific RGEN and ZFN were expressed in human cells. PCR was used to detect the induction of the 15-kb chromosomal deletions in these cells.

FIGS. 5A, 5B, 5C, and 5D show RGEN-induced Foxn1 gene targeting in mice. FIG. 5A: A schematic diagram depicting target DNA (SEQ ID NO: 132) and a sgRNA specific to exon 2 of the mouse Foxn1 gene (SEQ ID NO: 133). PAM in exon 2 is shown in a box and the sequence in the sgRNA that is complementary to exon 2 is underlined. Triangles indicate cleavage sites. FIG. 5B: Representative T7E1 assays demonstrating gene-targeting efficiencies of Cas9 mRNA plus Foxn1-specific sgRNA that were delivered via intra-cytoplasmic injection into one-cell stage mouse embryos. Numbers indicate independent founder mice generated from the highest dose. Arrows indicate bands cleaved by T7E1. FIG. 5C: DNA sequences of wild-type (WT) Foxn1 (SEQ ID NO: 134) and mutant alleles (SEQ ID NOS. 135-141) observed in three Foxn1 mutant founders identified in FIG. 5B. DNA sequences of mutant alleles in founder #108: -44 (SEQ ID NO: 135), -23 (SEQ ID NO: 136), -17 (SEQ ID NO: 137), and +1 (SEQ ID NO: 138). DNA sequences of mutant alleles in founder #111: +1 (SEQ ID NO: 138) and -11 (SEQ ID NO: 139). DNA sequences of mutant alleles in founder #114: -6 (SEQ ID NO: 140), -17 (SEQ ID NO: 137), and -8 (SEQ ID NO: 141). The number of occurrences is shown in parentheses. FIG. 5D: PCR genotyping of F1 progenies derived from crossing Foxn1 founder #108 and wild-type FVB/NTac. Note the segregation of the mutant alleles found in Foxn1 founder #108 in the progenies.

FIGS. 6A, 6B, and 6C show Foxn1 gene targeting in mouse embryos by intra-cytoplasmic injection of Cas9 mRNA and Foxn1-sgRNA. FIG. 6A: A representative result of a T7E1 assay monitoring the mutation rate after injecting the highest dose. Arrows indicate bands cleaved by T7E1. FIG. 6B: A summary of T7E1 assay results. Mutant fractions among in vitro cultivated embryos obtained after intra-cytoplasmic injection of the indicated RGEN doses are indicated. FIG. 6C: DNA sequences of wild-type (WT) Foxn1 (SEQ ID NO: 143) and Foxn1 mutant alleles (SEQ ID NOS. 144-152) identified from a subset of T7E1-positive mutant embryos. The DNA sequences of the mutant alleles are: Δ 11 (SEQ ID NO: 144), Δ 11+ Δ 17 (SEQ ID NO: 145) Δ 57 (SEQ ID NO: 146), Δ 17 (SEQ ID NO: 147), +1 (SEQ ID NO: 148), Δ 12 (SEQ ID NO: 149), Δ 72 (SEQ ID NO: 150), Δ 25 (SEQ ID NO: 151), Δ 24 (SEQ ID NO: 152). The target sequence of the wild-type allele is denoted in box.

FIGS. 7A, 7B, and 7C show Foxn1 gene targeting in mouse embryos using the recombinant Cas9 protein: Foxn1-sgRNA complex. FIG. 7A and FIG. 7B are representative T7E1 assays results and their summaries. Embryos were cultivated in vitro after they underwent pronuclear (FIG. 7A) or intra-cytoplasmic injection (FIG. 7B). Underlined numbers indicate T7E1-positive mutant founder mice. FIG. 7C: DNA sequences of wild-type (WT) Foxn1 (SEQ ID NO: 153) and Foxn1 mutant alleles (SEQ ID NOS. 154-166) identified from the in vitro cultivated embryos that were

obtained by the pronucleus injection of recombinant Cas9 protein: Foxn1-sgRNA complex at the highest dose. The target sequence of the wild-type allele is denoted in box. The DNA sequences of the mutant alleles are: $\Delta 18$ (SEQ ID NO: 154), $\Delta 20$ (SEQ ID NO: 155), $\Delta 19$ (SEQ ID NO: 156), $\Delta 17$ (SEQ ID NO: 157), $\Delta 11$ (SEQ ID NO: 158), $\Delta 3+1$ (SEQ ID NO: 159), $\Delta 2$ (SEQ ID NO: 160), +1, Embryo 1 (SEQ ID NO: 161), +1, Embryo 10 (SEQ ID NO: 162), $\Delta 6$ (SEQ ID NO: 163), $\Delta 5$ (SEQ ID NO: 164), $\Delta 28$ (SEQ ID NO: 165), and $\Delta 126$ (SEQ ID NO: 166).

FIGS. 8A, 8B, and 8C show Germ-line transmission of the mutant alleles found in Foxn1 mutant founder #12. FIG. 8A: wild type *f*PCR analysis. FIG. 8B: Foxn1 mutant founder #12 *f*PCR analysis. FIG. 8C: PCR genotyping of wild-type FVB/NTac, the founder mouse, and their F1 progenies.

FIGS. 9A and 9B show Genotypes of embryos generated by crossing Prkdc mutant founders. Prkdc mutant founders $\delta 25$ and $\eta 15$ were crossed and E13.5 embryos were isolated. FIG. 9A: *f*PCR analysis of wild-type, founder $\delta 25$, and founder $\eta 15$. Note that, due to the technical limitations of *f*PCR analysis, these results showed small differences from the precise sequences of the mutant alleles; e.g., from the sequence analysis, $\Delta 269/\Delta 61/\text{WT}$ and $\Delta 5+1/+7/+12/\text{WT}$ were identified in founders $\delta 25$ and $\eta 15$, respectively. FIG. 9B: Genotypes of the generated embryos.

FIGS. 10A, 10B, 10C, 10D, and 10E show Cas9 protein/sgrRNA complex induced targeted mutation at CCR5 gene (FIGS. 10A-10C) and ABCC11 gene (FIGS. 10D-10E). FIG. 10A: Results of a T7E1 assay monitoring the mutation rate at CCR5 locus after introducing Cas9 protein and sgRNA or Cas9 protein and crRNA+tracrRNA into K562 cells. FIG. 10B: Results of a T7E1 assay using $1/5$ scaled down doses of Cas9 protein and sgRNA. FIG. 10C: Wild-type (WT) CCR5 sequence (SEQ ID NO: 114) and Cas protein induced mutant sequences (SEQ ID NOs. 167-171 and 115) identified in CCR5 locus. The DNA sequences of the mutant sequences are: -4 (SEQ ID NO: 167), $\Delta 4$ (SEQ ID NO: 168), -7 (SEQ ID NO: 169), -1 (SEQ ID NO: 170), +1 (SEQ ID NO: 115), and -17, +1 (SEQ ID NO: 171). FIG. 10D: Results of a T7E1 assay monitoring the mutation rate at ABCC11 locus after introducing Cas9 protein and sgRNA into K562 cells. FIG. 10E: Wild-type (WT) ABCC11 sequence (SEQ ID NO: 172) and Cas9 protein induced mutant sequences (SEQ ID NOs. 173-176) identified in ABCC11 locus. The DNA sequences of the mutant sequences are: -6 (SEQ ID NO: 173), -3 (SEQ ID NO: 174), -29 (SEQ ID NO: 175), -20 (SEQ ID NO: 176), and -256 (TTCTC).

FIG. 11 shows recombinant Cas9 protein-induced mutations in Arabidopsis protoplasts.

FIG. 12 shows wild type BR11 sequence (SEQ ID NO: 177) and recombinant Cas9 protein-induced mutant sequences (SEQ ID NOs. 178-181) in the Arabidopsis BR11 gene. The DNA sequences of the mutant sequences are: -7 (SEQ ID NO: 178), -224 (SEQ ID NO: 179), -223 (SEQ ID NO: 180), and -223, +62 (SEQ ID NO: 181).

FIG. 13 shows T7E1 assay showing endogenous CCR5 gene disruption in 293 cells by treatment of Cas9-mal-9R4L and sgRNA/C9R4LC complex.

FIGS. 14A and 14B show mutation frequencies at on-target and off-target sites of RGENs reported in Fu et al. (2013). T7E1 assays analyzing genomic DNA from K562 cells (R) transfected serially with 20 μg of Cas9-encoding plasmid and with 60 μg and 120 μg of in vitro transcribed GX19 crRNA and tracrRNA, respectively (1×10^6 cells), or (D) co-transfected with 1 μg of Cas9-encoding plasmid and

1 μg of GX19 sgRNA expression plasmid (2×10^5 cells). FIG. 14A: VEGFA site 1 on target sequence (SEQ ID NO: 182) and off target sequences, OT1-3 (SEQ ID NO: 183) and OT1-11 (SEQ ID NO: 184). VEGFA site 2 on target sequence (SEQ ID NO: 185) and off target sequences OT2-1 (SEQ ID NO: 186), OT2-9 (SEQ ID NO: 187) and OT2-24 (SEQ ID NO: 188). FIG. 14B: VEGFA site 3 on target sequence (SEQ ID NO: 189) and off target sequence OT3-18 (SEQ ID NO: 190) and EMX1 on target sequence (SEQ ID NO: 191) and off target sequence OT4-1 (SEQ ID NO: 192).

FIGS. 15A and 15B show comparison of guide RNA structure. Mutation frequencies of the RGENs reported in Fu et al. (2013) were measured at on-target and off-target sites using the T7E1 assay. K562 cells were co-transfected with the Cas9-encoding plasmid and the plasmid encoding GX19 sgRNA or GGX20 sgRNA. Off-target sites (OT1-3 etc.) are labeled as in Fu et al. (2013). FIG. 15A: VEGFA site 1 on target sequence (SEQ ID NO: 182) and off target sequences OT1-3 (SEQ ID NO: 183 and OT1-11 (SEQ ID NO: 184). VEGFA site 2 on target sequence (SEQ ID NO: 185) and off target sequences OT2-1 (SEQ ID NO: 186), OT2-9 (SEQ ID NO: 187), and OT2-24 (SEQ ID NO: 188). FIG. 15B: VEGFA site 3 on target sequence (SEQ ID NO: 189) and off target sequence OT3-18 (SEQ ID NO: 190) and EMX1 on target sequence (SEQ ID NO: 191) and off target sequence OT4-1 (SEQ ID NO: 192).

FIGS. 16A, 16B, 16C, and 16D show that in vitro DNA cleavage by Cas9 nickases. FIG. 16A: Schematic overview of the Cas9 nuclease and the paired Cas9 nickase. The PAM sequences and cleavage sites are shown in box. FIG. 16B: Target sites in the human AAVS1 locus. The position of each target site is shown in triangle. FIG. 16C: Schematic overview of DNA cleavage reactions. FAM dyes (shown in box) were linked to both 5' ends of the DNA substrate. FIG. 16D: DSBs and SSBs analyzed using fluorescent capillary electrophoresis. Fluorescently-labeled DNA substrates were incubated with Cas9 nucleases or nickases before electrophoresis.

FIGS. 17A and 17B show comparison of Cas9 nuclease and nickase behavior. FIG. 17A: On-target mutation frequencies associated with Cas9 nucleases (WT), nickases (D10A), and paired nickases at the following target sequences of the AAVS1 locus: S1 (SEQ ID NO: 193), S2 (SEQ ID NO: 194), S3 (SEQ ID NO: 195), S4 (SEQ ID NO: 196), S5 (SEQ ID NO: 197), S6 (SEQ ID NO: 198), AS1 (SEQ ID NO: 199), AS2 (SEQ ID NO: 200), and AS3 (SEQ ID NO: 201). Paired nickases that would produce 5' overhangs or 3' overhangs are indicated. FIG. 17B: Analysis of off-target effects of Cas9 nucleases and paired nickases. A total of seven potential off-target sites (SEQ ID NOs. 202-208) for three sgRNAs were analyzed. The mutation frequency for the S2 on-target sequence (SEQ ID NO: 194) was compared to the off-target sequences, S2 Off-1 (SEQ ID NO: 202) and S2 Off-2 (SEQ ID NO: 203). The mutation frequency for the S3 on-target sequence (SEQ ID NO: 195) was compared to the off-target sequences, S3 Off-1 (SEQ ID NO: 204) and S3 Off-2 (SEQ ID NO: 205). The mutation frequency for the AS2 on-target sequence (SEQ ID NO: 198) was compared to the off-target sequences, AS2 Off-1 (SEQ ID NO: 206), AS2 Off-6 (SEQ ID NO: 207), and AS2 Off-9 (SEQ ID NO: 208).

FIGS. 18A, 18B, 18C, and 18D show paired Cas9 nickases tested at other endogenous human loci. The sgRNA target sites at the human CCR5 locus (FIG. 18A; SEQ ID NO: 209) and the BRCA2 locus (FIG. 18C; SEQ ID NO: 210). PAM sequences are indicated in a box. Genome editing activities at CCR5 (FIG. 18B) and BRCA2 (FIG.

18D) target sites were detected by the T7E1 assay. The repair of two nicks that would produce 5' overhangs led to the formation of indels much more frequently than did those producing 3' overhangs.

FIGS. 19A and 19B show that paired Cas9 nickases mediate homologous recombination. FIG. 19A: Strategy to detect homologous recombination. Donor DNA included an XbaI restriction enzyme site between two homology arms, whereas the endogenous target site lacked this site. A PCR assay was used to detect sequences that had undergone homologous recombination. To prevent amplification of contaminating donor DNA, primers specific to genomic DNA were used. FIG. 19B: Efficiency of homologous recombination. Only amplicons of a region in which homologous recombination had occurred could be digested with XbaI; the intensities of the cleavage bands were used to measure the efficiency of this method.

FIGS. 20A, 20B, 20C, and 20D show DNA splicing induced by paired Cas9 nickases. FIG. 20A: The target sites of paired nickases in the human AAVS1 locus. The distances between the AS2 site and each of the other sites are shown. Arrows indicate PCR primers. FIG. 20B: Genomic deletions detected using PCR. Asterisks indicate deletion-specific PCR products. FIG. 20C: DNA sequences of wild-type (WT) (SEQ ID NO: 211 and 332) and the following deletion-specific PCR products (SEQ ID Nos. 212-218) obtained using AS2 sgRNAs or deletion-specific PCR products (SEQ ID NOs. 219-224) using L1 sgRNAs. Target site PAM sequences are shown in box and sgRNA-matching sequences are shown in capital letters. Intact sgRNA-matching sequences are underlined. FIG. 20D: A schematic model of paired Cas9 nickase-mediated chromosomal deletions. Newly-synthesized DNA strands are shown in box.

FIGS. 21A, 21B, and 21C show that paired Cas9 nickases do not induce translocations. FIG. 21A: Schematic overview of chromosomal translocations between the on-target and off-target sites. FIG. 21B: PCR amplification to detect chromosomal translocations. FIG. 21C: Translocations induced by Cas9 nucleases but not by the nickase pair.

FIGS. 22A and 22B show a conceptual diagram of the T7E1 and RFLP assays. FIG. 22A: Comparison of assay cleavage reactions in four possible scenarios after engineered nuclease treatment in a diploid cell: (A) wild type, (B) a monoallelic mutation, (C) different biallelic mutations (hetero), and (D) identical biallelic mutations (homo). Black lines represent PCR products derived from each allele; dashed and dotted boxes indicate insertion/deletion mutations generated by NHEJ. FIG. 22B: Expected results of T7E1 and RGEN digestion resolved by electrophoresis.

FIG. 23 shows in vitro cleavage assay of a linearized plasmid containing the C4BPB target site bearing indels. DNA sequences of individual plasmid substrates (upper panel): WT (SEQ ID NO: 104), I1 (SEQ ID NO: 225), I2 (SEQ ID NO: 226), I3 (SEQ ID NO: 227), D1 (SEQ ID NO: 228), D2 (SEQ ID NO: 229), and D3 (SEQ ID NO: 230). The PAM sequence is underlined. Inserted bases are shown in box. Arrows (bottom panel) indicate expected positions of DNA bands cleaved by the wild-type-specific RGEN after electrophoresis.

FIGS. 24A and 24B show genotyping of mutations induced by engineered nucleases in cells via RGEN-mediated RFLP. FIG. 24A: Genotype of C4BPB wild type (SEQ ID NO: 231) and the following mutant K562 cell clones: +3 (SEQ ID NO: 232), -12 (SEQ ID NO: 233), -9 (SEQ ID NO: 234), -8 (SEQ ID NO: 235), -36 (SEQ ID NO: 236), +1 (SEQ ID NO: 237), +1 (SEQ ID NO: 238), +67 (SEQ ID NO: 239), -7, +1 (SEQ ID NO: 240), -94 (SEQ ID NO:

241). FIG. 24B: Comparison of the mismatch-sensitive T7E1 assay with RGEN-mediated RFLP analysis. Black arrows indicate the cleavage product by treatment of T7E1 enzyme or RGENs.

FIGS. 25A, 25B, and 25C show genotyping of RGEN-induced mutations via the RGEN-RFLP technique. FIG. 25A: Analysis of C4BPB-disrupted clones using RGEN-RFLP and T7E1 assays. Arrows indicate expected positions of DNA bands cleaved by RGEN or T7E1. FIG. 25B: Quantitative comparison of RGEN-RFLP analysis with T7E1 assays. Genomic DNA samples from wild-type and C4BPB-disrupted K562 cells were mixed in various ratios and subjected to PCR amplification. FIG. 25C: Genotyping of RGEN-induced mutations in the HLA-B gene in HeLa cells with RFLP and T7E1 analyses.

FIGS. 26A and 26B show genotyping of mutations induced by engineered nucleases in organisms via RGEN-mediated RFLP. FIG. 26A: Genotype of Pibf1 wild-type (WT) (SEQ ID NO: 242) and the following mutant founder mice: #1 (SEQ ID NO: 243 and SEQ ID NO: 244), #3 (SEQ ID NO: 245 and SEQ ID NO: 246), #4 (SEQ ID NO: 247 and SEQ ID NO: 242), #5 (SEQ ID NO: 246 and SEQ ID NO: 242), #6 (SEQ ID NO: 248 and SEQ ID NO: 249), #8 (SEQ ID NO: 250 and SEQ ID NO: 251), and #11 (SEQ ID NO: 252 and SEQ ID NO: 250). FIG. 26B: Comparison of the mismatch-sensitive T7E1 assay with RGEN-mediated RFLP analysis. Black arrows indicate the cleavage product by treatment of T7E1 enzyme or RGENs.

FIG. 27 shows RGEN-mediated genotyping of ZEN-induced mutations at a wild-type CCR5 sequence (SEQ ID NO: 253). The ZEN target site is shown in box. Black arrows indicate DNA bands cleaved by T7E1.

FIG. 28 shows polymorphic sites in a region of the human HLA-B gene (SEQ ID NO: 254). The sequence, which surrounds the RGEN target site, is that of a PCR amplicon from HeLa cells. Polymorphic positions are shown in box. The RGEN target site and the PAM sequence are shown in dashed and bolded box, respectively. Primer sequences are underlined.

FIGS. 29A and 29B show genotyping of oncogenic mutations via RGEN-RFLP analysis. FIG. 29A: A recurrent mutation (c.133-135 deletion of TCT; SEQ ID NO: 256) in the human CTNNB1 gene in HCT116 cells was detected by RGENs. The wild-type CTNNB1 sequence is represented by SEQ ID NO: 255. HeLa cells were used as a negative control. FIG. 29B: Genotyping of the KRAS substitution mutation (c.34 G>A) in the A549 cancer cell line with RGENs that contain mismatched guide RNA that are WT-specific (SEQ ID NO: 257) or mutant-specific (SEQ ID NO: 258). Mismatched nucleotides are shown in box. HeLa cells were used as a negative control. Arrows indicate DNA bands cleaved by RGENs. DNA sequences confirmed by Sanger sequencing are shown: wild-type (SEQ ID NO: 259) and c. 34G>A (SEQ ID NO: 260).

FIGS. 30A, 30B, 30C, and 30D show genotyping of the CCR5 delta32 allele in HEK293T cells via RGEN-RFLP analysis. FIG. 30A: RGEN-RFLP assays of cell lines. DNA sequences of the wild-type CCR5 locus (SEQ ID NO: 262) and delta 32 mutation (SEQ ID NO: 261) are shown. K562, SKBR3, and HeLa cells were used as wild-type controls. Arrows indicate DNA bands cleaved by RGENs. FIG. 30B: DNA sequence of wild-type (SEQ ID NO: 263) and delta32 CCR5 alleles (SEQ ID NO: 264). Both on-target and off-target sites of RGENs used in RFLP analysis are underlined. A single-nucleotide mismatch between the two sites is shown in box. The PAM sequence is underlined. FIG. 30C: In vitro cleavage of plasmids harboring WT or del32 CCR5

alleles using the wild-type-specific RGEN. FIG. 30D Confirming the presence of an off-target site of the CCR5-delta32-specific RGEN at the CCR5 locus. In vitro cleavage assays of plasmids harboring either on-target (SEQ ID NO: 265) or off-target sequences (SEQ ID NO: 266) using various amounts of the del32-specific RGEN.

FIGS. 31A and 31B show genotyping of a KRAS point mutation (c.34 G>A). FIG. 31A: RGEN-RFLP analysis of the KRAS mutation (c.34 G>A) in cancer cell lines. PCR products from Hela cells (used as a wild-type control) or A549 cells, which are homozygous for the point mutation, were digested with RGENs with perfectly matched crRNA specific to the wild-type sequence (SEQ ID NO: 259) or the mutant sequence (SEQ ID NO: 260). KRAS genotypes in these cells were confirmed by Sanger sequencing. FIG. 31B: Plasmids harboring either the wild-type (SEQ ID NO: 259) or mutant KRAS sequences (SEQ ID NO: 260) were digested using RGENs with perfectly matched crRNAs or attenuated, one-base mismatched crRNAs: m7 (SEQ ID NO: 267), m6 (SEQ ID NO: 257), m5 (SEQ ID NO: 268), m4 (SEQ ID NO: 269), m8 (SEQ ID NO: 260), m7,8 (SEQ ID NO: 270), m6, 8 (SEQ ID NO: 258), m5, 8 (SEQ ID NO: 271), and m4, 8 (SEQ ID NO: 272). Attenuated crRNAs that were chosen for genotyping are labeled in box above the gels.

FIGS. 32A and 32B show genotyping of a PIK3CA point mutation (c.3140 A>G). FIG. 32A: RGEN-RFLP analysis of the PIK3CA mutation (c.3140 A>G) in cancer cell lines. PCR products from Hela cells (used as a wild-type control) or HCT116 cells that are heterozygous for the point mutation were digested with RGENs with perfectly matched crRNA specific to the wild-type sequence (SEQ ID NO: 273) or the mutant sequence (SEQ ID NO: 274). PIK3CA genotypes in these cells were confirmed by Sanger sequencing. FIG. 32B: Plasmids harboring either the wild-type PIK3CA sequence (SEQ ID NO: 273) or mutant PIK3CA sequence (SEQ ID NO: 274) were digested using RGENs with perfectly matched crRNAs or attenuated, one-base mismatched crRNAs: m5 (SEQ ID NO: 275), m6 (SEQ ID NO: 276), m7 (SEQ ID NO: 277), m10 (SEQ ID NO: 278), m13 (SEQ ID NO: 279), m16 (SEQ ID NO: 280), m19 (SEQ ID NO: 281), m4 (SEQ ID NO: 274), m4, 5 (SEQ ID NO: 282), m4, 6 (SEQ ID NO: 283), m4, 7 (SEQ ID NO: 284), m4,10 (SEQ ID NO: 285), m4, 13 (SEQ ID NO: 286), m4, 16 (SEQ ID NO: 287), and m4, 19 (SEQ ID NO: 288). Attenuated crRNAs that were chosen for genotyping are labeled in box above the gels.

FIGS. 33A, 33B, 33C, and 33D show genotyping of recurrent point mutations in cancer cell lines. FIG. 33A: RGEN-RFLP assays to distinguish between a wild-type IDH gene sequence (SEQ ID NO: 289) and a recurrent oncogenic point mutation sequence in the IDH gene (c. 394c>T; SEQ ID NO: 290). RGENs with attenuated, one-base mismatched crRNAs, SEQ ID NO: 291 (WT-Specific RNA) and SEQ ID NO: 292 (Mutant-Specific RNA), distinguished the wild type and mutant IDH sequences. FIG. 33B: RGEN-RFLP assays to distinguish between a wild-type PIK3CA gene sequence (SEQ ID NO: 271) and a recurrent oncogenic point mutation sequence in the PIK3CA gene (c.3140A>G; SEQ ID NO: 273). RGENs with attenuated, one-base mismatched crRNAs, SEQ ID NO: 275 (WT-Specific RNA) and SEQ ID NO: 284 (Mutant-Specific RNA), distinguished the wild type and mutant PIK3CA sequences. FIG. 33C: RGEN-RFLP assays to distinguish between a wild-type NRAS gene sequence (SEQ ID NO: 293) and a recurrent oncogenic point mutation sequence in the NRAS gene (c.181C>A; SEQ ID NO: 294). RGENs with perfectly matched crRNAs, SEQ ID

NO: 293 (WT-Specific RNA) and SEQ ID NO: 294 (Mutant-Specific RNA), distinguished the wild type and mutant NRAS sequences. FIG. 33D: RGEN-RFLP assays to distinguish between a wild-type BRAF gene sequence (SEQ ID NO: 295) and a recurrent oncogenic point mutation sequence in the BRAF gene (c.1799T>A; SEQ ID NO: 296). RGENs with perfectly matched crRNAs, SEQ ID NO: 295 (WT-Specific RNA) and SEQ ID NO: 296 (Mutant-Specific RNA), distinguished the wild type and mutant BRAF sequences. Genotypes of each cell line confirmed by Sanger sequencing are shown. Mismatched nucleotides are shown in box. Black arrows indicate DNA bands cleaved by RGENs.

BEST MODE FOR CARRYING OUT THE INVENTION

In accordance with one aspect of the invention, the present invention provides a composition for cleaving target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein. In addition, the present invention provides a use of the composition for cleaving target DNA in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

In the present invention, the composition is also referred to as a RNA-guided endonuclease (RGEN) composition.

ZFNs and TALENs enable targeted mutagenesis in mammalian cells, model organisms, plants, and livestock, but the mutation frequencies obtained with individual nucleases are widely different from each other. Furthermore, some ZFNs and TALENs fail to show any genome editing activities. DNA methylation may limit the binding of these engineered nucleases to target sites. In addition, it is technically challenging and time-consuming to make customized nucleases.

The present inventors have developed a new RNA-guided endonuclease composition based on Cas protein to overcome the disadvantages of ZFNs and TALENs.

Prior to the present invention, an endonuclease activity of Cas proteins has been known. However, it has not been known whether the endonuclease activity of Cas protein would function in an eukaryotic cell because of the complexity of the eukaryotic genome. Further, until now, a composition comprising Cas protein or Cas protein-encoding nucleic acid and a guide RNA specific for the target DNA to cleave a target DNA in eukaryotic cells or organisms has not been developed.

Compared to ZFNs and TALENs, the present RGEN composition based on Cas protein can be more readily customized because only the synthetic guide RNA component is replaced to make a new genome-editing nuclease. No sub-cloning steps are involved to make customized RNA guided endonucleases. Furthermore, the relatively small size of the Cas gene (for example, 4.2 kbp for Cas9) as compared to a pair of TALEN genes (~6 kbp) provides an advantage for this RNA-guided endonuclease composition in some applications such as virus-mediated gene delivery. Further, this RNA-guided endonuclease does not have off-target effects and thus does not induce unwanted mutations, deletion, inversions, and duplications. These features make the present RNA-guided endonuclease composition a scalable, versatile, and convenient tool for genome engineering in eukaryotic cells and organisms. In addition, RGEN can be designed to target any DNA sequence, almost any single nucleotide polymorphism or small insertion/deletion (indel)

can be analyzed via RGEN-mediated RFLP. The specificity of RGENs is determined by the RNA component that hybridizes with a target DNA sequence of up to 20 base pairs (bp) in length and by the Cas9 protein that recognizes the protospacer-adjacent motif (PAM). RGENs are readily reprogrammed by replacing the RNA component. Therefore, RGENs provide a platform to use simple and robust RFLP analysis for various sequence variations.

The target DNA may be an endogenous DNA, or artificial DNA, preferably, endogenous DNA.

As used herein, the term “Cas protein” refers to an essential protein component in the CRISPR/Cas system, forms an active endonuclease or nickase when complexed with two RNAs termed CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA).

The information on the gene and protein of Cas are available from GenBank of National Center for Biotechnology Information (NCBI), without limitation.

The CRISPR-associated (cas) genes encoding Cas proteins are often associated with CRISPR repeat-spacer arrays. More than forty different Cas protein families have been described. Of these protein families, Cas1 appears to be ubiquitous among different CRISPR/Cas systems. There are three types of CRISPR-Cas system. Among them, Type II CRISPR/Cas system involving Cas9 protein and crRNA and tracrRNA is representative and is well known. Particular combinations of cas genes and repeat structures have been used to define 8 CRISPR subtypes (Ecoli, Ypest, Nmeni, Dvulg, Tneap, Hmari, Aperm, and Mtube).

The Cas protein may be linked to a protein transduction domain. The protein transduction domain may be poly-arginine or a TAT protein derived from HIV, but it is not limited thereto.

The present composition may comprise Cas component in the form of a protein or in the form of a nucleic acid encoding Cas protein.

In the present invention, Cas protein may be any Cas protein provided that it has an endonuclease or nickase activity when complexed with a guide RNA.

Preferably, Cas protein is Cas9 protein or variants thereof.

The variant of the Cas9 protein may be a mutant form of Cas9 in which the catalytic aspartate residue is changed to any other amino acid. Preferably, the other amino acid may be an alanine, but it is not limited thereto.

Further, Cas protein may be the one isolated from an organism such as *Streptococcus* sp., preferably *Streptococcus pyogenes* or a recombinant protein, but it is not limited thereto.

The Cas protein derived from *Streptococcus pyogenes* may recognize NGG trinucleotide. The Cas protein may comprise an amino acid sequence of SEQ ID NO: 109, but it is not limited thereto.

The term “recombinant” when used with reference, e.g., to a cell, nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, a recombinant Cas protein may be generated by reconstituting Cas protein-encoding sequence using the human codon table.

As for the present invention, Cas protein-encoding nucleic acid may be a form of vector, such as plasmid comprising Cas-encoding sequence under a promoter such as CMV or CAG. When Cas protein is Cas9, Cas9 encoding sequence may be derived from *Streptococcus* sp., and preferably derived from *Streptococcus pyogenes*. For example,

Cas9 encoding nucleic acid may comprise the nucleotide sequence of SEQ ID. NO: 1. Moreover, Cas9 encoding nucleic acid may comprise the nucleotide sequence having homology of at least 50% to the sequence of SEQ ID NO: 1, preferably at least 60, 70, 80, 90, 95, 97, 98, or 99% to the SEQ ID NO:1, but it is not limited thereto. Cas9 encoding nucleic acid may comprise the nucleotide sequence of SEQ ID NOs.108, 110, 106, or 107.

As used herein, the term “guide RNA” refers to a RNA which is specific for the target DNA and can form a complex with Cas protein and bring Cas protein to the target DNA.

In the present invention, the guide RNA may consist of two RNA, i.e., CRISPR RNA (crRNA) and transactivating crRNA (tracrRNA) or be a single-chain RNA (sgRNA) produced by fusion of an essential portion of crRNA and tracrRNA.

The guide RNA may be a dualRNA comprising a crRNA and a tracrRNA.

If the guide RNA comprises the essential portion of crRNA and tracrRNA and a portion complementary to a target, any guide RNA may be used in the present invention.

The crRNA may hybridize with a target DNA.

The RGEN may consist of Cas protein, and dualRNA (invariable tracrRNA and target-specific crRNA), or Cas protein and sgRNA (fusion of an essential portion of invariable tracrRNA and target-specific crRNA), and may be readily reprogrammed by replacing crRNA.

The guide RNA further comprises one or more additional nucleotides at the 5' end of the single-chain guide RNA or the CRNA of the dualRNA.

Preferably, the guide RNA further comprises 2-additional guanine nucleotides at the 5' end of the single-chain guide RNA or the crRNA of the dualRNA.

The guide RNA may be transferred into a cell or an organism in the form of RNA or DNA that encodes the guide RNA. The guide RNA may be in the form of an isolated RNA, RNA incorporated into a viral vector, or is encoded in a vector. Preferably, the vector may be a viral vector, plasmid vector, or agrobacterium vector, but it is not limited thereto.

A DNA that encodes the guide RNA may be a vector comprising a sequence coding for the guide RNA. For example, the guide RNA may be transferred into a cell or organism by transfecting the cell or organism with the isolated guide RNA or plasmid DNA comprising a sequence coding for the guide RNA and a promoter.

Alternatively, the guide RNA may be transferred into a cell or organism using virus-mediated gene delivery.

When the guide RNA is transfected in the form of an isolated RNA into a cell or organism, the guide RNA may be prepared by in vitro transcription using any in vitro transcription system known in the art. The guide RNA is preferably transferred to a cell in the form of isolated RNA rather than in the form of plasmid comprising encoding sequence for a guide RNA. As used herein, the term “isolated RNA” may be interchangeable to “naked RNA”. This is cost- and time-saving because it does not require a step of cloning. However, the use of plasmid DNA or virus-mediated gene delivery for transfection of the guide RNA is not excluded.

The present RGEN composition comprising Cas protein or Cas protein-encoding nucleic acid and a guide RNA can specifically cleave a target DNA due to a specificity of the guide RNA for a target and an endonuclease or nickase activity of Cas protein.

As used herein, the term “cleavage” refers to the breakage of the covalent backbone of a nucleotide molecule.

In the present invention, a guide RNA may be prepared to be specific for any target which is to be cleaved. Therefore, the present RGEN composition can cleave any target DNA by manipulating or genotyping the target-specific portion of the guide RNA.

The guide RNA and the Cas protein may function as a pair. As used herein, the term "paired Cas nickase" may refer to the guide RNA and the Cas protein functioning as a pair. The pair comprises two guide RNAs. The guide RNA and Cas protein may function as a pair, and induce two nicks on different DNA strand. The two nicks may be separated by at least 100 bps, but are not limited thereto.

In the Example, the present inventors confirmed that paired Cas nickase allow targeted mutagenesis and large deletions of up to 1-kbp chromosomal segments in human cells. Importantly, paired nickases did not induce indels at off-target sites at which their corresponding nucleases induce mutations. Furthermore, unlike nucleases, paired nickases did not promote unwanted translocations associated with off-target DNA cleavages. In principle, paired nickases double the specificity of Cas9-mediated mutagenesis and will broaden the utility of RNA-guided enzymes in applications that require precise genome editing such as gene and cell therapy.

In the present invention, the composition may be used in the genotyping of a genome in the eukaryotic cells or organisms in vitro.

In one specific embodiment, the guide RNA may comprise the nucleotide sequence of Seq ID. No. 1, wherein the portion of nucleotide position 3~22 is a target-specific portion and thus, the sequence of this portion may be changed depending on a target.

As used herein, a eukaryotic cell or organism may be yeast, fungus, protozoa, plant, higher plant, and insect, or amphibian cells, or mammalian cells such as CHO, HeLa, HEK293, and COS-1, for example, cultured cells (in vitro), graft cells and primary cell culture (in vitro and ex vivo), and in vivo cells, and also mammalian cells including human, which are commonly used in the art, without limitation.

In one specific embodiment, it was found that Cas9 protein/single-chain guide RNA could generate site-specific DNA double-strand breaks in vitro and in mammalian cells, whose spontaneous repair induced targeted genome mutations at high frequencies.

Moreover, it was found that gene-knockout mice could be induced by the injection of Cas9 protein/guide RNA complexes or Cas9 mRNA/guide RNA into one-cell stage embryo and germ-line transmittable mutations could be generated by Cas9/guide RNA system.

Using Cas protein rather than a nucleic acid encoding Cas protein to induce a targeted mutagenesis is advantageous because exogenous DNA is not introduced into an organism. Thus, the composition comprising Cas protein and a guide RNA may be used to develop therapeutics or value-added crops, livestock, poultry, fish, pets, etc.

In accordance with another aspect of the invention, the present invention provides a composition for inducing targeted mutagenesis in eukaryotic cells or organisms, comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein. In addition, the present invention provides a use of the composition for inducing targeted mutagenesis in eukaryotic cells or organisms, comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

In accordance with another aspect of the invention, the present invention provides a kit for cleaving a target DNA or inducing targeted mutagenesis in eukaryotic cells or organisms comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

The kit may comprise a guide RNA and Cas protein-encoding nucleic acid or Cas protein as separate components or as one composition.

The present kit may comprise some additional components necessary for transferring the guide RNA and Cas component to a cell or an organism. For example, the kit may comprise an injection buffer such as DEPC-treated injection buffer, and materials necessary for analysis of mutation of a target DNA, but are not limited thereto.

In accordance with another aspect, the present invention provides a method for preparing a eukaryotic cell or organism comprising Cas protein and a guide RNA comprising a step of co-transfecting or serial-transfecting the eukaryotic cell or organism with a Cas protein-encoding nucleic acid or Cas protein, and a guide RNA or DNA that encodes the guide RNA.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

In the present invention, a Cas protein-encoding nucleic acid or Cas protein and a guide RNA or DNA that encodes the guide RNA may be transferred into a cell by various methods known in the art, such as microinjection, electroporation, DEAE-dextran treatment, lipofection, nanoparticle-mediated transfection, protein transduction domain mediated transduction, virus-mediated gene delivery, and PEG-mediated transfection in protoplast, and so on, but are not limited thereto. Also, a Cas protein encoding nucleic acid or Cas protein and a guide RNA may be transferred into an organism by various method known in the art to administer a gene or a protein such as injection. A Cas protein-encoding nucleic acid or Cas protein may be transferred into a cell in the form of complex with a guide RNA, or separately. Cas protein fused to a protein transduction domain such as Tat can also be delivered efficiently into cells.

Preferably, the eukaryotic cell or organism is co-transfected or serial-transfected with a Cas9 protein and a guide RNA.

The serial-transfection may be performed by transfection with Cas protein-encoding nucleic acid first, followed by second transfection with naked guide RNA. Preferably, the second transfection is after 3, 6, 12, 18, 24 hours, but it is not limited thereto.

In accordance with another aspect, the present invention provides a eukaryotic cell or organism comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

The eukaryotic cells or organisms may be prepared by transferring the composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein into the cell or organism.

The eukaryotic cell may be yeast, fungus, protozoa, higher plant, and insect, or amphibian cells, or mammalian cells such as CHO, HeLa, HEK293, and COS-1, for example, cultured cells (in vitro), graft cells and primary cell culture (in vitro and ex vivo), and in vivo cells, and also mammalian cells including human, which are commonly

used in the art, without limitation. Further the organism may be yeast, fungus, protozoa, plant, higher plant, insect, amphibian, or mammal.

In accordance with another aspect of the invention, the present invention provides a method for cleaving a target DNA or inducing targeted mutagenesis in eukaryotic cells or organisms, comprising a step of treating a cell or organism comprising a target DNA with a composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

The step of treating a cell or organism with the composition may be performed by transferring the present composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein into the cell or organism.

As described in the above, such transfer may be performed by microinjection, transfection, electroporation, and so on.

In accordance with another aspect of the invention, the present invention provides an embryo comprising a genome edited by the present RGEN composition comprising a guide RNA specific for target DNA or DNA that encodes the guide RNA, and Cas protein-encoding nucleic acid or Cas protein.

Any embryo can be used in the present invention, and for the present invention, the embryo may be an embryo of a mouse. The embryo may be produced by injecting PMSG (Pregnant Mare Serum Gonadotropin) and hCG (human Chorionic Gonadotropin) into a female mouse of 4 to 7 weeks and the super-ovulated female mouse may be mated to males, and the fertilized embryos may be collected from oviducts.

The present RGEN composition introduced into an embryo can cleave a target DNA complementary to the guide RNA by the action of Cas protein and cause a mutation in the target DNA. Thus, the embryo into which the present RGEN composition has been introduced has an edited genome.

In one specific embodiment, it was found that the present RGEN composition could cause a mutation in a mouse embryo and the mutation could be transmitted to offspring.

A method for introducing the RGEN composition into the embryo may be any method known in the art, such as microinjection, stem cell insertion, retrovirus insertion, and so on. Preferably, a microinjection technique can be used.

In accordance with another aspect, the present invention provides a genome-modified animal obtained by transferring the embryo comprising a genome edited by the present RGEN composition into the oviducts of an animal.

In the present invention, the term "genome-modified animal" refers to an animal of which genome has been modified in the stage of embryo by the present RGEN composition and the type of the animal is not limited.

The genome-modified animal has mutations caused by a targeted mutagenesis based on the present RGEN composition. The mutations may be any one of deletion, insertion, translocation, inversion. The site of mutation depends on the sequence of guide RNA of the RGEN composition.

The genome-modified animal having a mutation of a gene may be used to determine the function of the gene.

In accordance with another aspect of the invention, the present invention provides a method of preparing a genome-modified animal comprising a step of introducing the present RGEN composition comprising a guide RNA specific for the target DNA or DNA that encodes the guide RNA and Cas protein-encoding nucleic acid or Cas protein into an embryo

of an animal; and a step of transferring the embryo into a oviduct of pseudopregnant foster mother to produce a genome-modified animal.

The step of introducing the present RGEN composition may be accomplished by any method known in the art such as microinjection, stem cell insertion, retroviral insertion, and so on.

In accordance with another aspect of the invention, the present invention provides a plant regenerated from the genome-modified protoplasts prepared by the method for eukaryotic cells comprising the RGEN composition.

In accordance with another aspect of the invention, the present invention provides a composition for genotyping mutations or variations in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence Cas protein. In addition, the present invention provides a composition for genotyping nucleic acid sequences in pathogenic microorganisms in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence and Cas protein.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

As used herein the term "genotyping" refers to the "Restriction fragment length polymorphism (RFLP) assay".

RFLP may be used in 1) the detection of indel in cells or organisms induced by the engineered nucleases, 2) the genotyping naturally-occurring mutations or variations in cells or organisms, or 3) the genotyping the DNA of infected pathogenic microorganisms including virus or bacteria, etc.

The mutations or variation may be induced by engineered nucleases in cells.

The engineered nuclease may be a Zinc Finger Nuclease (ZFNs), Transcription Activator-Like Effector Nucleases (TALENs), or RGENs, but it is not limited thereto.

As used herein the term "biological sample" includes samples for analysis, such as tissues, cells, whole blood, serum, plasma, saliva, sputum, cerebrospinal fluid or urine, but is not limited thereto.

The mutations or variation may be a naturally-occurring mutations or variations.

The mutations or variations are induced by the pathogenic microorganisms. Namely, the mutations or variations occur due to the infection of pathogenic microorganisms, when the pathogenic microorganisms are detected, the biological sample is identified as infected.

The pathogenic microorganisms may be virus or bacteria, but are not limited thereto.

Engineered nuclease-induced mutations are detected by various methods, which include mismatch-sensitive Surveyor or T7 endonuclease I (T7E1) assays, RFLP analysis, fluorescent PCR, DNA melting analysis, and Sanger and deep sequencing. The T7E1 and Surveyor assays are widely used but often underestimate mutation frequencies because the assays detect heteroduplexes (formed by the hybridization of mutant and wild-type sequences or two different mutant sequences); they fail to detect homoduplexes formed by the hybridization of two identical mutant sequences. Thus, these assays cannot distinguish homozygous biallelic mutant clones from wild-type cells nor heterozygous biallelic mutants from heterozygous monoallelic mutants (FIG. 22). In addition, sequence polymorphisms near the nuclease target site can produce confounding results because the enzymes can cleave heteroduplexes formed by hybridization of these different wild-type alleles. RFLP analysis is free of these limitations and therefore is a method of choice. Indeed, RFLP analysis was one of the first methods used to detect

engineered nuclease-mediated mutations. Unfortunately, however, it is limited by the availability of appropriate restriction sites.

In accordance with another aspect of the invention, the present invention provides a kit for genotyping mutations or variations in an isolated biological sample, comprising the composition for genotyping mutations or variations in an isolated biological sample. In addition, the present invention provides a kit for genotyping nucleic acid sequences in pathogenic microorganisms in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence and Cas protein.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

In accordance with another aspect of the invention, the present invention provides a method of genotyping mutations or variations in an isolated biological sample, using the composition for genotyping mutations or variations in an isolated biological sample. In addition, the present invention provides a method of genotyping nucleic acid sequences in pathogenic microorganisms in an isolated biological sample, comprising a guide RNA specific for the target DNA sequence and Cas protein.

A guide RNA, Cas protein-encoding nucleic acid or Cas protein are as described in the above.

MODE FOR THE INVENTION

Hereinafter, the present invention will be described in more detail with reference to Examples. However, these Examples are for illustrative purposes only, and the invention is not intended to be limited by these Examples.

Example 1: Genome Editing Assay

1-1. DNA Cleavage Activity of Cas9 Protein

Firstly, the DNA cleavage activity of Cas9 derived from *Streptococcus pyogenes* in the presence or absence of a chimeric guide RNA in vitro was tested.

To this end, recombinant Cas9 protein that was expressed in and purified from *E. coli* was used to cleave a predigested or circular plasmid DNA that contained the 23-base pair (bp) human CCR5 target sequence. A Cas9 target sequence consists of a 20-bp DNA sequence complementary to crRNA or a chimeric guide RNA and the trinucleotide (5'-NGG-3') protospacer adjacent motif (PAM) recognized by Cas9 itself (FIG. 1A).

Specifically, the Cas9-coding sequence (4,104 bp), derived from *Streptococcus pyogenes* strain M1 GAS (NC_002737.1), was reconstituted using the human codon usage table and synthesized using oligonucleotides. First, 1-kb DNA segments were assembled using overlapping ~35-mer oligonucleotides and Phusion™ polymerase (New England Biolabs) and cloned into T-vector (SolGent). A full-length Cas9 sequence was assembled using four 1-kbp DNA segments by overlap PCR. The Cas9-encoding DNA segment was subcloned into p3s, which was derived from pcDNA3.1 (Invitrogen). In this vector, a peptide tag (NH₂-GGSGPPKKRKYVPYDVPDYA-COOH, SEQ ID NO: 2) containing the HA epitope and a nuclear localization signal (NLS) was added to the C-terminus of Cas9. Expression and nuclear localization of the Cas9 protein in HEK 293T cells were confirmed by western blotting using anti-HA antibody (Santa Cruz).

Then, the Cas9 cassette was subcloned into pET28-b(+) and transformed into BL21 (DE3). The expression of Cas9 was induced using 0.5 mM IPTG for 4 h at 25° C. The Cas9

protein containing the His6-tag at the C terminus was purified using Ni-NTA agarose resin (Qiagen) and dialyzed against 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT, and 10% glycerol (1). Purified Cas9 (50 nM) was incubated with super-coiled or pre-digested plasmid DNA (300 ng) and chimeric RNA (50 nM) in a reaction volume of 20 µl in NEB buffer 3 for 1 h at 37° C. Digested DNA was analyzed by electrophoresis using 0.8% agarose gels.

Cas9 cleaved the plasmid DNA efficiently at the expected position only in the presence of the synthetic RNA and did not cleave a control plasmid that lacked the target sequence (FIG. 1B).

1-2. DNA Cleavage by Cas9/guide RNA Complex in Human Cells

A RFP-GFP reporter was used to investigate whether the Cas9/guide RNA complex can cleave the target sequence incorporated between the RFP and GFP sequences in mammalian cells.

In this reporter, the GFP sequence is fused to the RFP sequence out-of-frame (2). The active GFP is expressed only when the target sequence is cleaved by site-specific nucleases, which causes frameshifting small insertions or deletions (indels) around the target sequence via error-prone non-homologous end-joining (NHEJ) repair of the double-strand break (DSB) (FIG. 2).

The RFP-GFP reporter plasmids used in this study were constructed as described previously (2). Oligonucleotides corresponding to target sites (Table 1) were synthesized (Macrogen) and annealed. The annealed oligonucleotides were ligated into a reporter vector digested with EcoRI and BamHI.

HEK 293T cells were co-transfected with Cas9-encoding plasmid (0.8 µg) and the RFP-GFP reporter plasmid (0.2 µg) in a 24-well plate using Lipofectamine™ 2000 (Invitrogen).

Meanwhile, the in vitro transcribed chimeric RNA had been prepared as follows. RNA was in vitro transcribed through run-off reactions using the MEGAscript™ T7 kit (Ambion) according to the manufacturer's manual. Templates for RNA in vitro transcription were generated by annealing two complementary single strand DNAs or by PCR amplification (Table 1). Transcribed RNA was resolved on a 8% denaturing urea-PAGE gel. The gel slice containing RNA was cut out and transferred to probe elution buffer. RNA was recovered in nuclease-free water followed by phenol:chloroform extraction, chloroform extraction, and ethanol precipitation. Purified RNAs were quantified by spectrometry.

At 12 h post transfection, chimeric RNA (1 µg) prepared by in vitro transcription was transfected using Lipofectamine 2000.

At 3d post-transfection, transfected cells were subjected to flow cytometry and cells expressing both RFP and GFP were counted.

It was found that GFP-expressing cells were obtained only when the cells were transfected first with the Cas9 plasmid and then with the guide RNA 12 h later (FIG. 2), demonstrating that RGENs could recognize and cleave the target DNA sequence in cultured human cells. Thus GFP-expressing cells were obtained by serial-transfection of the Cas9 plasmid and the guide RNA rather than co-transfection.

TABLE 1

Gene		sequence (5' to 3')	SEQ ID NO.
Oligonucleotides used for the construction of the reporter plasmid			
CCR5	F	AATTCATGACATCAATTATTATACATCGGAGGAG	3
	R	GATCCTCCTCCGATGATAATAATTGATGTCATG	4
Primers used in the T7E1 assay			
CCR5	F1	CTCCATGGTGCTATAGAGCA	5
	F2	GAGCCAAGCTCTCCATCTAGT	6
	R	GCCCTGTCAAGAGTTGACAC	7
C4BPB	F1	TATTTGGCTGGTTGAAAGGG	8
	R1	AAAGTCATGAAATAAACACACCCA	9
	F2	CTGCATTGATATGGTAGTACCATG	10
	R2	GCTGTTTCATTGCAATGGAATG	11
Primers used for the amplification of off-target sites			
ADCY5	F1	GCTCCACCTTAGTGCTCTG	12
	R1	GGTGGCAGGAACCTGTATGT	13
	F2	GTCATTGGCCAGAGATGTGGA	14
	R2	GTCCCATGACAGGCGTGAT	15
KCNJ6	F	GCCTGGCCAAGTTTCAGTTA	16
	R1	TGGAGCCATTGGTTTGATC	17
	R2	CCAGAACTAAGCCGTTTCTGAC	18
CNTNAP2	F1	ATCACCGACAACCAAGTTTCC	19
	F2	TGCAGTGCAGACTCTTTCCA	20
	R	AAGGACACAGGCAACTGAA	21
N/A Chr. 5	F1	TGTGGAACGAGTGGTGACAG	22
	R1	GCTGGATTAGGAGGCGAGATT	23
	F2	GTGCTGAGAACGCTTCATAGAG	24
	R2	GGACCAAAGGACATCTTCTCAC	25
Primers used for the detection of chromosomal deletions			
Deletion	F	CCACATCTCGTTCTCGGTTT	26
	R	TCACAAGCCACAGATATTT	27

1-3. Targeted Disruption of Endogenous Genes in Mammalian Cells by RGEN

To test whether RGENs could be used for targeted disruption of endogenous genes in mammalian cells, genomic DNA isolated from transfected cells using T7 endonuclease I (T7E1), a mismatch-sensitive endonuclease that specifically recognizes and cleaves heteroduplexes formed by the hybridization of wild-type and mutant DNA sequences was analyzed (3).

To introduce DSBs in mammalian cells using RGENs, 2×10^6 K562 cells were transfected with 20 μ g of Cas9-encoding plasmid using the 4D-Nucleofector™, SF Cell Line 4D-Nucleofector® X Kit, Program FF-120 (Lonza) according to the manufacturer's protocol. For this experiment, K562 (ATCC, CCL-243) cells were grown in RPMI-1640 with 10% FBS and the penicillin/streptomycin mix (100 U/ml and 100 μ g/ml, respectively).

After 24 h, 10-40 μ g of in vitro transcribed chimeric RNA was nucleofected into 1×10^6 K562 cells. The in vitro transcribed chimeric RNA had been prepared as described in the Example 1-2.

Cells were collected two days after RNA transfection and genomic DNA was isolated. The region including the target site was PCR-amplified using the primers described in Table 1. The amplicons were subjected to the T7E1 assay as described previously (3). For sequencing analysis, PCR products corresponding to genomic modifications were purified and cloned into the T-Blunt vector using the T-Blunt

PCR Cloning Kit (SolGent). Cloned products were sequenced using the M13 primer.

It was found that mutations were induced only when the cells were transfected serially with Cas9-encoding plasmid and then with guide RNA (FIG. 3). Mutation frequencies (Indels (%) in FIG. 3A) estimated from the relative DNA band intensities were RNA-dosage dependent, ranging from 1.3% to 5.1%. DNA sequencing analysis of the PCR amplicons corroborated the induction of RGEN-mediated mutations at the endogenous sites. Indels and microhomologies, characteristic of error-prone NHEJ, were observed at the target site. The mutation frequency measured by direct sequencing was 7.3% (=7 mutant clones/96 clones), on par with those obtained with zinc finger nucleases (ZFNs) or transcription-activator-like effector nucleases (TALENs).

Serial transfection of Cas9 plasmid and guide RNA was required to induce mutations in cells. But when plasmids that encode guide RNA, serial transfection was unnecessary and cells were co-transfected with Cas9 plasmid and guide RNA-encoding plasmid.

In the meantime, both ZFNs and TALENs have been successfully developed to disrupt the human CCR5 gene (3-6), which encodes a G-protein-coupled chemokine receptor, an essential co-receptor of HIV infection. A CCR5-specific ZEN is now under clinical investigation in the US for the treatment of AIDS (7). These ZFNs and TALENs, however, have off-target effects, inducing both local mutations at sites whose sequences are homologous to the on-target sequence (6, 8-10) and genome rearrangements that arise from the repair of two concurrent DSBs induced at on-target and off-target sites (11-12). The most striking off-target sites associated with these CCR5-specific engineered nucleases reside in the CCR2 locus, a close homolog of CCR5, located 15-kbp upstream of CCR5. To avoid off-target mutations in the CCR2 gene and unwanted deletions, inversions, and duplications of the 15-kbp chromosomal segment between the CCR5 on-target and CCR2 off-target sites, the present inventors intentionally chose the target site of our CCR5-specific RGEN to recognize a region within the CCR5 sequence that has no apparent homology with the CCR2 sequence.

The present inventors investigated whether the CCR5-specific RGEN had off-target effects. To this end, we searched for potential off-target sites in the human genome by identifying sites that are most homologous to the intended 23-bp target sequence. As expected, no such sites were found in the CCR2 gene. Instead, four sites, each of which carries 3-base mismatches with the on-target site, were found (FIG. 4A). The T7E1 assays showed that mutations were not detected at these sites (assay sensitivity, ~0.5%), demonstrating exquisite specificities of RGENS (FIG. 4B). Furthermore, PCR was used to detect the induction of chromosomal deletions in cells separately transfected with plasmids encoding the ZEN and RGEN specific to CCR5. Whereas the ZFN induced deletions, the RGEN did not (FIG. 4C).

Next, RGENs was reprogrammed by replacing the CCR5-specific guide RNA with a newly-synthesized RNA designed to target the human C4BPB gene, which encodes the beta chain of C4b-binding protein, a transcription factor. This RGEN induced mutations at the chromosomal target site in K562 cells at high frequencies (FIG. 3B). Mutation frequencies measured by the T7E1 assay and by direct sequencing were 14% and 8.3% (=4 mutant clones/48 clones), respectively. Out of four mutant sequences, two clones contained a single-base or two-base insertion precisely at the cleavage site, a pattern that was also observed

at the CCR5 target site. These results indicate that RGENs cleave chromosomal target DNA at expected positions in cells.

Example 2: Proteinaceous RGEN-Mediated Genome Editing

RGENs can be delivered into cells in many different forms. RGENs consist of Cas9 protein, crRNA, and tracrRNA. The two RNAs can be fused to form a single-chain guide RNA (sgRNA). A plasmid that encodes Cas9 under a promoter such as CMV or CAG can be transfected into cells. crRNA, tracrRNA, or sgRNA can also be expressed in cells using plasmids that encode these RNAs. Use of plasmids, however, often results in integration of the whole or part of the plasmids in the host genome. The

As shown in FIG. 10, Cas9 protein/sgRNA complex induced targeted mutation at the CCR5 locus at frequencies that ranged from 4.8 to 38% in a sgRNA or Cas9 protein dose-dependent manner, on par with the frequency obtained with Cas9 plasmid transfection (45%). Cas9 protein/crRNA/tracrRNA complex was able to induce mutations at a frequency of 9.4%. Cas9 protein alone failed to induce mutations. When 2×10^5 cells were transfected with $\frac{1}{5}$ scale-downed doses of Cas9 protein and sgRNA, mutation frequencies at the CCR5 locus ranged from 2.7 to 57% in a dose-dependent manner, greater than that obtained with co-transfection of Cas9 plasmid and sgRNA plasmid (32%).

We also tested Cas9 protein/sgRNA complex that targets the ABCC11 gene and found that this complex induced indels at a frequency of 35%, demonstrating general utility of this method.

TABLE 2

Sequences of guide RNA		
RNA Target type	RNA sequence (5' to 3')	SEQ ID Length NO
CCR5 sgRNA	GGUGACAUCAAUUUUUAUCAUGUUUUUAGAGCUAG104 AAAUAGCAAGUUAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	bp 28
crRNA	GGUGACAUCAAUUUUUAUCAUGUUUUUAGAGCUAU GCUGUUUUU	44 bp 29
tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGUUAAAUA GGCUAGUCGGUUUUAUCAUUGAAAAAGUGGCACCG AGUCGGUGCUUUUUUU	86 bp 30

bacterial sequences incorporated in plasmid DNA can cause unwanted immune response in vivo. Cells transfected with plasmid for cell therapy or animals and plants derived from DNA-transfected cells must go through a costly and lengthy regulation procedure before market approval in most developed countries. Furthermore, plasmid DNA can persist in cells for several days post-transfection, aggravating off-target effects of RGENS.

Here, we used recombinant Cas9 protein complexed with in vitro transcribed guide RNA to induce targeted disruption of endogenous genes in human cells. Recombinant Cas9 protein fused with the hexa-histidine tag was expressed in and purified from *E. coli* using standard Ni ion affinity chromatography and gel filtration. Purified recombinant Cas9 protein was concentrated in storage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol). Cas9 protein/sgRNA complex was introduced directly into K562 cells by nucleofection: 1×10^6 K562 cells were transfected with 22.5-225 (1.4-14 μ M) of Cas9 protein mixed with 100 μ g (29 μ M) of in vitro transcribed sgRNA (or crRNA 40 μ g and tracrRNA 80 μ g) in 100 μ l solution using the 4D-Nucleofector, SF Cell Line 4D-Nucleofector X Kit, Program FF-120 (Lonza) according to the manufacturer's protocol. After nucleofection, cells were placed in growth media in 6-well plates and incubated for 48 hr. When 2×10^5 K562 cells were transfected with $\frac{1}{5}$ scale-downed protocol, 4.5-45 μ g of Cas9 protein mixed with 6-60 μ g of in vitro transcribed sgRNA (or crRNA 8 μ g and tracrRNA 16 μ g) were used and nucleofected in 20 μ l solution. Nucleofected cells were then placed in growth media in 48-well plates. After 48 hr, cells were collected and genomic DNA was isolated. The genomic DNA region spanning the target site was PCR-amplified and subjected to the T7E1 assay.

Example 3: RNA-Guided Genome Editing in Mice

To examine the gene-targeting potential of RGENs in pronuclear (PN)-stage mouse embryos, the forkhead box N1 (Foxn1) gene, which is important for thymus development and keratinocyte differentiation (Nehls et al., 1996), and the protein kinase, DNA activated, catalytic polypeptide (Prkdc) gene, which encodes an enzyme critical for DNA DSB repair and recombination (Taccioli et al., 1998) were used.

To evaluate the genome-editing activity of the Foxn1-RGEN, we injected Cas9 mRNA (10-ng/ μ l solution) with various doses of the sgRNA (FIG. 5a) into the cytoplasm of PN-stage mouse embryos, and conducted T7 endonuclease I (T7E1) assays (Kim et al. 2009) using genomic DNAs obtained from in vitro cultivated embryos (FIG. 6a).

Alternatively, we directly injected the RGEN in the form of recombinant Cas9 protein (0.3 to 30 ng/ μ l) complexed with the two-fold molar excess of Foxn1-specific sgRNA (0.14 to 14 ng/ μ l) into the cytoplasm or pronucleus of one-cell mouse embryos, and analyzed mutations in the Foxn1 gene using in vitro cultivated embryos (FIG. 7).

Specifically, Cas9 mRNA and sgRNAs were synthesized in vitro from linear DNA templates using the mMACHINE T7 Ultra kit (Ambion) and MEGAShortscript T7 kit (Ambion), respectively, according to the manufacturers' instructions, and were diluted with appropriate amounts of diethyl pyrocarbonate (DEPC, Sigma)-treated injection buffer (0.25 mM EDTA, 10 mM Tris, pH 7.4). Templates for sgRNA synthesis were generated using oligonucleotides listed in Table 3. Recombinant Cas9 protein was obtained from ToolGen, Inc.

25

TABLE 3

RNA Name	Direction	Sequence (5' to 3')	SEQ ID NO
Foxn1 #1 sgRNA	F	GAAATTAATACGACTCACTATAGGCAGTCTGAGC TCACACTTCCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	31
Foxn1 #2 sgRNA	F	GAAATTAATACGACTCACTATAGGACTTCCAGGC TCCACCCGACGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	32
Foxn1 #3 sgRNA	F	GAAATTAATACGACTCACTATAGGCCAGGCTCCA CCCGACTGGAGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	33
Foxn1 #4 sgRNA	F	GAAATTAATACGACTCACTATAGGACTGGAGGGC GAACCCCAAGGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	34
Foxn1 #5 sgRNA	F	GAAATTAATACGACTCACTATAGGACCCCAAGGG GACCTCATGCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	35
Prkdc #1 sgRNA	F	GAAATTAATACGACTCACTATAGGTTAGTTTTTT CCAGAGACTTGTGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	36
Prkdc #2 sgRNA	F	GAAATTAATACGACTCACTATAGGTTGGTTGCT TGTGTTTATCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	37
Prkdc #3 sgRNA	F	GAAATTAATACGACTCACTATAGGCACAAGCAA CCAAAGTCTCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	38
Prkdc #4 sgRNA	F	GAAATTAATACGACTCACTATAGGCCCTCAATGCT AAGCGACTTCGTTTTAGAGCTAGAAATAGCAAGT TAAAATAAGGCTAGTCCG	39

All animal experiments were performed in accordance with the Korean Food and Drug Administration (KFDA) guidelines. Protocols were reviewed and approved by the Institutional Animal Care and Use Committees (IACUC) of the Laboratory Animal Research Center at Yonsei University (Permit Number: 2013-0099). All mice were maintained in the specific pathogen-free facility of the Yonsei Laboratory Animal Research Center. FVB/NTac (Taconic) and ICR mouse strains were used as embryo donors and foster mothers, respectively. Female FVB/NTac mice (7-8 weeks old) were super-ovulated by intra-peritoneal injections of 5 IU pregnant mare serum gonadotropin (PMSG, Sigma) and 5 IU human chorionic gonadotropin (hCG, Sigma) at 48-hour intervals. The super-ovulated female mice were mated to FVB/NTac stud males, and fertilized embryos were collected from oviducts.

Cas9 mRNA and sgRNAs in M2 medium (Sigma) were injected into the cytoplasm of fertilized eggs with well-recognized pronuclei using a Piezo-driven micromanipulator (Prime Tech).

In the case of injection of recombinant Cas9 protein, the recombinant Cas9 protein: Foxn1-sgRNA complex was diluted with DEPC-treated injection buffer (0.25 mM EDTA, 10 mM Tris, pH 7.4) and injected into male pronuclei using a TransferMan NK2 micromanipulator and a Femto-Jet® microinjector (Eppendorf).

The manipulated embryos were transferred into the oviducts of pseudopregnant foster mothers to produce live animals, or were cultivated in vitro for further analyses.

To screen F0 mice and in vitro cultivated mouse embryos with RGEN-induced mutations, T7E1 assays were per-

26

formed as previously described using genomic DNA samples from tail biopsies and lysates of whole embryos (Cho et al., 2013).

Briefly, the genomic region encompassing the RGEN target site was PCR-amplified, melted, and re-annealed to form heteroduplex DNA, which was treated with T7 endonuclease 1 (New England Biolabs), and then analyzed by agarose gel electrophoresis. Potential off-target sites were identified by searching with bowtie 0.12.9 and were also similarly monitored by T7E1 assays. The primer pairs used in these assays were listed in Tables 4 and 5.

TABLE 4

Primers used in the T7E1 assay			
Gene	Direction	Sequence (5' to 3')	SEQ ID NO
Foxn1	F1	GTCTGTCTATCATCTCTCCCTTCTCTCC	40
	F2	TCCTAATCCGATGGCTAGCTCCAG	41
	R1	ACGAGCAGCTGAAGTTAGCATGC	42
	R2	CTACTCAATGCTCTTAGAGCTACCAGGCTTGC	43
Prkdc	F	GACTGTGTGGGGAGGGCCG	44
	F2	GGGAGGGCCGAAGCTTATTTTTG	45
	R1	CCTGAAGACTGAAGTTGGCAGAAGTGA	46
	R2	CTTAGGGCTTCTCTACAATCAG	47

TABLE 5

Primers used for amplification of off-target sites				
Gene	Notation	Direction	Sequence (5' to 3')	SEQ ID NO
Foxn1	off 1	F	CTCGTGTGTAGCCCTGAC	48
		R	AGACTGGCTGGAACTCACAG	49
	off 2	F	CACTAAAGCCTGTCAGGAAGCCG	50
		R	CTGTGGAGAGCACACAGCAGC	51
	off 3	F	GCTGCGACCTGAGACCATG	52
		R	CTTCAATGGCTTCTGCTTAGGCTAC	53
	off 4	F	GGTTCAGATGAGGCCATCCTTTC	54
		R	CCTGATCTGCAGGCTTAACCCCTTG	55
Prkdc	off 1	F	CTCACCTGCACATCATGTGG	56
		R	GGCATCCACCCTATGGGGTC	57
	off 2	F	GCCTTGACCTAGAGCTTAAAGAGCC	58
		R	GGTCTTGTAGCAGGAAGGACACTG	59
	off 3	F	AAAACCTGCTTGATGGGATATGTGG	60
		R	CTCTCACTGGTTATCTGTGCTCCTTC	61
	off 4	F	GGATCAATAGGTGGTGGGGGATG	62
		R	GTGAATGACACAATGTGACAGCTTCA	63
	off 5	F	CACAAGACAGACCTCTCAACATTCAG	64
		R	GTGCATGCATATAATCCATTCTGATT	65
	off 6	F1	GGGAGGCAGAGGCAGGT	66
		F2	GGATCTGTGAGTTGAGGCCA	67
		R1	GCTCCAGAACCTACTCTTAGGCTC	68

Mutant founders identified by the T7E1 assay were further analyzed by fPCR. Appropriate regions of genomic DNA were sequenced as described previously (Sung et al.,

2013). For routine PCR genotyping of F1 progenies, the following primer pairs were used for both wild-type and mutant alleles:

(SEQ ID NO: 69)
5'-CTACTCCCTCCGAGTCTGA-3'
and
(SEQ ID NO: 70)
5'-CCAGGCCTAGGTTCCAGGTA-3',
for the Foxn1 gene
(SEQ ID NO: 71)
5'-CCCCAGCATTGCAGATTTC-3'
and
(SEQ ID NO: 72)
5'-AGGGCTTCTTCTCTACAATCACG-3'.
for Prkdc gene

In the case of injection of Cas9 mRNA, mutant fractions (the number of mutant embryos/the number of total embryos) were dose-dependent, ranging from 33% (1 ng/μl sgRNA) to 91% (100 ng/μl) (FIG. 6b). Sequence analysis confirmed mutations in the Foxn1 gene; most mutations

were small deletions (FIG. 6c), reminiscent of those induced by ZFNs and TALENS (Kim et al., 2013).

In the case of injection of Cas9 protein, these injection doses and methods minimally affected the survival and development of mouse embryos in vitro: over 70% of RGEN-injected embryos hatched out normally in both experiments. Again, mutant fractions obtained with Cas9 protein injection were dose-dependent, and reached up to 88% at the highest dose via pronucleus injection and to 71% via intra-cytoplasmic injection (FIGS. 7a and 7b). Similar to the mutation patterns induced by Cas9 mRNA plus sgRNA (FIG. 6c), those induced by the Cas9 protein-sgRNA complex were mostly small deletions (FIG. 7c). These results clearly demonstrate that RGENs have high gene-targeting activity in mouse embryos.

Encouraged by the high mutant frequencies and low cytotoxicity induced by RGENS, we produced live animals by transferring the mouse embryos into the oviducts of pseudopregnant foster mothers.

Notably, the birth rates were very high, ranging from 58% to 73%, and were not affected by the increasing doses of Foxn1-sgRNA (Table 6).

TABLE 6

RGEN-mediated gene-targeting in FVB/NTaC mice						
Target Gene	Cas9 mRNA + sgRNA (ng/μl)	Injected embryos	Transfer red embryos (%)	Total newborns (%)	Live newborns * (%)	Founders † (%)
Foxn1	10 + 1	76	62 (82)	45 (73)	31 (50)	12 (39)
	10 + 10	104	90 (87)	52 (58)	58 (64)	33 (57)
	10 + 100	100	90 (90)	62 (69)	58 (64)	54 (93)
Prkdc	Total	280	242 (86)	159 (66)	147 (61)	99 (67)
	50 + 50	73	58 (79)	35 (60)	33 (57)	11 (33)
	50 + 100	79	59 (75)	22 (37)	21 (36)	7 (33)
	50 + 250	94	73 (78)	37 (51)	37 (51)	21 (57)
Total	246	190 (77)	94 (49)	91 (48)	39 (43)	

40

Out of 147 newborns, we obtained 99 mutant founder mice. Consistent with the results observed in cultivated embryos (FIG. 6b), mutant fractions were proportional to the doses of Foxn1-sgRNA, and reached up to 938 (100 ng/μl Foxn1-sgRNA) (Tables 6 and 7, FIG. 5b).

TABLE 7

DNA sequences of Foxn1 mutant alleles identified from a subset of T7E1-positive mutant founders			
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 134)	del + ins	#	Founder mice
ACTTCCAGGC-----AACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 297)	Δ19	1	20
ACTTCCAGGC-----GAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 298)	Δ18	1	115
ACTTCCAGGCTCC----- (SEQ ID NO: 299)	Δ60	1	19
ACTTCCAGGCTCC----- (SEQ ID NO: 300)	Δ44	1	108
ACTTCCAGGCTCC-----CAAGGGGA CCTCATGCAGG (SEQ ID NO: 301)	Δ21	1	64

TABLE 7-continued

DNA sequences of Foxn1 mutant alleles identified from a subset of T7E1-positive mutant founders			
	del + ins	#	Founder mice
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 134)			
ACTTCCAGGCTCC-----TTAGGAGGGCGAACCCCA AGGGGACCTCA (SEQ ID NO: 302)	Δ12 + 6	1	126
ACTTCCAGGCTCCACC----- --TCATGCAGG (SEQ ID NO: 303)	Δ28	1	5
ACTTCCAGGCTCCACCC-----CCAA GGGACCTCATG (SEQ ID NO: 304)	Δ21 + 4	1	61
ACTTCCAGGCTCCACCC-----AAGGGGA CCTCATGCAGG (SEQ ID NO: 305)	Δ18	2	95, 29
ACTTCCAGGCTCCACCC-----CAAGGGGA CCTCATGCAGG (SEQ ID NO: 306)	Δ17	7	12, 14, 27, 66, 108, 114, 126
ACTTCCAGGCTCCACCC-----ACCCAAGGGG ACCTCATGCAG (SEQ ID NO: 307)	Δ15 + 1	1	32
ACTTCCAGGCTCCACCC-----CACCCAAGGG GACCTCATGCA (SEQ ID NO: 308)	Δ15 + 2	1	124
ACTTCCAGGCTCCACCC-----ACCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 309)	Δ13	1	32
ACTTCCAGGCTCCACCC-----GGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 310)	Δ8	1	110
ACTTCCAGGCTCCACCCCT-----GGGGA CCTCATGCAGG (SEQ ID NO: 311)	Δ20 + 1	1	29
ACTTCCAGGCTCCACCCG-----AACCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 312)	Δ11	1	111
ACTTCCAGGCTCCACCCGA-----A CCTCATGCAGG (SEQ ID NO: 313)	Δ22	1	79
ACTTCCAGGCTCCACCCGA-----GGGGA CCTCATGCAGG (SEQ ID NO: 314)	Δ18	2	13, 127
ACTTCCAGGCTCCACCCCA-----AGGGGA CCTCATGCAGG (SEQ ID NO: 315)	Δ17	1	24
ACTTCCAGGCTCCACCCGA-----ACCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 316)	Δ11	5	14, 53, 58, 69, 124
ACTTCCAGGCTCCACCCGA-----GACCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 317)	Δ10	1	14
ACTTCCAGGCTCCACCCGA-----GGGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 318)	Δ5	3	53, 79, 115
ACTTCCAGGCTCCACCCGAC----- -CTCATGCAGG (SEQ ID NO: 319)	Δ23	1	108
ACTTCCAGGCTCCACCCGAC-----CCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 320)	Δ11	1	3
ACTTCCAGGCTCCACCCGAC-----GAAGGGCCCCA AGGGGACCTCA (SEQ ID NO: 321)	Δ11 + 6	1	66
ACTTCCAGGCTCCACCCGAC-----GAACCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 322)	Δ8	2	3, 66
ACTTCCAGGCTCCACCCGAC-----GGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 323)	Δ5	1	27
ACTTCCAGGCTCCACCCGAC--GTGCTTGAGGGCGAACCCCA AGGGGACCTCA (SEQ ID NO: 324)	Δ2 + 6	2	5
ACTTCCAGGCTCCACCCGACT-----CACTATCTTCTGGGC TCCTCCATGTC (SEQ ID NO: 325)	Δ6 + 25	2	21, 114

TABLE 7-continued

DNA sequences of Foxn1 mutant alleles identified from a subset of T7E1-positive mutant founders			
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGA CCTCATGCAGG (SEQ ID NO: 134)	del + ins	#	Founder mice
ACTTCCAGGCTCCACCCGACT---TGGCGAACCCCAAGGGG ACCTCATGCAG (SEQ ID NO: 326)	Δ4 + 1	1	53
ACTTCCAGGCTCCACCCGACT--TGCGAGGGCGAACCCCAAGG GGACCTCATGC (SEQ ID NO: 327)	Δ2 + 3	1	126
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGG ACCTCATGCAG (SEQ ID NO: 328)	+1	15	3, 5, 12, 19, 29, 55, 56, 61, 66, 68, 81, 108, 111, 124, 127
ACTTCCAGGCTCCACCCGACTTTGGAGGGCGAACCCCAAGGG GACCTCATGCA (SEQ ID NO: 329)	+2	2	79, 120
ACTTCCAGGCTCCACCCGACTGTTGGAGGGCGAACCCCAAGG GGACCTCATGC (SEQ ID NO: 330)	+3	1	55
ACTTCCAGGCTCCACCCGACTGGAG (+455) GGCGAACCCCA AGGGGACCTCC (SEQ ID NO: 331)	+455	1	13

To generate Prkdc-targeted mice, we applied a 5-fold higher concentration of Cas9 mRNA (50 ng/μl) with increasing doses of Prkdc-sgRNA (50, 100, and 250 ng/μl). Again, the birth rates were very high, ranging from 51% to 60%, enough to produce a sufficient number of newborns for the analysis (Table 6). The mutant fraction was 57% (21 mutant founders among 37 newborns) at the maximum dose of Prkdc-sgRNA. These birth rates obtained with RGENs were approximately 2- to 10-fold higher than those with TALENs reported in our previous study (Sung et al., 2013). These results demonstrate that RGENs are potent gene-targeting reagents with minimal toxicity.

To test the germ-line transmission of the mutant alleles, we crossed the Foxn1 mutant founder #108, a mosaic with four different alleles (FIG. 5c, and Table 8) with wild-type mice, and monitored the genotypes of F1 offspring.

TABLE 8

Genotypes of Foxn1 mutant mice			
Founder NO.	sgRNA (ng/ml)	Genotyping Summary	Detected alleles
58*	1	not determined	Δ11
19	100	bi-allelic	Δ60/+1
20	100	bi-allelic	Δ67/Δ19
13	100	bi-allelic	Δ18/+455
32	10	bi-allelic (heterozygote)	Δ13/Δ15 + 1
115	10	bi-allelic (heterozygote)	Δ18/Δ5
111	10	bi-allelic (heterozygote)	Δ11/+1
110	10	bi-allelic (homozygote)	Δ8/Δ8
120	10	bi-allelic (homozygote)	+2/+2
81	100	heterozygote	+1/WT
69	100	homozygote	Δ11/Δ11
55	1	mosaic	Δ18/Δ1/+1/+3
56	1	mosaic	Δ127/Δ41/Δ2/+1
127	1	mosaic	Δ18/+1/WT

TABLE 8-continued

Genotypes of Foxn1 mutant mice			
Founder NO.	sgRNA (ng/ml)	Genotyping Summary	Detected alleles
53	1	mosaic	Δ11/Δ5/Δ4 + 1/WT
27	10	mosaic	Δ17/Δ5/WT
29	10	mosaic	Δ18/Δ20 + 1/+1
95	10	mosaic	Δ18/Δ14/Δ8/Δ4
108	10	mosaic	+1/Δ17/Δ23/Δ44
114	10	mosaic	Δ17/Δ8/Δ6 + 25
124	10	mosaic	Δ11/Δ15 + 2/+1
126	10	mosaic	Δ17/Δ2 + 3/Δ12 + 6
12	100	mosaic	Δ30/Δ28/Δ17/+1
5	100	mosaic	Δ28/Δ11/Δ2 + 6/+1
14	100	mosaic	Δ17/Δ11/Δ10
21	100	mosaic	Δ127/Δ41/Δ2/Δ6 + 25
24	100	mosaic	Δ17/+1/WT
64	100	mosaic	Δ31/Δ21/+1/WT
68	100	mosaic	Δ17/Δ11/+1/WT
79	100	mosaic	Δ22/Δ5/+2/WT
61	100	mosaic	Δ21 + 4/Δ6/+1/+9
66**	100	mosaic	Δ17/Δ8/Δ11 + 6/+1/WT
3	100	mosaic	Δ11/Δ8/+1

50 Underlined alleles were sequenced. Alleles in red, detected by sequencing, but not by PCR.
*only one clone sequenced.
**Not determined by PCR.

As expected, all the progenies were heterozygous mutants possessing the wild-type allele and one of the mutant alleles (FIG. 5d). We also confirmed the germ-line transmission in independent founder mice of Foxn1 (FIG. 8) and Prkdc (FIG. 9). To the best of our knowledge, these results provide the first evidence that RGEN-induced mutant alleles are stably transmitted to F1 progenies in animals.

Example 4: RNA-Guided Genome Editing in Plants

4-1. Production of Cas9 Protein

65 The Cas9 coding sequence (4104 bps), derived from *Streptococcus pyogenes* strain M1 GAS (NC_002737.1), was cloned to pET28-b(+) plasmid. A nuclear targeting

sequence (NLS) was included at the protein N terminus to ensure the localization of the protein to the nucleus. pET28-b(+) plasmid containing Cas9 ORF was transformed into BL21 (DE3). Cas9 was then induced using 0.2 mM IPTG for 16 hrs at 18° C. and purified using Ni-NTA agarose beads (Qiagen) following the manufacturer's instructions. Purified Cas9 protein was concentrated using Ultracel-100K (Millipore).

4-2. Production of Guide RNA

The genomic sequence of the Arabidopsis gene encoding the BRI1 was screened for the presence of a NGG motif, the so called protospacer adjacent motif (PAM), in an exon which is required for Cas9 targeting. To disrupt the BRI1 gene in Arabidopsis, we identified two RGEN target sites in an exon that contain the NGG motif. sgRNAs were produced in vitro using template DNA. Each template DNA was generated by extension with two partially overlapped oligonucleotides (Macrogen, Table X1) and Phusion polymerase (Thermo Scientific) using the following conditions—98° C. 30 sec {98° C. 10 sec, 54° C. 20 sec, 72° C. 2 min}×20, 72° C. 5 min.

TABLE 9

Oligonucleotides for the production of the template DNA for in vitro transcription		
Oligonucleotides	Sequence (5'-3')	SEQ ID NO
BRI1 target 1 (Forward)	GAAATTAATACGACTCACTATAGGTTGAAAGAT GGAAGCGCGGTTTATAGACTAGAAATAGCAAGT TAAATAAGGCTAGTCCG	73
BRI1 target 2 (Forward)	GAAATTAATACGACTCACTATAGGTGAAACTAAA CTGGTCCACAGTTTATAGACTAGAAATAGCAAGT TAAATAAGGCTAGTCCG	74
Universal (Reverse)	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGT TGATAACGGACTAGCCTTATTTAACTTGC	75

The extended DNA was purified and used as a template for the in vitro production of the guide RNA's using the MEGAshortscript T7 kit (Life Technologies). Guide RNA were then purified by Phenol/Chloroform extraction and ethanol precipitation. To prepare Cas9/sgRNA complexes, 10 µl of purified Cas9 protein (12 µg/µl) and 4 µl each of two sgRNAs (11 µg/µl) were mixed in 20 µl NEB3 buffer (New England Biolabs) and incubated for 10 min at 37° C.

4-3. Transfection of Cas9/sgRNA Complex to Protoplast

The leaves of 4-week-old Arabidopsis seedlings grown aseptically in petri dishes were digested in enzyme solution (1% cellulose R10, 0.5% macerozyme R10, 450 mM mannitol, 20 mM MES pH 5.7 and CPW salt) for 8–16 hrs at 25° C. with 40 rpm shaking in the dark. Enzyme/protoplast solutions were filtered and centrifuged at 100×g for 3–5 min. Protoplasts were re-suspended in CPW solution after counting cells under the microscope (×100) using a hemacytometer. Finally, protoplasts were re-suspended at 1×10⁶/ml in MMG solution (4 mM HEPES pH 5.7, 400 mM mannitol and 15 mM MgCl₂). To transfect the protoplasts with Cas9/sgRNA complex, 200 µL (200,000 protoplasts) of the protoplast suspension were gently mixed with 3.3 or 10 µL of Cas9/sgRNA complex [Cas9 protein (6 µg/µL) and two sgRNAs (2.2 µg/µL each)] and 200 µL of 40% polyethylene glycol transfection buffer (40% PEG4000, 200 mM mannitol and 100 mM CaCl₂) in 2 ml tubes. After 5–20 min

incubation at room temperature, transfection was stopped by adding wash buffer with W5 solution (2 mM MES pH 5.7, 154 mM NaCl, 125 mM CaCl₂ and 5 mM KCl). Protoplasts were then collected by centrifugation for 5 min at 100×g, washed with 1 ml of W5 solution, centrifuged for another 5 min at 100×g. The density of protoplasts was adjusted to 1×10⁵/ml and they were cultured in modified KM 8p liquid medium with 400 mM glucose.

4-4. Detection of Mutations in Arabidopsis Protoplasts and Plants

After 24 hr or 72 hr post-transfection, protoplasts were collected and genomic DNA was isolated. The genomic DNA region spanning the two target sites was PCR-amplified and subjected to the T7E1 assay. As shown in FIG. 11, indels were induced by RGENs at high frequencies that ranged from 50% to 70%. Surprisingly, mutations were induced at 24 hr post-transfection. Apparently, Cas9 protein functions immediately after transfection. PCR products were purified and cloned into T-Blunt PCR Cloning Kit (Solgent). Plasmids were purified and subjected to Sanger sequencing with M13F primer. One mutant sequence had a 7-bp deletion at one site (FIG. 12). The other three mutant sequences had deletions of ~220-bp DNA segments between the two RGEN site.

Example 5: Cas9 Protein Transduction Using a Cell-Penetrating Peptide or Protein Transduction Domain

5-1. Construction of His-Cas9-Encoding Plasmid

Cas9 with a cysteine at the C-terminal was prepared by PCR amplification using the previously described Cas9 plasmid {Cho, 2013 #166} as the template and cloned into pET28-(a) vector (Novagen, Merck Millipore, Germany) containing His-tag at the N-terminus.

5-2. Cell Culture

293T (Human embryonic kidney cell line), and HeLa (human ovarian cancer cell line) were grown in DMEM (GIBCO-BRL Rockville) supplemented with 10% FBS and 1% penicillin and streptomycin.

5-3. Expression and Purification of Cas9 Protein

To express the Cas9 protein, *E. coli* BL21 cells were transformed with the pET28-(a) vector encoding Cas9 and plated onto Luria-Bertani (LB) agar medium containing 50 µg/mL kanamycin (Amresco, Solon, OH). Next day, a single colony was picked and cultured in LB broth containing 50 µg/mL kanamycin at 37° C. overnight. Following day, this starter culture at 0.1 OD₆₀₀ was inoculated into Luria broth containing 50 µg/mL kanamycin and incubated for 2 hrs at 37° C. until OD₆₀₀ reached to 0.6–0.8. To induce Cas9 protein expression, the cells were cultured at 30° C. overnight after addition of isopropyl-β-D-thiogalactopyranoside (IPTG) (Promega, Madison, WI) to the final concentration of 0.5 mM.

The cells were collected by centrifugation at 4000 rpm for 15–20 mins, resuspended in a lysis buffer (20 mM Tris-Cl pH8.0, 300 mM NaCl, 20 mM imidazole, 1× protease inhibitor cocktail, 1 mg/ml lysozyme), and lysed by sonication (40% duty, 10 sec pulse, 30 sec rest, for 10 mins on ice). The soluble fraction was separated as the supernatant after centrifugation at 15,000 rpm for 20 mins at 4° C. Cas9 protein was purified at 4° C. using a column containing Ni-NTA agarose resin (QIAGEN) and AKTA prime instrument (AKTA prime, GE Healthcare, UK). During this chromatography step, soluble protein fractions were loaded onto Ni-NTA agarose resin column (GE Healthcare, UK) at the flow rate of 1 mL/min. The column was washed with a

washing buffer (20 mM Tris-Cl pH8.0, 300 mM NaCl, 20 mM imidazole, 1× protease inhibitor cocktail) and the bound protein was eluted at the flow rate of 0.5 ml/min with an elution buffer (20 mM Tris-Cl pH8.0, 300 mM NaCl, 250 mM imidazole, 1× protease inhibitor cocktail). The pooled eluted fraction was concentrated and dialyzed against storage buffer (50 mM Tris-HCl, pH8.0, 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 20% Glycerol). Protein concentration was quantitated by Bradford assay (Biorad, Hercules, CA) and purity was analyzed by SDS-PAGE using bovine serum albumin as the control.

5-4. Conjugation of Cas9 to 9R4L

1 mg Cas9 protein diluted in PBS at the concentration of 1 mg/mL and 50 µg of maleimide-9R4L peptide in 25 µL DW (Peptron, Korea) were gently mixed using a rotor at room temperature for 2 hrs and at 4° C. overnight. To remove unconjugated maleimide-9R4L, the samples were dialyzed using 50 kDa molecular weight cutoff membrane against of DPBS (pH 7.4) at 4° C. for 24 hrs. Cas9-9R4L protein was collected from the dialysis membrane and the protein amount was determined using Bradford assay.

5-5. Preparation of sgRNA-9R4L

sgRNA (1 µg) was gently added to various amounts of C9R4LC peptide (ranging from 1 to 40 weight ratio) in 100 µL of DPBS (pH 7.4). This mixture was incubated at room temperature for 30 mins and diluted to 10 folds using RNase-free deionized water. The hydrodynamic diameter and z-potential of the formed nanoparticles were measured using dynamic light scattering (Zetasizer-nano analyzer ZS; Malvern instruments, Worcestershire, UK).

5-6. Cas9 Protein and sgRNA Treatments

Cas9-9R4L and sgRNA-C9R4LC were treated to the cells as follows: 1 µg of sgRNA and 15 µg of C9R4LC peptide were added to 250 mL of OPTIMEM medium and incubated at room temperature for 30 mins. At 24 hrs after seeding, cells were washed with OPTIMEM medium and treated with sgRNA-C9R4LC complex for 4 hrs at 37° C. Cells were washed again with OPTIMEM medium and treated with Cas9-9R4L for 2 hrs at 37° C. After treatment, culture media was replaced with serum-containing complete medium and incubated at 37° C. for 24 hrs before the next treatment. Same procedure was followed for multiple treatments of Cas9 and sgRNA for three consecutive days.

5-7. Cas9-9R4L and sgRNA-9R4L can Edit Endogenous Genes in Cultured Mammalian Cells without the Use of Additional Delivery Tools

To determine whether Cas9-9R4L and sgRNA-9R4L can edit endogenous genes in cultured mammalian cells without the use of additional delivery tools, we treated 293 cells with Cas9-9R4L and sgRNA-9R4L targeting the CCR5 gene and analyzed the genomic DNA. T7E1 assay showed that 9% of CCR5 gene was disrupted in cells treated with both Cas9-9R4L and sgRNA-9R4L and that the CCR5 gene disruption was not observed in control cells including those untreated, treated with either Cas9-9R or sgRNA-9R4L, or treated with both unmodified Cas-9 and sgRNA (FIG. 13), suggesting that the treatment with Cas9-9R4L protein and sgRNA conjugated with 9R4L, but not unmodified Cas9 and sgRNA, can lead to efficient genome editing in mammalian cells.

Example 6: Control of Off-Target Mutation According to Guide RNA Structure

Recently, three groups reported that RGENs had off-target effects in human cells. To our surprise, RGENs induced mutations efficiently at off-target sites that differ by 3 to 5

nucleotides from on-target sites. We noticed, however, that there were several differences between our RGENs and those used by others. First, we used dualRNA, which is crRNA plus tracrRNA, rather than single-guide RNA (sgRNA) that is composed of essential portions of crRNA and tracrRNA. Second, we transfected K562 cells (but not HeLa cells) with synthetic crRNA rather than plasmids encoding crRNA. HeLa cells were transfected with crRNA-encoding plasmids. Other groups used sgRNA-encoding plasmids. Third, our guide RNA had two additional guanine nucleotides at the 5' end, which are required for efficient transcription by T7 polymerase in vitro. No such additional nucleotides were included in the sgRNA used by others. Thus, the RNA sequence of our guide RNA can be shown as 5'-GGX₂₀, whereas 5'-GX₁₉, in which X₂₀ or GX₁₉ corresponds to the 20-bp target sequence, represents the sequence used by others. The first guanine nucleotide is required for transcription by RNA polymerase in cells. To test whether off-target RGEN effects can be attributed to these differences, we chose four RGENs that induced off-target mutations in human cells at high frequencies (13). First, we compared our method of using in vitro transcribed dualRNA with the method of transfecting sgRNA-encoding plasmids in K562 cells and measured mutation frequencies at the on-target and off-target sites via the T7E1 assay. Three RGENs showed comparable mutation frequencies at on-target and off-target sites regardless of the composition of guide RNA. Interestingly, one RGEN (VEFGA site 1) did not induce indels at one validated off-target site, which differs by three nucleotides from the on-target site (termed OT1-11, FIG. 14), when synthetic dualRNA was used. But the synthetic dualRNA did not discriminate the other validated off-target site (OT1-3), which differs by two nucleotides from the on-target site.

Next, we tested whether the addition of two guanine nucleotides at the 5' end of sgRNA could make RGENs more specific by comparing 5'-GGX₂₀ (or 5'-GGGX₁₉) sgRNA with 5'-GX₁₉ sgRNA. Four GX₁₉ sgRNAs complexed with Cas9 induced indels equally efficiently at on-target and off-target sites, tolerating up to four nucleotide mismatches. In sharp contrast, GGX₂₀ sgRNAs discriminated off-target sites effectively. In fact, the T7E1 assay barely detected RGEN-induced indels at six out of the seven validated off-target sites when we used the four GGX₂₀ sgRNAs (FIG. 15). We noticed, however, that two GGX₂₀ sgRNAs (VEGFA sites 1 and 3) were less active at on-target sites than were the corresponding GX₁₉ sgRNAs. These results show that the extra nucleotides at the 5' end can affect mutation frequencies at on-target and off-target sites, perhaps by altering guide RNA stability, concentration, or secondary structure.

These results suggest that three factors—the use of synthetic guide RNA rather than guide RNA-encoding plasmids, dualRNA rather than sgRNA, and GGX₂₀ sgRNA rather than GX₁₉ sgRNA—have cumulative effects on the discrimination of off-target sites.

Example 7: Paired Cas9 Nickases

In principle, single-strand breaks (SSBs) cannot be repaired by error-prone NHEJ but still trigger high fidelity homologous recombination (HDR) or base excision repair. But nickase-induced targeted mutagenesis via HDR is much less efficient than is nuclease-induced mutagenesis. We reasoned that paired Cas9 nickases would produce composite DSBs, which trigger DNA repair via NHEJ or HDR,

leading to efficient mutagenesis (FIG. 16A). Furthermore, paired nickases would double the specificity of Cas9-based genome editing.

We first tested several Cas9 nucleases and nickases designed to target sites in the AAVS1 locus (FIG. 16B) in vitro via fluorescent capillary electrophoresis. Unlike Cas9 nucleases that cleaved both strands of DNA substrates, Cas9 nickases composed of guide RNA and a mutant form of Cas9 in which a catalytic aspartate residue is changed to an alanine (D10A Cas9) cleaved only one strand, producing site-specific nicks (FIG. 16C, D). Interestingly, however, some nickases (AS1, AS2, AS3, and S6 in FIG. 17A) induced indels at target sites in human cells, suggesting that nicks can be converted to DSBs, albeit inefficiently, in vivo. Paired Cas9 nickases producing two adjacent nicks on opposite DNA strands yielded indels at frequencies that ranged from 14% to 91%, comparable to the effects of paired nucleases (FIG. 17A). The repair of two nicks that would produce 5' overhangs led to the formation of indels much more frequently than those producing 3' overhangs at three genomic loci (FIG. 17A and FIG. 18). In addition, paired nickases enabled targeted genome editing via homology-directed repair more efficiently than did single nickases (FIG. 19).

We next measured mutation frequencies of paired nickases and nucleases at off-target sites using deep sequencing. Cas9 nucleases complexed with three sgRNAs induced off-target mutations at six sites that differ by one or two nucleotides from their corresponding on-target sites with frequencies that ranged from 0.5% to 10% (FIG. 17B). In contrast, paired Cas9 nickases did not produce indels above the detection limit of 0.1% at any of the six off-target sites. The S2 Off-1 site that differs by a single nucleotide at the first position in the PAM (i.e., N in NGG) from its on-target site can be considered as another on-target site. As expected, the Cas9 nuclease complexed with the S2 sgRNA was equally efficient at this site and the on-target site. In sharp contrast, D10A Cas9 complexed with the S2 and AS2 sgRNAs discriminated this site from the on-target site by a factor of 270 fold. This paired nickase also discriminated the AS2 off-target sites (Off-1 and Off-9 in FIG. 17B) from the on-target site by factors of 160 fold and 990 fold, respectively.

Example 8: Chromosomal DNA Splicing Induced by Paired Cas9 Nickases

Two concurrent DSBs produced by engineered nucleases such as ZENs and TALENs can promote large deletions of the intervening chromosomal segments has been reported. We tested whether two SSBs induced by paired Cas9 nickases can also produce deletions in human cells. We used PCR to detect deletion events and found that seven paired nickases induced deletions of up to 1.1-kbp chromosomal segments as efficiently as paired Cas9 nucleases did (FIG. 20A, B). DNA sequences of the PCR products confirmed the deletion events (FIG. 20C). Interestingly, the sgRNA-matching sequence remained intact in two out of seven deletion-specific PCR amplicons (underlined in FIG. 20C). In contrast, Cas9 nuclease pairs did not produce sequences that contained intact target sites. This finding suggests that two distant nicks were not converted to two separate DSBs to promote deletions of the intervening chromosomal segment. In addition, it is unlikely that two nicks separated by more than a 100 bp can produce a composite DSB with large overhangs under physiological conditions because the melting temperature is very high.

We propose that two distant nicks are repaired by strand displacement in a head-to-head direction, resulting in the formation of a DSB in the middle, whose repair via NHEJ causes small deletions (FIG. 20D). Because the two target sites remain intact during this process, nickases can induce SSBs again, triggering the cycle repeatedly until the target sites are deleted. This mechanism explains why two offset nicks producing 5' overhangs but not those producing 3' overhangs induced indels efficiently at three loci.

We then investigated whether Cas9 nucleases and nickases can induce unwanted chromosomal translocations that result from NHEJ repair of on-target and off-target DNA cleavages (FIG. 21A). We were able to detect translocations induced by Cas9 nucleases using PCR (FIG. 21B, C). No such PCR products were amplified using genomic DNA isolated from cells transfected with the plasmids encoding the AS2+S3 Cas9 nickase pair. This result is in line with the fact that both AS2 and S3 nickases, unlike their corresponding nucleases, did not produce indels at off-target sites (FIG. 17B).

These results suggest that paired Cas9 nickases allow targeted mutagenesis and large deletions of up to 1-kbp chromosomal segments in human cells. Importantly, paired nickases did not induce indels at off-target sites at which their corresponding nucleases induce mutations. Furthermore, unlike nucleases, paired nickases did not promote unwanted translocations associated with off-target DNA cleavages. In principle, paired nickases double the specificity of Cas9-mediated mutagenesis and will broaden the utility of RNA-guided enzymes in applications that require precise genome editing such as gene and cell therapy. One caveat to this approach is that two highly active sgRNAs are needed to make an efficient nickase pair, limiting targetable sites. As shown in this and other studies, not all sgRNAs are equally active. When single clones rather than populations of cells are used for further studies or applications, the choice of guide RNAs that represent unique sequences in the genome and the use of optimized guide RNAs would suffice to avoid off-target mutations associated with Cas9 nucleases. We propose that both Cas9 nucleases and paired nickases are powerful options that will facilitate precision genome editing in cells and organisms.

Example 9: Genotyping with CRISPR/Cas-Derived RNA-Guided Endonucleases

Next, we reasoned that RGENs can be used in Restriction fragment length polymorphism (RFLP) analysis, replacing conventional restriction enzymes. Engineered nucleases including RGENs induce indels at target sites, when the DSBs caused by the nucleases are repaired by the error-prone non-homologous end-joining (NHEJ) system. RGENs that are designed to recognize the target sequences cannot cleave mutant sequences with indels but will cleave wild-type target sequences efficiently.

9-1. RGEN Components

crRNA and tracrRNA were prepared by in vitro transcription using MEGAscript T7 kit (Ambion) according to the manufacturer's instruction. Transcribed RNAs were resolved on a 8% denaturing urea-PAGE gel. The gel slice containing RNA was cut out and transferred to elution buffer. RNA was recovered in nuclease-free water followed by phenol:chloroform extraction, chloroform extraction, and ethanol precipitation. Purified RNA was quantified by spectrometry. Templates for crRNA were prepared by annealing an oligonucleotide whose sequence is shown as 5'-GAAATTAATACGACTCACTATAGGX₂₀GTTT-

TAGAGCTATGCTGTTT-3' (SEQ ID NO: 76), in which X₂₀ is the target sequence, and its complementary oligonucleotide. The template for tracrRNA was synthesized by extension of forward and reverse oligonucleotides (5'-GAAATTAATACGACTCACTATAGGAACCAT-CAAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCG-3' (SEQ ID NO: 77) and 5'-AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAACT TGC-TATG-3' (SEQ ID NO: 78)) using Phusion polymerase (New England Biolabs).

9-2. Recombinant Cas9 Protein Purification

The Cas9 DNA construct used in our previous Example, which encodes Cas9 fused to the His6-tag at the C terminus, was inserted in the pET-28a expression vector. The recombinant Cas9 protein was expressed in *E. coli* strain BL21 (DE3) cultured in LB medium at 25° C. for 4 hours after induction with 1 mM IPTG. Cells were harvested and resuspended in buffer containing 20 mM Tris PH 8.0, 500 mM NaCl, 5 mM imidazole, and 1 mM PMSF. Cells were frozen in liquid nitrogen, thawed at 4° C., and sonicated. After centrifugation, the Cas9 protein in the lysate was bound to Ni-NTA agarose resin (Qiagen), washed with buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, and 20 mM imidazole, and eluted with buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, and 250 mM imidazole. Purified Cas9 protein was dialyzed against 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT, and 10% glycerol and analyzed by SDS-PAGE.

9-3. T7 Endonuclease I Assay

The T7E1 assay was performed as following. In brief, PCR products amplified using genomic DNA were denatured at 95° C., reannealed at 16° C., and incubated with 5 units of T7 Endonuclease I (New England BioLabs) for 20 min at 37° C. The reaction products were resolved using 2 to 2.5% agarose gel electrophoresis.

9-4. RGEN-RFLP Assay

PCR products (100-150 ng) were incubated for 60 min at 37° C. with optimized concentrations (Table 10) of Cas9 protein, tracrRNA, crRNA in 10 µl NEB buffer 3 (1x). After the cleavage reaction, RNase A (4 µg) was added, and the reaction mixture was incubated for 30 min at 37° C. to remove RNA. Reactions were stopped with 6x stop solution buffer containing 30% glycerol, 1.2% SDS, and 100 mM EDTA. Products were resolved with 1-2.5% agarose gel electrophoresis and visualized with EtBr staining.

TABLE 10

Concentration of RGEN components in RFLP assays			
Target Name	Cas9 (ng/µl)	crRNA (ng/µl)	tracrRNA (ng/µl)
C4BPB	100	25	60
PIBF-NGG-RGEN	100	25	60
HLA-B	1.2	0.3	0.7
CCR5-ZFN	100	25	60
CTNNB1 Wild type specific	30	10	20
CTNNB1 mutant specific	30	10	20
CCR5 WT-specific	100	25	60
CCR5 Δ32-specific	10	2.5	6
KRAS WT specific (wt)	30	10	20
KRAS mutant specific (m8)	30	10	20
KRAS WT specific (m6)	30	10	20
KRAS mutant specific (m6, 8)	30	10	20
PIK3CA WT specific (wt)	100	25	60

TABLE 10-continued

Concentration of RGEN components in RFLP assays			
Target Name	Cas9 (ng/µl)	crRNA (ng/µl)	tracrRNA (ng/µl)
PIK3CA mutant specific (m4)	30	10	20
PIK3CA WT specific (m7)	100	25	60
PIK3CA mutant specific (m4, 7)	30	10	20
BRAF WT-specific	30	10	20
BRAF mutant-specific	100	25	60
NRAS WT-specific	100	25	60
NRAS mutant-specific	30	10	20
IDH WT-specific	30	10	20
IDH mutant-specific	30	10	20
PIBF-NAG-RGEN	30	10	60

TABLE 11

Primers				
Gene (site)	Direction	Sequence (5' to 3')	SEQ ID NO	
CCR5 (RGEN)	F1	CTCCATGGTGCTATAGAGCA	79	
	F2	GAGCCAAGCTCTCCATCTAGT	80	
	R	GCCCTGTCAAGAGTTGACAC	81	
CCR5 (ZFN)	F	GCACAGGGTGAACAAGATGGA	82	
	R	GCCAGGTACCTATCGATTGTCAGG	83	
CCR5 (del132)	F	GAGCCAAGCTCTCCATCTAGT	84	
	R	ACTCTGACTG GGTCCACCAGC	85	
C4BPB	F1	TATTTGGCTGGTTGAAAGGG	86	
	R1	AAAGTCATGAAATAAACACACCCCA	87	
	F2	CTGCATTGATATGGTAGTACCATG	88	
C4BPB	R2	GCTGTTCAATGCAATGGAAATG	89	
	CTNNB1	F	ATGGAGTTGGACATGGCCATGG	90
R		ACTCACTATCCACAGTTCAGCATTTACC	91	
KRAS	F	TGGAGATAGCTGTGACGAACTTT	92	
	R	CAACAA AGCAAAGGTAAAGTTGGTAATAG	93	
PIK3CA	F	GGTTTCAGGAGATGTGTTACAAGGC	94	
	R	GATTGTGCAATTCCTATGCAATCGGTC	95	
NRAS	F	CACTGGGTACTTAATCTGTAGCCCTC	96	
	R	GTTTCCAAGTCATTCACAGTAGC	97	
IDH1	F	CATCACTGCAGTTGTAGGTTATAACTATCC	98	
	R	TTGAAAACCACAGATCTGGTTGAACC	99	
BRAF	F	GGAGTGCCCAAGAGAATATCTGG	100	
	R	CTGAAACTGGTTTCAAATATTCGTTTAAAGG	101	
PIBF	F	GCTCTGTATGCCCTGTAGTAGG	102	
	R	TTTGCACTGCACCTTACCTTTG	103	

9-5. Plasmid Cleavage Assay

Restriction enzyme-treated linearized plasmid (100 ng) was incubated for 60 min at 37° C. with Cas9 protein (0.1 µg), tracrRNA (60 ng), and crRNA (25 ng) in 10 µl NEB 3 buffer (1x). Reactions were stopped with 6x stop solution containing 30% glycerol, 1.2% SDS, and 100 mM EDTA. Products were resolved with 1% agarose gel electrophoresis and visualized with EtBr staining.

9-6. Strategy of RFLP

New RGENs with desired DNA specificities can be readily created by replacing crRNA; no de novo purification of custom proteins is required once recombinant Cas9 protein is available. Engineered nucleases, including

RGENs, induce small insertions or deletions (indels) at target sites when the DSBs caused by the nucleases are repaired by error-prone non-homologous end-joining (NHEJ). RGENs that are designed to recognize the target sequences cleave wild-type sequences efficiently but cannot cleave mutant sequences with indels (FIG. 22).

We first tested whether RGENs can differentially cleave plasmids that contain wild-type or modified C4BPB target sequences that harbor 1- to 3-base indels at the cleavage site. None of the six plasmids with these indels were cleaved by a C4BPB-specific RGEN5 composed of target-specific crRNA, tracrRNA, and recombinant Cas9 protein (FIG. 23). In contrast, the plasmid with the intact target sequence was cleaved efficiently by this RGEN.

9-7. Detection of Mutations Induced by the Same RGENs Using RGEN-Mediated RFLP

Next, to test the feasibility of RGEN-mediated RFLP for detection of mutations induced by the same RGENs, we utilized gene-modified K562 human cancer cell clones established using an RGEN targeting C4BPB gene (Table 12).

TABLE 12

Target sequence of RGENs used in this study		
Gene	Target sequence	SEQ ID NO
human C4BPB	<u>AATGACCACTACATCCTCAAGGG</u>	104
mouse Pibf1	<u>AGATGATGTCTCATCATCAGAGG</u>	105

C4BPB mutant clones used in this study have various mutations ranging from 94 bp deletion to 67 bp insertion (FIG. 24A). Importantly, all mutations occurred in mutant clones resulted in the loss of RGEN target site. Among 6 C4BPB clones analyzed, 4 clones have both wildtype and mutant alleles (+/-) and 2 clones have only mutant alleles (-/-).

The PCR products spanning the RGEN target site amplified from wildtype K562 genomic DNA were digested completely by the RGEN composed of target-specific crRNA, tracrRNA, and recombinant Cas9 protein expressed in and purified from *E. coli* (FIG. 24B/Lane 1). When the C4BPB mutant clones were subjected to RFLP analysis using the RGEN, PCR amplicons of +/- clones that contained both wildtype and mutant alleles were partially digested, and those of -/- clones that did not contain the wildtype allele were not digested at all, yielding no cleavage products corresponding to the wildtype sequence (FIG. 24B). Even a single-base insertion at the target site blocked the digestion (#12 and #28 clones) of amplified mutant alleles by the C4BPB RGEN, showing the high specificity of RGEN-mediated RFLP. We subjected the PCR amplicons to the mismatch-sensitive T7E1 assay in parallel (FIG. 24B). Notably, the T7E1 assay was not able to distinguish -/- clones from +/- clones. To make it matters worse, the T7E1 assay cannot distinguish homozygous mutant clones that contain the same mutant sequence from wildtype clones, because annealing of the same mutant sequence will form a homoduplex. Thus, RGEN-mediated RFLP has a critical advantage over the conventional mismatch-sensitive nuclease assay in the analysis of mutant clones induced by engineered nucleases including ZENs, TALENs and RGENs.

9-8. Quantitative Assay for RGEN-RFLP Analysis

We also investigated whether RGEN-RFLP analysis is a quantitative method. Genomic DNA samples isolated from the C4BPB null clone and the wild-type cells were mixed at various ratios and used for PCR amplifications. The PCR products were subjected to RGEN genotyping and the T7E1 assay in parallel (FIG. 25b). As expected, DNA cleavage by the RGEN was proportional to the wild type to mutant ratio. In contrast, results of the T7E1 assay correlated poorly with mutation frequencies inferred from the ratios and were inaccurate, especially at high mutant %, a situation in which complementary mutant sequences can hybridize with each other to form homoduplexes.

9-9. Analysis of Mutant Mouse Founders Using a RGEN-Mediated RFLP Genotyping

We also applied RGEN-mediated RFLP genotyping (RGEN genotyping in short) to the analysis of mutant mouse founders that had been established by injection of TALENs into mouse one-cell embryos (FIG. 26A). We designed and used an RGEN that recognized the TALEN target site in the Pibf1 gene (Table 10). Genomic DNA was isolated from a wildtype mouse and mutant mice and subjected to RGEN genotyping after PCR amplification. RGEN genotyping successfully detected various mutations, which ranged from one to 27-bp deletions (FIG. 26B). Unlike the T7E1 assay, RGEN genotyping enabled differential detection of +/- and -/- founder.

9-10. Detection of Mutations Induced in Human Cells by a CCR5-Specific ZEN Using RGENs

In addition, we used RGENs to detect mutations induced in human cells by a CCR5-specific ZFN, representing yet another class of engineered nucleases (FIG. 27). These results show that RGENs can detect mutations induced by nucleases other than RGENs themselves. In fact, we expect that RGENs can be designed to detect mutations induced by most, if not all, engineered nucleases. The only limitation in the design of an RGEN genotyping assay is the requirement for the GG or AG (CC or CT on the complementary strand) dinucleotide in the PAM sequence recognized by the Cas9 protein, which occurs once per 4 bp on average. Indels induced anywhere within the seed region of several bases in crRNA and the PAM nucleotides are expected to disrupt RGEN-catalyzed DNA cleavage. Indeed, we identified at least one RGEN site in most (98%) of the ZEN and TALEN sites.

9-11. Detection of Polymorphisms or Variations Using RGEN

Next, we designed and tested a new RGEN that targets a highly polymorphic locus, HLA-B, that encodes Human Leukocyte Antigen B (a.k.a. MHC class I protein) (FIG. 28). HeLa cells were transfected with RGEN plasmids, and the genomic DNA was subjected to T7E1 and RGEN-RFLP analyses in parallel. T7E1 produced false positive bands that resulted from sequence polymorphisms near the target site (FIG. 25c). As expected, however, the same RGEN used for gene disruption cleaved PCR products from wild-type cells completely but those from RGEN-transfected cells partially, indicating the presence of RGEN-induced indels at the target site. This result shows that RGEN-RFLP analysis has a clear advantage over the T7E1 assay, especially when it is not known whether target genes have polymorphisms or variations in cells of interest.

9-12. Detection of Recurrent Mutations Found in Cancer and Naturally-Occurring Polymorphisms Through RGEN-RFLP Analysis

RGEN-RFLP analysis has applications beyond genotyping of engineered nuclease-induced mutations. We sought to

use RGEN genotyping to detect recurrent mutations found in cancer and naturally-occurring polymorphisms. We chose the human colorectal cancer cell line, HCT116, which carries a gain-of-function 3-bp deletion in the oncogenic CTNNB1 gene encoding beta-catenin. PCR products amplified from HCT116 genomic DNA were cleaved partially by both wild-type-specific and mutant-specific RGENs, in line with the heterozygous genotype in HCT116 cells (FIG. 29a). In sharp contrast, PCR products amplified from DNA from HeLa cells harboring only wild-type alleles were digested completely by the wild-type-specific RGEN and were not cleaved at all by the mutation-specific RGEN.

We also noted that HEK293 cells harbor the 32-bp deletion (del32) in the CCR5 gene, which encodes an essential co-receptor of HIV infection: Homozygous del32 CCR5 carriers are immune to HIV infection. We designed one RGEN specific to the del32 allele and the other to the wild-type allele. As expected, the wild-type-specific RGEN cleaved the PCR products obtained from K562, SKBR3, or HeLa cells (used as wild-type controls) completely but those from HEK293 cells partially (FIG. 30a), confirming the presence of the uncleavable del32 allele in HEK293 cells. Unexpectedly, however, the del32-specific RGEN cleaved the PCR products from wild-type cells as efficiently as those from HEK293 cells. Interestingly, this RGEN had an off-target site with a single-base mismatch immediately downstream of the on-target site (FIG. 30). These results suggest that RGENs can be used to detect naturally-occurring indels but cannot distinguish sequences with single nucleotide polymorphisms or point mutations due to their off-target effects.

To genotype oncogenic single-nucleotide variations using RGENs, we attenuated RGEN activity by employing a single-base mismatched guide RNA instead of a perfectly-matched RNA. RGENs that contained the perfectly-matched guide RNA specific to the wild-type sequence or mutant sequence cleaved both sequences (FIGS. 31a and 32a). In contrast, RGENs that contained a single-base mismatched guide RNA distinguished the two sequences, enabling genotyping of three recurrent oncogenic point mutations in the KRAS, PIK3CA, and IDH1 genes in human cancer cell lines (FIG. 29b and FIGS. 33a, b). In addition, we were able to detect point mutations in the BRAF and NRAS genes using RGENs that recognize the NAG PAM sequence (FIGS. 33c, d). We believe that we can use RGEN-RFLP to genotype almost any, if not all, mutations or polymorphisms in the human and other genomes.

The above data proposes RGENs as providing a platform to use simple and robust RFLP analysis for various sequence

variations. With high flexibility in reprogramming target sequence, RGENs can be used to detect various genetic variations (single nucleotide variations, small insertion/deletions, structural variations) such as disease-related recurring mutations, genotypes related to drug-response by a patient and also mutations induced by engineered nucleases in cells. Here, we used RGEN genotyping to detect mutations induced by engineered nucleases in cells and animals. In principle, one could also use RGENs that will specifically detect and cleave naturally-occurring variations and mutations.

Based on the above description, it should be understood by those skilled in the art that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention without departing from the technical idea or essential features of the invention as defined in the following claims. In this regard, the above-described examples are for illustrative purposes only, and the invention is not intended to be limited by these examples. The scope of the present invention should be understood to include all of the modifications or modified form derived from the meaning and scope of the following claims or its equivalent concepts.

REFERENCES

1. M. Jinek et al., *Science* 337, 816 (Aug. 17, 2012).
2. H. Kim, E. Um, S. R. Cho, C. Jung, J. S. Kim, *Nat Methods* 8, 941 (November 2011).
3. H. J. Kim, H. J. Lee, H. Kim, S. W. Cho, J. S. Kim, *Genome Res* 19, 1279 (July 2009).
4. E. E. Perez et al., *Nat Biotechnol* 26, 808 (July 2008).
5. J. C. Miller et al., *Nat Biotechnol* 29, 143 (February 2011).
6. C. Mussolino et al., *Nucleic Acids Res* 39, 9283 (November 2011).
7. J. Cohen, *Science* 332, 784 (May 13, 2011).
8. V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, *Nat Methods* 8, 765 (September 2011).
9. R. Gabriel et al., *Nat Biotechnol* 29, 816 (September 2011).
10. E. Kim et al., *Genome Res*, (Apr. 20, 2012).
11. H. J. Lee, J. Kweon, E. Kim, S. Kim, J. S. Kim, *Genome Res* 22, 539 (March 2012).
12. H. J. Lee, E. Kim, J. S. Kim, *Genome Res* 20, 81 (January 2010).
13. Fu Y, Foden J A, Khayter C, Maeder M L, Reyon D, Joung J K, Sander J D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotech advance online publication* (2013)

SEQUENCE LISTING

```

Sequence total quantity: 332
SEQ ID NO: 1          moltype = DNA length = 4107
FEATURE              Location/Qualifiers
misc_feature         1..4107
                     note = Description of Artificial Sequence: Synthetic
                     polynucleotide
misc_feature         1..4107
                     note = Cas9-coding sequence
source               1..4107
                     mol_type = other DNA
                     organism = synthetic construct

SEQUENCE: 1
atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gccagcaag aagttcaagg tgctgggcaa caccgaccgc 120
cacagcatca agaagaacct gatcggcgcc ctgctgtctc acagcggcga gaccgccgag 180

```

-continued

```

gccacccgcc tgaagcgcac cgcccgcgc cgctacacc gccgcaagaa ccgatctgc 240
tacctgcagg agatcttcag caacgagatg gccaaagtg acgacagctt cttccaccgc 300
ctggaggaga gcttcctggg ggaggaggac aagaagcac agcggcacc catctctgcg 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgcgcaag 420
aagctggggg acagcacoga caaggcggac ctgctcctga tctacctggc cctggcccac 480
atgatcaagt tccgcgcca cttcctgatc gagggcgacc tgaacccoga caacagcgac 540
gtggacaagc tgttcatcca gctgggtgcag acctacaacc agctgttcca ggagaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc 660
cgctcggaga acctgatcgc ccagctgccc ggcgagaaga agaaccggct gttcggcaac 720
ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcca cctggccgag 780
gacgccaagc tgacgtgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc 900
ctgctgagcg acctcctggc gctgaacacc gagatcacca agggccccct gagcgccagc 960
atgatcaagc gatcagcoga gcaccaccag gacctgaccc tgetgaaagg cctgggtgcg 1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacatcg acggcggcgc gagccaggag gagtctaca agtctatcaa gcccactctg 1140
gagaagatgg acggcacoga ggagctgctg gtgaagctga accgcgagga cctgctgcgc 1200
aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgcac 1260
gccatcctgc gcccaccaga ggactctac cccttctga aggacaaccg cgagaagatc 1320
gagaagatcc tgaccttcgc catcccctac tacgtgggccc ccttgcccgc cggcaacagc 1380
cgcttcgocct ggatgaacccg caagagcgag gagaccatca cccctggaa cttcggaggag 1440
gtggtggaca agggcgccag gcgccagagc ttcctogagc gcatgaccaa cttcgacaag 1500
aacctgcccc acgagaaggt gctgcccacg cacagcctgc tgtacgagta cttcacccgtg 1560
tacaacgagc tgaccaaggt gaatgacgtg accgagggca tgcgcaagcc cgcttctctg 1620
agcggcgagc agaagaaggg catcgtggac ctgctgttca agaccaaccg caaggtgacc 1680
gtgaagcagc tgaaggaagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740
agcggcgtgg aggaccgctt caacgcccagc ctgggcaact accacgacct gctgaagatc 1800
atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggactctgtg 1860
ctgacccctga ccctgttcca ggaccgagc atgatcgagg agcgcctgaa gacctacgcc 1920
cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggctggggc 1980
cgcttgagcc gcaagcttat caacggcatc cgcgacaagc agagcgcaaa gacctcctg 2040
gacttctcga agaagcagcg cttcgccaac cgcaacttca tgacgtgat ccacgacgac 2100
agcctgacct tcaaggagga catccagaag gcccaaggtga gcggccaggg cgacagcctg 2160
cacgagcaca tcgccaacct ggccggcagc cccgcatca agaagggcat cctgcagacc 2220
gtgaaggtgg tggacgagct ggtgaaggtg atgggcccgc acaagcccga gaacatcgtg 2280
atcgagatgg cccgagagaa ccagaccacc cagaagggcc agaagaacag ccgagagcgc 2340
atgaagcgca tcaaggaggg catcaaggag ctgggcaagc agatcctgaa ggagcaaccc 2400
gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460
gacatgtacg tggaccagga gctggacatc aaccgctcga gcgactacga cgtggaccac 2520
atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gaccgcagc 2580
gacaagaacc gcggcaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaa 2640
aactactggc gccagctgct gaacgccaag ctgatcacc agcgcaagtt cgacaacctg 2700
accaagggcg accgcccggg cctgagcgag ctggacaagg ccgcttcat caagcgccag 2760
ctggtggaga cccgcccagc caccaagcac gtggcccaga tccctggacag ccgcatgaa 2820
accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc 2880
aagctggtga gcaacttccg caaggacttc cagttctaca agtgctcoga gatcaacaac 2940
taccaccaag cccacgagc ctacctgaac gccgtggtg gcaccgccc gatcaagaag 3000
taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag 3060
atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtaact cttctacagc 3120
aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc 3180
cccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgagcattc 3240
gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tctgtaagaa gaccgaggtg 3300
cagaccggcg gcttcagcaa ggagagcatc ctgcccgaag gcaacagcga caagctgatc 3360
gcccgaaga aggactggga ccccagaag tacggcggct tcgacagccc caccgtggcc 3420
tacagcgtgc tgggtggggc caaggtggag aagggcaaga gcaagaagct gaagagcgtg 3480
aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540
ttcctggagg ccaagggtta caaggaggtg aagaaggacc tgatcatcaa gctgcccagg 3600
tacagcctgt tcgagctgga gaacggcccg aagcgcctgc tggccagcgc cggcgagctg 3660
cagaagggca acgagctggc cctgcccagc aagtacgtga acttctgta cctggccagc 3720
cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840
atcctggccg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag 3900
cccatcggcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctggggccc 3960
cccggcctc tcaagtaact cgacaccacc atcgaccgca agcgtacac cagcaccag 4020
gaggtgctgg acgcccacct gatccaccag agcatcaccg gtctgtacga gaccgcatc 4080
gacctgagcc agctggggcg cgactaa 4107

```

```

SEQ ID NO: 2 moltype = AA length = 21
FEATURE Location/Qualifiers
REGION 1..21
note = Description of Artificial Sequence: Synthetic peptide
REGION 1..21
note = Peptide tag
source 1..21
mol_type = protein
organism = synthetic construct

```

```

SEQUENCE: 2
GGSGPPKKKR KVYPYDVPDY A

```


-continued

FEATURE	Location/Qualifiers	
misc_feature	1..24	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..24	
	note = R1 primer for C4BPB	
source	1..24	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 9		
aaagtcatga aataaacaca ccca		24
SEQ ID NO: 10	moltype = DNA length = 24	
FEATURE	Location/Qualifiers	
misc_feature	1..24	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..24	
	note = F2 primer for C4BPB	
source	1..24	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 10		
ctgcattgat atggtagtag catg		24
SEQ ID NO: 11	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
misc_feature	1..21	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..21	
	note = R2 primer for C4BPB	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 11		
gctgttcatt gcaatggaat g		21
SEQ ID NO: 12	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..20	
	note = F1 primer for ADCY5	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 12		
gctccacact tagtgctctg		20
SEQ ID NO: 13	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	
misc_feature	1..20	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..20	
	note = R1 primer for ADCY5	
source	1..20	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 13		
ggtggcagga acctgtatgt		20
SEQ ID NO: 14	moltype = DNA length = 21	
FEATURE	Location/Qualifiers	
misc_feature	1..21	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..21	
	note = F2 primer for ADCY5	
source	1..21	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 14		
gtcattggcc agagatgtgg a		21
SEQ ID NO: 15	moltype = DNA length = 20	
FEATURE	Location/Qualifiers	

-continued

misc_feature	1..20	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..20	
		note = R2 primer for ADCY5
source	1..20	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 15		
gtcccatgac aggcgtgtat		20
SEQ ID NO: 16		moltype = DNA length = 20
FEATURE		Location/Qualifiers
misc_feature	1..20	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..20	
		note = F primer for KCNJ6
source	1..20	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 16		
gctggccea gtttcagtta		20
SEQ ID NO: 17		moltype = DNA length = 20
FEATURE		Location/Qualifiers
misc_feature	1..20	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..20	
		note = R1 primer for KCNJ6
source	1..20	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 17		
tggagccatt ggtttgcac		20
SEQ ID NO: 18		moltype = DNA length = 22
FEATURE		Location/Qualifiers
misc_feature	1..22	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..22	
		note = R2 primer for KCNJ6
source	1..22	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 18		
ccagaactaa gccgtttctg ac		22
SEQ ID NO: 19		moltype = DNA length = 20
FEATURE		Location/Qualifiers
misc_feature	1..20	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..20	
		note = F1 primer for CNTNAP2
source	1..20	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 19		
atcaccgaca accagtttcc		20
SEQ ID NO: 20		moltype = DNA length = 20
FEATURE		Location/Qualifiers
misc_feature	1..20	
		note = Description of Artificial Sequence: Synthetic oligonucleotide
misc_feature	1..20	
		note = F2 primer for CNTNAP2
source	1..20	
		mol_type = other DNA
		organism = synthetic construct
SEQUENCE: 20		
tgcagtgcag actctttcca		20
SEQ ID NO: 21		moltype = DNA length = 20
FEATURE		Location/Qualifiers
misc_feature	1..20	

-continued

	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..20	
source	note = R primer for CNTNAP2 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 21		
aaggacacag ggcaactgaa		20
SEQ ID NO: 22	moltype = DNA length = 20 Location/Qualifiers	
FEATURE	1..20	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..20	
source	note = F1 primer for N/A Chr. 5 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 22		
tgtggaacga gtggtgacag		20
SEQ ID NO: 23	moltype = DNA length = 22 Location/Qualifiers	
FEATURE	1..22	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..22	
source	note = R1 primer for N/A Chr. 5 1..22 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 23		
gctggattag gaggcaggat tc		22
SEQ ID NO: 24	moltype = DNA length = 22 Location/Qualifiers	
FEATURE	1..22	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..22	
source	note = F2 primer for N/A Chr. 5 1..22 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 24		
gtgctgagaa cgcttcatag ag		22
SEQ ID NO: 25	moltype = DNA length = 23 Location/Qualifiers	
FEATURE	1..23	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..23	
source	note = R2 primer for N/A Chr. 5 1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 25		
ggaccaaacc acattcttct cac		23
SEQ ID NO: 26	moltype = DNA length = 20 Location/Qualifiers	
FEATURE	1..20	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
misc_feature	1..20	
source	note = F primer for deletion 1..20 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 26		
ccacatctcg ttctcggtt		20
SEQ ID NO: 27	moltype = DNA length = 20 Location/Qualifiers	
FEATURE	1..20	
misc_feature	note = Description of Artificial Sequence: Synthetic	

-continued

```

misc_feature      oligonucleotide
                  1..20
                  note = R primer for deletion
source           1..20
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 27
tcacaagccc acagatattt                               20

SEQ ID NO: 28      moltype = RNA length = 105
FEATURE           Location/Qualifiers
misc_feature      1..105
                  note = Description of Artificial Sequence: Synthetic
                  polynucleotide
misc_feature      1..105
                  note = sgRNA for CCR5
source           1..105
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 28
ggtgacatca attattatac atgttttaga gctagaaata gcaagttaaa ataaggctag 60
tccgttatca acttgaaaaa gtggcaccga gtcggtgctt ttttt                    105

SEQ ID NO: 29      moltype = RNA length = 44
FEATURE           Location/Qualifiers
misc_feature      1..44
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..44
                  note = crRNA for CCR5
source           1..44
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 29
ggtgacatca attattatac atgttttaga gctatgctgt tttg                               44

SEQ ID NO: 30      moltype = RNA length = 86
FEATURE           Location/Qualifiers
misc_feature      1..86
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..86
                  note = tracrRNA for CCR5
source           1..86
                  mol_type = other RNA
                  organism = synthetic construct

SEQUENCE: 30
ggaaccattc aaaacagcat agcaagttaa aataaggcta gtccgttatc aacttgaaaa 60
agtgcaccg agtcggtgct tttttt                    86

SEQ ID NO: 31      moltype = DNA length = 86
FEATURE           Location/Qualifiers
misc_feature      1..86
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..86
                  note = Foxn1 #1 sgRNA
source           1..86
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 31
gaaattaata cgactcacta taggcagtct gacgtcacac ttccgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg                    86

SEQ ID NO: 32      moltype = DNA length = 86
FEATURE           Location/Qualifiers
misc_feature      1..86
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..86
                  note = Foxn1 #2 sgRNA
source           1..86
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 32
gaaattaata cgactcacta taggacttcc aggtccacc cgacgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg                    86

SEQ ID NO: 33      moltype = DNA length = 86

```

-continued

FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Foxn1 #3 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 33
gaaattaata cgactcacta taggccaggc tccacccgac tggagtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 34 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Foxn1 #4 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 34
gaaattaata cgactcacta taggactgga gggcgaaccc caaggtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 35 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Foxn1 #5 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 35
gaaattaata cgactcacta taggacccca aggggacctc atgcgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 36 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Prkdc #1 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 36
gaaattaata cgactcacta taggttagtt tttccagag acttgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 37 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Prkdc #2 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 37
gaaattaata cgactcacta taggttggtt tgcttggtt tatcgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 38 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
note = Description of Artificial Sequence: Synthetic
oligonucleotide

misc_feature 1..86
note = Prkdc #3 sgRNA

source 1..86
mol_type = other DNA
organism = synthetic construct

-continued

SEQUENCE: 38
gaaattaata cgactcacta taggcacaag caaaccaaag tctcgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 39 moltype = DNA length = 86
FEATURE Location/Qualifiers
misc_feature 1..86
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..86
 note = Prkdc #4 sgRNA
source 1..86
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 39
gaaattaata cgactcacta taggcctcaa tgctaagcga cttcgtttta gagctagaaa 60
tagcaagtta aaataaggct agtccg 86

SEQ ID NO: 40 moltype = DNA length = 29
FEATURE Location/Qualifiers
misc_feature 1..29
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..29
 note = F1 primer for Foxn1
source 1..29
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 40
gtctgtctat catctcttcc cttctctcc 29

SEQ ID NO: 41 moltype = DNA length = 25
FEATURE Location/Qualifiers
misc_feature 1..25
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..25
 note = F2 primer for Foxn1
source 1..25
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 41
tccctaatacc gatggctagc tccag 25

SEQ ID NO: 42 moltype = DNA length = 23
FEATURE Location/Qualifiers
misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..23
 note = R1 primer for Foxn1
source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 42
acgagcagct gaagttagca tgc 23

SEQ ID NO: 43 moltype = DNA length = 32
FEATURE Location/Qualifiers
misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..32
 note = R2 primer for Foxn1
source 1..32
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 43
ctactcaatg ctcttagagc taccaggctt gc 32

SEQ ID NO: 44 moltype = DNA length = 20
FEATURE Location/Qualifiers
misc_feature 1..20
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..20
 note = F primer for Prkdc
source 1..20
 mol_type = other DNA

-continued

organism = synthetic construct
 SEQUENCE: 44
 gactgttggtg gggaggccg 20

SEQ ID NO: 45 moltype = DNA length = 24
 FEATURE Location/Qualifiers
 misc_feature 1..24
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..24
 note = F2 primer for Prkdc
 source 1..24
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 45
 gggaggccg aaagtcttat ttg 24

SEQ ID NO: 46 moltype = DNA length = 28
 FEATURE Location/Qualifiers
 misc_feature 1..28
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..28
 note = R1 primer for Prkdc
 source 1..28
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 46
 cctgaagact gaagttggca gaagtggag 28

SEQ ID NO: 47 moltype = DNA length = 27
 FEATURE Location/Qualifiers
 misc_feature 1..27
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..27
 note = R2 primer for Prkdc
 source 1..27
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 47
 ctttaggct tcttctctac aatcacg 27

SEQ ID NO: 48 moltype = DNA length = 38
 FEATURE Location/Qualifiers
 misc_feature 1..38
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..38
 note = F primer for Foxn1
 source 1..38
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 48
 ctcggtgtgt agccctgacc tcggtgtgta gccctgac 38

SEQ ID NO: 49 moltype = DNA length = 21
 FEATURE Location/Qualifiers
 misc_feature 1..21
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..21
 note = R primer for Foxn1
 source 1..21
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 49
 agactggcct ggaactcaca g 21

SEQ ID NO: 50 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..23
 note = F primer for Foxn1
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

-continued

SEQUENCE: 50
cactaaagcc tgtcaggaag ccg 23

SEQ ID NO: 51 moltype = DNA length = 21
FEATURE Location/Qualifiers
misc_feature 1..21
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..21
note = R primer for Foxn1
source 1..21
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 51
ctgtggagag cacacagcag c 21

SEQ ID NO: 52 moltype = DNA length = 19
FEATURE Location/Qualifiers
misc_feature 1..19
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..19
note = F primer for Foxn1
source 1..19
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 52
gctgcgacct gagaccatg 19

SEQ ID NO: 53 moltype = DNA length = 26
FEATURE Location/Qualifiers
misc_feature 1..26
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..26
note = R primer for Foxn1
source 1..26
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 53
cttcaatggc ttctgetta ggctac 26

SEQ ID NO: 54 moltype = DNA length = 23
FEATURE Location/Qualifiers
misc_feature 1..23
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..23
note = F primer for Foxn1
source 1..23
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 54
ggttcagatg aggccatcct ttc 23

SEQ ID NO: 55 moltype = DNA length = 24
FEATURE Location/Qualifiers
misc_feature 1..24
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..24
note = R primer for Foxn1
source 1..24
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 55
cctgatctgc aggcttaacc cttg 24

SEQ ID NO: 56 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_feature 1..22
note = Description of Artificial Sequence: Synthetic
oligonucleotide
misc_feature 1..22
note = F primer for Prkdc
source 1..22
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 56

-continued

```

ctcacctgca catcacatgt gg                                22

SEQ ID NO: 57      moltype = DNA length = 20
FEATURE           Location/Qualifiers
misc_feature      1..20
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..20
                  note = R primer for Prkdc
source           1..20
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 57
ggcatccacc ctatggggtc                                20

SEQ ID NO: 58      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..25
                  note = F primer for Prkdc
source           1..25
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 58
gccttgacct agagcttaaa gagcc                          25

SEQ ID NO: 59      moltype = DNA length = 25
FEATURE           Location/Qualifiers
misc_feature      1..25
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..25
                  note = R primer for Prkdc
source           1..25
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 59
ggtcttgta gcaggaagga cactg                            25

SEQ ID NO: 60      moltype = DNA length = 27
FEATURE           Location/Qualifiers
misc_feature      1..27
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..27
                  note = F primer for Prkdc
source           1..27
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 60
aaaactctgc ttgatgggat atgtggg                        27

SEQ ID NO: 61      moltype = DNA length = 26
FEATURE           Location/Qualifiers
misc_feature      1..26
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..26
                  note = R primer for Prkdc
source           1..26
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 61
ctctcactgg ttatctgtgc tccttc                          26

SEQ ID NO: 62      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..23
                  note = F primer for Prkdc
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 62
ggatcaatag gtggtggggg atg                             23

```


-continued

source 1..21
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 80
 gagccaagct ctccatctag t 21

SEQ ID NO: 81 moltype = DNA length = 20
 FEATURE Location/Qualifiers
 misc_feature 1..20
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..20
 note = Primer
 source 1..20
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 81
 gccctgtcaa gagttgacac 20

SEQ ID NO: 82 moltype = DNA length = 22
 FEATURE Location/Qualifiers
 misc_feature 1..22
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..22
 note = Primer
 source 1..22
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 82
 gcacagggtg gaacaagatg ga 22

SEQ ID NO: 83 moltype = DNA length = 24
 FEATURE Location/Qualifiers
 misc_feature 1..24
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..24
 note = Primer
 source 1..24
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 83
 gccaggtacc tatcgattgt cagg 24

SEQ ID NO: 84 moltype = DNA length = 21
 FEATURE Location/Qualifiers
 misc_feature 1..21
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..21
 note = Primer
 source 1..21
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 84
 gagccaagct ctccatctag t 21

SEQ ID NO: 85 moltype = DNA length = 20
 FEATURE Location/Qualifiers
 misc_feature 1..20
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..20
 note = Primer
 source 1..20
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 85
 actctgactg ggtcaccagc 20

SEQ ID NO: 86 moltype = DNA length = 20
 FEATURE Location/Qualifiers
 misc_feature 1..20
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..20
 note = Primer
 source 1..20

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 86
tatttgctg gttgaaaggg                20

SEQ ID NO: 87      moltype = DNA length = 24
FEATURE           Location/Qualifiers
misc_feature      1..24
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..24
                  note = Primer
source            1..24
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 87
aaagtcatga aataaacaca ccca          24

SEQ ID NO: 88      moltype = DNA length = 24
FEATURE           Location/Qualifiers
misc_feature      1..24
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..24
                  note = Primer
source            1..24
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 88
ctgcattgat atggtagtag catg          24

SEQ ID NO: 89      moltype = DNA length = 21
FEATURE           Location/Qualifiers
misc_feature      1..21
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..21
                  note = Primer
source            1..21
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 89
gctgttcatt gcaatggaat g             21

SEQ ID NO: 90      moltype = DNA length = 22
FEATURE           Location/Qualifiers
misc_feature      1..22
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..22
                  note = Primer
source            1..22
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 90
atggagttgg acatggccat gg            22

SEQ ID NO: 91      moltype = DNA length = 28
FEATURE           Location/Qualifiers
misc_feature      1..28
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..28
                  note = Primer
source            1..28
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 91
actcactatc cacagttcag catttacc      28

SEQ ID NO: 92      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
misc_feature      1..23
                  note = Primer
source            1..23
                  mol_type = other DNA

```

-continued

organism = synthetic construct
 SEQUENCE: 92
 tggagatagc tgtcagcaac ttt 23

SEQ ID NO: 93 moltype = DNA length = 29
 FEATURE Location/Qualifiers
 misc_feature 1..29
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..29
 note = Primer
 source 1..29
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 93
 caacaaagca aaggtaaagt tgtaatatag 29

SEQ ID NO: 94 moltype = DNA length = 25
 FEATURE Location/Qualifiers
 misc_feature 1..25
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..25
 note = Primer
 source 1..25
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 94
 ggtttcagga gatgtgttac aaggc 25

SEQ ID NO: 95 moltype = DNA length = 27
 FEATURE Location/Qualifiers
 misc_feature 1..27
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..27
 note = Primer
 source 1..27
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 95
 gattgtgcaa ttcctatgca atcggtc 27

SEQ ID NO: 96 moltype = DNA length = 25
 FEATURE Location/Qualifiers
 misc_feature 1..25
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..25
 note = Primer
 source 1..25
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 96
 cactgggtac ttaatctgta gcctc 25

SEQ ID NO: 97 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..23
 note = Primer
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 97
 ggttccaagt cattcccagt agc 23

SEQ ID NO: 98 moltype = DNA length = 30
 FEATURE Location/Qualifiers
 misc_feature 1..30
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 misc_feature 1..30
 note = Primer
 source 1..30
 mol_type = other DNA
 organism = synthetic construct

-continued

SEQUENCE: 98
catcactgca gttgtaggtt ataactatcc 30

SEQ ID NO: 99 moltype = DNA length = 26
FEATURE Location/Qualifiers
misc_feature 1..26
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..26
 note = Primer
source 1..26
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 99
ttgaaaacca cagatctggt tgaacc 26

SEQ ID NO: 100 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_feature 1..22
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..22
 note = Primer
source 1..22
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 100
ggagtgccaa gagaatatct gg 22

SEQ ID NO: 101 moltype = DNA length = 32
FEATURE Location/Qualifiers
misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..32
 note = Primer
source 1..32
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 101
ctgaaaactgg tttcaaaata ttcgtttaa gg 32

SEQ ID NO: 102 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_feature 1..22
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..22
 note = Primer
source 1..22
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 102
gctctgtatg ccctgtagta gg 22

SEQ ID NO: 103 moltype = DNA length = 22
FEATURE Location/Qualifiers
misc_feature 1..22
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..22
 note = Primer
source 1..22
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 103
ttgcatctg accttacctt tg 22

SEQ ID NO: 104 moltype = DNA length = 23
FEATURE Location/Qualifiers
misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
misc_feature 1..23
 note = Target sequence of RGEN
source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 104

-continued

```

aatgaccact acatcctcaa ggg 23

SEQ ID NO: 105      moltype = DNA length = 23
FEATURE            Location/Qualifiers
misc_feature        1..23
                    note = Description of Artificial Sequence: Synthetic
                    oligonucleotide
misc_feature        1..23
                    note = Target sequence of RGEN
source              1..23
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 105
agatgatgtc tcatcatcag agg 23

SEQ ID NO: 106      moltype = DNA length = 4170
FEATURE            Location/Qualifiers
misc_feature        1..4170
                    note = Description of Artificial Sequence: Synthetic
                    polynucleotide
misc_feature        1..4170
                    note = Cas9-coding sequence in p3s-Cas9HC (humanized,
                    C-term tagging, human cell experiments)
source              1..4170
                    mol_type = other DNA
                    organism = synthetic construct

SEQUENCE: 106
atggacaaga agtacagcat cggcctggac atcgggtacca acagcgtggg ctgggcccgtg 60
atcaccgcagc agtacaaggt gccagcaag aagttcaagg tgctgggcaa caccgaccgc 120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgcccag 180
gccaccgcc tgaagcgcac gcgccgccg cgctacaccc gccgcaagaa ccgcatctgc 240
tacctgcagg agatcttcag caacgagatg gccaaagtg acgacagctt cttccaccgc 300
ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgcaccc catctctcgc 360
aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccattacca cctgcgcaag 420
aagctggtgg acagcaccga caaggccgac ctgcccctga tctacctggc cctggcccac 480
atgatcaagt tccgcgccca cttcctgatc gaggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttoga ggagaacccc 600
atcaacgcca gggcggtgga gcccaagccc atcctgagcg cccgcctgag caagagccgc 660
cgccctggaga acctgatcgc ccagctgccc ggcgagaaga agaaccgccc gttcggcaac 720
ctgatcgccc tgagcctggg cctgaccccc aactcaaga gcaacttoga cctggccgag 780
gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttc ctggcccga agaacctgag cgacgccatc 900
ctgctgagcg acatcctgcg cgtgaacacc gagatcacca agccccccc gagcgccagc 960
atgatcaagc gctacgacga gcaccaccag gacctgacc tgctgaaggg cctgggtgcg 1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacatcg accgcccggc cagccaggag gagttctaca agtctatcaa gccctactg 1140
gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgccc 1200
aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgccc 1260
gccatcctgc tgcgcaagga ggactcttac cccttctga aggacaaccg cgagaagatc 1320
gagaagatcc gacctctccg catcccctac tacgtggggc ccctggcccc cggaacagc 1380
cgcttcgccc ggtgacccc caagagcgag gagaccatca cccctggaa cttcggagag 1440
gtggtggaaca agggcccccag gcgccagagc ttcctcgagc gcatgaccaa cttcgacaag 1500
aacctgcccc acgagaaggt gctgcccgaag cacagcctcg tgtagcagta cttcaccgtg 1560
tacaacgagc tgaaccaaggt gaagtacgtg accgaggcca tgcgcaagcc cgccttctg 1620
agcggcgagc agaagaaggt caactgtggac ctgctgttca agaccaaacc caagtgacc 1680
gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740
agcggcgctg aggaccgctt caacgccagc ctgggcaacc accacagctt gctgaagatc 1800
atcaaggaca aggacttccc ggacaacgag gagaacgagg acatcctgga ggacatgctg 1860
ctgaccctga cctgttctga ggaccgagc atgatcgagg agcgcctgaa gacctacgcc 1920
cacctgtctc acgacaaggt gatgaagcag ctgaagcacc gccgctacac cggctggggc 1980
cgcctgagcc gcaagcttat caacgcatc cgccacaagc agagcggcaa gacctactg 2040
gacttctcga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100
agcctgacct tcaaggagga catccagaag gccacagtgga gcccggaggg cgacagcctg 2160
cacgagcaca tcgccaacct gcccgcgagc cccgccatca agaagggcat cctgcagacc 2220
gtgaaggtgg tggacgagct ggtgaaggtg atgggcccgc acaagcccga gaacatcgtg 2280
atcgagatgg cccgagagaa ccagaccacc cagaagggcc agaagaacag ccgagagcgc 2340
atgaagcgca tccaggaggg catcaaggag ctgggagacc agatcctgaa ggagcacc 2400
gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccc 2460
gacatgtacg tggaccagga gctggacatc aaccgcccga gcgactacga cgtggaccac 2520
atcgtgcccc agagcttccc gaaggacgac agcatcgaca acaaggtgct gaccgcccgc 2580
gacaagaacc gccgcaagag cgacaacgtg cccagcgagg aggtgggtaa gaagatgaag 2640
aactactggc gccagctgct gaacgccaag ctgatcacc agcgcgaagt cgacaacctg 2700
accaagggcg accgcccggc cctgagcgag ctggacaagg cccgcttcat caagcggcag 2760
ctggtggaga cccgccagat caccagcac gtggcccaga tctggacag ccgcatgaac 2820
accaagtacg acgagaacga caagctgatc cgcgagtgga aggtgatcac cctgaagagc 2880
aagctggtga gcgacttccc caaggacttc cagttctaca aggtgcccga gatcaacaac 2940
taccaccagc cccacgacgc ctacctgaac gccgtggtg gcaccgccc gatcaagaag 3000
taccacaagc cggagagcga gttcgtgat ggcgactaca aggtgtacga cgtgcgcaag 3060
atgatcgcca agagcgagca ggagatcgcc aaggccaccg ccaagtactt cttctacagc 3120

```

-continued

```

aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc 3180
cccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc 3240
gccaccgtgc gcaaggtgct gagcatgcc caggtgaaca tcgtgaagaa gaccgaggtg 3300
cagaccggcg gcttcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgatc 3360
gcccgcaaga aggactggga ccccaagaag tacggcggct tcgacagccc caccgtggcc 3420
tacagcgtgc tgggtggggc caaggtggag aagggaaga gcaagaagct gaagagcgtg 3480
aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540
ttcctggagg ccaaggggta caaggagggt aagaaggacc tgatcatcaa gctgcccagg 3600
tacagcctgt tcgagctgga gaaccggccc aagcgcctgc tggccagcgc cggcgagctg 3660
cagaagggca acgagctggc ctgcccagc aagtagctga acttctgtg cctggccagc 3720
cactacgaga agctgaaggg cagcccggag gacaacgagc agaagcagct gttcgtggag 3780
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840
atcctggccg agcccaacct ggacaaggtg ctgagcgctt acaacaagca ccgcgacaag 3900
cccatccggc agcagggcga gaacatcatc cacctgttca ccctgaccaa cctggggcgc 3960
cccggccgct tcaagtaact cgacaccacc atcgaccgca agcgtctacc cagcaccagg 4020
gaggtgctgg agcggccctt gatccaccag agcatcaccg gtctgtacga gaccgcgatc 4080
gacctgagcc agctggggcg cagcggcctc caaagaaaaa gagaaaagta 4140
taccctacg acgtgcccga ctacgcctaa 4170

```

```

SEQ ID NO: 107      moltype = DNA length = 4194
FEATURE            Location/Qualifiers
misc_feature       1..4194
                   note = Description of Artificial Sequence: Synthetic
                   polynucleotide
misc_feature       1..4194
                   note = Cas9 coding sequence in p3s-Cas9HN (humanized codon,
                   N-term tagging (underlined), human cell experiments)
source             1..4194
                   mol_type = other DNA
                   organism = synthetic construct

```

```

SEQUENCE: 107
atggtgtacc cctacgagct gcccgactac gccgaattgc tcctcaaaaa gaagagaaaag 60
gtagggatcc gaattcccgg gaaaaaaccc gacaagaagt acagcatcgg cctggacatc 120
ggtaccaaca gcgtgggctg ggccgtgatc accgacgagt acaaggtgcc cagcaagaag 180
ttcaaggtgc tgggcaaacac cgaccgccac agcatcaaga agaacctgat cggcgccctg 240
ctggttcgaca gcgcgagagc cgccgaggcc accgcctcga agcgcaccgc ccgcccggcg 300
tacaccggcc gcaagaaccg catctgtatc ctgacggaga tcttcagcaa cgagatggcc 360
aaggtggagc acagcttctt ccaccgctct gaggagagct tcctgggtgga ggaggacaag 420
aagcacgagc gccaccctct ctccggcaac atcgtggagc aggtggccta ccacgagaag 480
taccaccaaa tctaccacct gcgcaagaag ctggtggaca gcaccgaaa ggccgacctg 540
cgctgatctt acctggccct ggcccacatg atcaaagttc ggggccactt cctgatcgag 600
ggcgacctga accccgacaa cagcgcagct gacaagctgt tcatccagct ggtgcagacc 660
tacaaccagc tgttcgagga gaaccctatc aacgcacgcg gcctggagcg caaggccatc 720
ctgagcggcc gcctgagcaa gagccggccc ctggagaacc tgatcgccca gctgcccggc 780
gagaagaaga acggcctggt cggcaacctg atcgccctga gcctgggccc gacccccacc 840
ttcaagagca acttcgacct ggccgaggac gccaaagctgc agctgagcaa ggacacctac 900
gacgacgacc tggacaacct gctggcccag atcggcgacc agtacgcccga cctgttctctg 960
gcccgaaga acctgagcgc cgccatcctg ctgagcgaca tcctgcgctg gaacaccgag 1020
atcaccaggc cccctctgag gcaccgcatg atcaagcgtc acgacgagca ccaccggagc 1080
ctgacctctg tgaaggccct ggtgcccagc cagctgcccg agaagtacaa ggagatcttc 1140
ttcgaccaga gcaagaacgg ctacgcccgc tacatcgagc gcgcgcccag ccaggaggag 1200
ttctacaagt tctcaagctc catcctggag aagatggagc gcaccgagga gctgtggtg 1260
aagctgaacc gcgaggacct gctgcccagc cagcgcacct tcgacaacgg cagcatcccc 1320
caccagatcc acctggggcg gctgcaagcc atcctgcccg gccaggagga cttctacccc 1380
ttcctgaagg acaaccgcaa aaagatcgag aagatcctga ccttccgcat cccctactac 1440
gtgggccccg tggcccggcg caacagcccg ttcgctcgtg tgaccggcaa gagcgaggag 1500
accatcacc cctggaactt cgaggagggt gtggacaagg gcgcccagcg ccagagcttc 1560
atcgagcgca tgaccaactt cgacaagaac ctgcccacag agaaggtgct gcccaagcac 1620
agcctgctgt acgagtactt caccgtgtac aacgagctga ccaaggtgaa gtacgtgacc 1680
gagggcatgc gcaagcccgc cttcctgagc ggcgagcaga agaaggccat cgtggacctg 1740
ctgttcaaga ccaaccgcaa ggtgaccgtg aagcagctga aggaggacta cttcaagaag 1800
atcgagtgct tcgacagcgt ggagatcagc ggcgtggagg accgcttcaa cgcagcctg 1860
ggcacctacc acgacctgct gaagatcatc aaggacaagg acttctcgtg caacgaggag 1920
aacgaggaca tcctggagga catcgtgctg accctgacct tgttcgagga cgcgagatg 1980
atcgaggagc gcctgaagac ctacgcccac ctggttcgag acaaggtgat gaagcagctg 2040
aagcggcccg gctacaccgg ctggggcccg ctgagcccga agcttataca cggcatcccc 2100
gacaagcaga gcggaagac catcctggac ttctgaaga gcgacggctt cgccaaccgc 2160
aactcatgac agctgatcca cagcagcagc ctgaccttca aggaggacct ccagaaggcc 2220
caggtgagcg gccaggcgga cagcctgcac gagcacatcg ccaacctggc cggcagcccc 2280
gccatcaaga agggcatcct gcagaccgtg aaggtgggtg acgagctggt gaaggtgatg 2340
ggccggccaca agcccagaa catcgtgatc gagatggccc gcgagaacca gaccaccag 2400
aagggccaga agaacagccc cgagcgcctg aagcgcctcg aggagggcat caaggagctg 2460
ggcagccaga tcctgaagga gcaccccgtg gagaacaccc agctgcagaa cgagaagctg 2520
tacctgtact acctgagaa ggcccggcag atgtacgtgg accagagctg ggacatcaac 2580
cgcctgagcg actacagct ggaccacatc gtgcccaga gcttctgaa ggacgacagc 2640
atcgacaaca aggtgctgac ccgcagcgac aagaaccggc gcaagagcga caactgccc 2700
agcagggagg tggtaagaa gatgaagaac tactggcgcc agctgctgaa cgccaagctg 2760
atcaccagc gcaggttcga caacctgacc aaggccgagc gcgccggcct gagcagctg 2820
gacaaggccg gcttcatcaa gcgcccagct gtggagacc gccagatcac caagcagctg 2880

```

-continued

```

ccccagatcc tggacagccg catgaacacc aagtaacgac agaacgacaa gctgatccgc 2940
gaggtgaagg tgatecaccct gaagagcaag ctgggtgagcg acttccgcaa ggacttccag 3000
ttctacaagg tgcgcgagat caacaactac caccacgccc acgacgccta cctgaacgcc 3060
gtggtgggca ccgcccctgat caagaagtac cccaagctgg agagcaggtt cgtgtacggc 3120
gactacaagg tgtacgacgt gcgcaagatg atcgccaaga gcgagcagga gatcggcaag 3180
gccaccgcca agtacttctt ctacagcaac atcatgaact tcttcaagac cgagatcacc 3240
ctggccaacg gcgagatccg caagcgccc ctgatcgaga ccaacggcga gaccggcgag 3300
atcgtgtggg acaagggccg cgacttcgccc accgtgcgca aggtgctgag catgcccag 3360
gtgaacatcg tgaagaagac cgaggtgcag accggcgctc tcagcaagga gagcatcctg 3420
cccaagcgca acagcgacaa cgtgatcgcc cgcaagaagg actgggacc caagaagtac 3480
ggcggcttcg acagcccac gctggcctac agcgtgctgg tggtaggcaa ggtggagaag 3540
ggcaagagca agaagctgaa gagcgtgaa gagctgctgg gcatcaccat catggagcgc 3600
agcagcttcg aagaagaacc ctcgacttc ctggaggcca agggctacaa ggaggtgaag 3660
aagacctga tcatcaagct gcccaagtac agcctgttcg agctggagaa cggcccagag 3720
cgcatgctgg ccagcgcggg cgagctgcag aaggccaacg agctggccct gccacgcaag 3780
tacgtgaact tcctgtacct ggccagccc tacgagaagc tgaagggcag ccccgaggac 3840
aacgagcaga agcagctggt cgtggagcag cacaagcact acctggacga gatcctcagag 3900
cagatcagcg agttcagcaa gcgcgtgatc ctggccgacg ccaacctgga caaggtgctg 3960
agcgcttaca acaagcacc gcacaagccc atccgcgagc aggcggagaa catcatccac 4020
ctgttcaacc tgaccaacct gggcgcccc gccgccttca agtacttoga caccaccatc 4080
gaccgcaagc gctacaccag caccaaggag gtgctggacg ccacctgat ccaccagagc 4140
atcaccggtc tgtacgagac ccgcatcgac ctgagccagc tggcgggcga ctaa 4194

```

```

SEQ ID NO: 108      moltype = DNA length = 4107
FEATURE
misc_feature       1..4107
                   note = Description of Artificial Sequence: Synthetic
                   polynucleotide
misc_feature       1..4107
                   note = Cas9-coding sequence in Streptococcus pyogenes
source             1..4107
                   mol_type = other DNA
                   organism = synthetic construct

```

```

SEQUENCE: 108
atggataaga aataactcaat aggccttagat atcggcacia atagcgtcgg atggcgcggtg 60
atcactgatg aatataaagg tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120
cacagtatca aaaaaaatct tataggggct cttttatttg acagtgagga gacagcggaa 180
gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
tatctacagg agattttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300
cttgaagagt ctttttgggt ggaagaagac aagaagcatg aacgctatcc tatttttggg 360
aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc ctttagcgcat 480
atgattaaat ttcgttggtc tttttgatt gagggagatt taaatcctga taatagtgat 540
gtggacaaac tattttacca gttggtacaa acctacaatc aattatttga agaaaacct 600
ataacgcaa gtggagtaga tgctaaagcg attctttctg cagcattgag taaatcaaga 660
cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaa 720
ctcattgctt tgcattggg tttgaccctc aattttaaat caaatttga tttggcagaa 780
gatgctaaat tacagctttc aaaagatact tacgatgatg attagataa tttattggcg 840
caaattggag atcaaatgct tgatttgggt ttggcagcta agaattatc agatgctatt 900
ttactttcag atatcctaag agttaaatact gaaataacta aggcctccct atcagcttca 960
atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
caacaacttc cgaaaaagta taaagaatc ttttttgatc aatcaaaaaa cggatagca 1080
ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
aagcaacgga cctttgacaa cggctctatt ccccatcaaa ttcacttggg tgagctgcat 1260
gctattttga gaagacaaga agacttttat ccatttttaa aagacaatcg tgagaagatt 1320
gaaaaaatct tgacttttcg aatctcttat tatgttggtc cattggcgcg tggcaatagt 1380
cgttttgcat ggtgactcgc gaagtctgaa gaaacaatta cccatgggaa ttttgaagaa 1440
gttgcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa 1500
aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacgggt 1560
tataacgaaat tgacaaaagc caaatatggt actgaaggaa tgcgaaaaac agcatttctt 1620
tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680
gttaagcaat taaaagaaga ttatttcaaaa aaaatagaat gttttgatag tgttgaatt 1740
tcaggagtgt aagatagatt taatgcttca ttaggtacct accatgattg gctaaaatt 1800
attaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattggt 1860
ttaacattga ccttatttga agatagggag atgatggagg aaagacttaa aacatagct 1920
cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980
cgtttgcctc gaaaattgat taaatgatt agggataagc aatctggcaa aacaattata 2040
gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat 2100
agtttgacat ttaagaaga cattcaaaaa gcacaagtggt ctggacaagg cgatagttta 2160
catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact 2220
gtaaaagtgt ttgatgaatt ggtcaaagta atggggcgcc ataagccaga aaatctcgtt 2280
attgaaatgg ccagtgaaaa tcagacaact caaaaaggccc agaaaaattc gcgagagcgt 2340
atgaaacgaa tcgaagaagg tatcaaaagaa tttaggaagtc agattcttaa agagcatcct 2400
gttgaaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460
gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcogatc 2520
attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtcct aacgcgttct 2580
gataaaaatc gtgtaaatc ggataacggt ccaagtgaag aagtagtcaa aaagatgaaa 2640
aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta 2700

```

-continued

```

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgcca 2760
ttggttgaaa ctcgccaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat 2820
actaataacg atgaaaatga taacttatt cgagaggtta aagtgtattc cttaaaaatc 2880
aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat 2940
taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000
tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060
atgattgcta agtctgagca agaaataggg aaagcaaccg caaaatattt cttttactct 3120
aataatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
cctctaactg aaactaatgg ggaactgga gaatttgtct gggataaagg gcgagatatt 3240
gccacagtgc gcaaatgatt gtccatgcc caagtcaata ttgtcaagaa aacagaagta 3300
cagacaggcg gattctccaa ggagtcaatt ttacaaaaa gaaattcgga caagcttatt 3360
gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtc aacggtagct 3420
tattcagtc tagtggttgc taaggtggaa aaagggaaat cgaagaagt aaaatcogtt 3480
aaagagttac tagggatcac aatgatggaa agaagttcct ttgaaaaaaa tccgattgac 3540
tttttagaag ctaaaggata taaggaagt aaaaaagact taatcatata actacctaaa 3600
tatagtcctt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660
caaaaagaaa atgagctggc tctgccaagc aaatatgtga atttttata tttagctagt 3720
cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780
cagcataagc attatttaga tgagattatt gagcaaatca gtgaatttct taagcgtggt 3840
attttagcag atgccaatct agataaagtt cttagtgcac ataacaaca tagagacaaa 3900
ccaatcagtg aacaagcaga aaatattatt cattattata cgttgacgaa tcttggagct 3960
cccgtgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
gaagtttag atgcccactc tatccatcaa tccatcactg gtctttatga aacacgcat 4080
gatttgagtc agctaggagg tgactaa 4107

```

```

SEQ ID NO: 109      moltype = AA length = 1368
FEATURE           Location/Qualifiers
REGION           1..1368
                  note = Description of Artificial Sequence: Synthetic
                  polypeptide
REGION           1..1368
                  note = Amino acid sequence of Cas9 from S.pyogenes
source           1..1368
                  mol_type = protein
                  organism = synthetic construct

```

```

SEQUENCE: 109
MDKKYSIGLD  IGTNSVGVAV  ITDEYKVPK  KFKVLGNTDR  HSIKKNLIGA  LLFDSGETAE  60
ATRLKRTARR  RYTRRKNRIC  YLQEIFSNEM  AKVDDSPFHR  LEESFLVEED  KKHERHPIFG  120
NIVDEVAYHE  KYPPIYHLRK  KLVDSDDKAD  LRLIYLALAH  MIKFRGHFLI  EGDLPDNDSD  180
VDKLFILQVQ  TYNQLFEENP  INASGVDAKA  ILSARLSKSR  RLENLIAQLP  GEKKNGLFGN  240
LIALSLGLTP  NFKSNFDLAE  DAKLQLSKDT  YDDDLNLLA  QIGDQYADLF  LAAKNLSDAI  300
LLSDILRVNT  EITKAPLSAS  MIKRYDEHQ  DLTLKALVR  QQLPEKYKEI  FPDQSKNGYA  360
GYIDGGASQE  EFKYKFIKPI  EKMDGTEELL  VKLNREDLLR  KQTFDNGSI  PHQIHLGELH  420
AILRQEDFY  PFLKDNREKI  EKILTFRIPY  YVGPLARGNS  RFAWMTRKSE  ETITPWNFEE  480
VVDKGASQ  FIERMTNFDK  NLPNEKVLPK  HSLLYEYFTV  YNELTKVKYV  TEGMRKPAPL  540
SGEQKKAIVD  LLPKTRKVT  VKQLKEDYFK  KIECFDSVEI  SGVEDRPNAS  LGTYHDLKI  600
IKDKDFLDNE  ENEDILEDIV  LTLTLFEDRE  MIEERLKTYA  HLPDDKVMKQ  LKRRRYTGWG  660
RLSRKLINGI  RDKQSGKITL  DPLKSDGFAN  RNFMQLIHDD  SLTFKEDIQK  AQVSGQDLS  720
HEHIANLAGS  PAIKKGLIQT  VKVDELVKV  MGRHKPENIV  IEMARENQTT  QKGQNSRER  780
MKRIEEDIKE  LGSQILKEHP  VENTQLQNEK  LYLYYLQNGR  DMYVDQELDI  NRLSDYDVDH  840
IVPQSFLLKDD  SIDNKVLTRS  DKNRGSDNV  PSEEVVKKMK  NYWRQLLNAK  LITQRKFDNL  900
TKAERGGLE  LDKAGFIKRO  LVETROITKH  VAQILDSDRM  TKYDENDKLI  REVKVIITLKS  960
KLVSDFRKDF  QFYKVVREINN  YHHAHDAYLN  AVVGTALIKK  YPKLESEFVY  GDYKVYDVRK  1020
MIAKSEQEIG  KATAKYPFYS  NIMNFPKTEI  TLANGEIRKR  PLIETNETG  EIVWDKGRDF  1080
ATVRKVLSP  QNVLVNKRTEV  QTGGFSKESI  LPKRNSDKLI  ARKKDWDPKK  YGGFDSPTVA  1140
YSVLVVAKVE  KGKSKLKS  V  KELLGITIME  RSSFEKNPID  FLEAKGYKEV  KKDIIKLPK  1200
YSLFELENGR  KRMLASAGEL  QKGNELALPS  KYVNFYLLAS  HYEKLGSP  DNEQKQLFVE  1260
QHKHYLDEII  EQISEFSKRV  ILADANLDKV  LSAYNKHRDK  PIREQAENII  HLFTLTNLGA  1320
PAAFKYFDTT  IDRKRYTSTK  EVLDATLIHQ  SITGLYETRI  DLSQLGGD  1368

```

```

SEQ ID NO: 110      moltype = DNA length = 4221
FEATURE           Location/Qualifiers
misc_feature     1..4221
                  note = Description of Artificial Sequence: Synthetic
                  polynucleotide
misc_feature     1..4221
                  note = Cas9-coding sequence in pET-Cas9N3T for the
                  production of recombinant Cas9 protein in E. coli
                  (humanized codon; hexa-His-tag and a nuclear localization
                  signal at the N terminus)
source           1..4221
                  mol_type = other DNA
                  organism = synthetic construct

```

```

SEQUENCE: 110
atggcgagca gccatcatca tcatcatcat gtgtaccctc acgacgtgcc cgactacgcc 60
gaattgcctc caaaaaagaa gagaaaggta gggatcgaga acctgtactc ccaggggcag 120
aagaagtaca gcactcggcct ggacatcggt accaacacgc tgggctgggc cgtgatcacc 180
gacgagtaca aggtgcccag caagaagttc aaggtgctgg gcaacaccga cgcgccacg 240
atcaagaaga acctgatcgg cgcctcgtct ttcgacagcg gcgagaccgc cgaggccacc 300

```

-continued

```

cgccctgaagc gcaccgccc cgcgccgtac acccgccgca agaaccgcat ctgtactctg 360
caggagatct tcagcaacga gatggccaag gtggaacgaca gctttttcca ccgocctggag 420
gagagcttcc tgggtggagga ggacaagaag cacgagcgcc accccatctt cggcaacatc 480
gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgcg caagaagtctg 540
gtggacagca ccgacaaggc cgacctgccc ctgatctacc tggccctggc ccacatgatc 600
aagttccgcg gccacttctt gatcgagggc gacctgaacc ccgacaacag cgacgtggac 660
aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggagaa ccccatcaac 720
gccagcggcg tggacgcca ggccatcctg agcgcgccgc tgagcaagag ccgccocctg 780
gagaacctga tcgcccagct gcccgcgag aagaagaacg gcctgttcgg caacctgatc 840
gccctgagcc tggcctgac ccccaacttc aagagcaact tcgacctggc cgaggagccc 900
aagctgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc 960
ggcgaccagt acgcccagct gttcctggcc gccaaagaac tgagcgagcg catcctgctg 1020
agcgacatcc tgcgcgtgaa caccgagatc accaaggccc ccctgagcgc cagcatgatc 1080
aagcgtctac ccgagagact ccaggacctg acctgctga aggcctggg gcgcccagcag 1140
ctgcccgaga agtacaagga gatcttcttc gaccagagca agaaccgcta cgcggctac 1200
atcgacggcg cgcgcaagca ggaggagtct tacaagtcca tcaagcccat cctggagaag 1260
atggacggca ccgagagact gctggtgaa gtgaaccgag aggaacctgt gcgcaagcag 1320
cgcaacctcg acaaccgag catccccac cagatccacc tggcgagctg gcacgccatc 1380
ctgcccgcgc aggaggact ctacccttc ctgaaggaca acccgagaa gatcgagaag 1440
atcctgacct tccgcaatccc ctactacgtg ggccccctgg cccgcccga cagcccctt 1500
gcctggatga cccgcaagag cgaggagacc atcaccccc ggaacctcga ggaggtggg 1560
gacaaggcgg ccagcgccca gagcttcat gagcgcgatg ccaacttcga caagaacctg 1620
cccaacgaga aggtgctgccc caagcacagc ctgctgtacg agtacttcc cgtgtacaac 1680
gagctgacca aggtgaagat cgtgaccgag ggcctgcgca agcccgcctt cctgagcggc 1740
gagcagaaga agggcatcgt ggacctgctg ttcaagacca accgcaaggt gacogtgaag 1800
cagctgaagg aggaactct caagaagatc gagtctctcg acagcgtgga gatcagcggc 1860
gtggaggacc gcttcaacgc cagcctgggc acctaccagc acctgctgaa gatcatcaag 1920
gacaaggact tccggcaaa cgaggagaac gaggacatcc tggaggacat cgtgctgacc 1980
ctgacctgtg tcgaggaccg cgagatgatc gagggagccc tgaagacctc cgcccacctg 2040
ttcgacgaca aggtgatgaa gcagctgaa cgccgcccgt acaccggctg gggcccctg 2100
agccgcaagc ttatcaacgg catcccgac aagcagagcg gcaagaccat cctggacctt 2160
ctgaagagcg accgctctgc caaccgcaac ttcatgagc tgatccacga cgacagcctg 2220
accttcaagg aggcacatcca gaaggcccag gtgagcggcc agggcgacag cctgcaagcag 2280
cacatcgcca acccgcccgg cagccccccc atcaagaagg gcatcctgca gaccgtgaag 2340
gtggtggagc agctgggtgaa ggtgatgggc cgccacaagc ccgagaacat cgtgatcgag 2400
atggcccggc agaaccagac caccagaag ggccagaaga acagcccgca ggcgatgaag 2460
cgctcgagc agggcatcaa ggagctgggc agccagatcc tgaaggagca cccogtggag 2520
aacaccacgc tgcagaacga gaagctgtac ctgtactacc tgcagaacgc ccgagacatg 2580
tacctggacc aggagctgga catcaaccgc ctgagcagct acgacgtgga ccacatcgtg 2640
ccccagagct tccgcaagga cgacagcatc gacaacaagg tgtgacccc cagcgacaag 2700
aaccgcccga agagcgacaa cgtgccaccg gaggaggtgg tgaagaagat gaagaactac 2760
tggcgccagc tgcgtaacgc caagctgatc acccagcga agttcgacaa cctgaccaag 2820
gccagcggcg cgcgcccagc cgagctggac aaggcccggc tcatcaagcg ccagctggtg 2880
gagaccgccc agatcaccaa gcacgtggcc cagatcctgg acagccgat gaacaccaag 2940
tacgacgaga acgacaagct gatccgagag gtgaaggtag tcacctgaa gagcaagctg 3000
gtgagcagct tccgcaagga cttccagttc tacaaggtgc gcgagatcaa caactaacac 3060
cacgcccagc acgctacct gaacgcccgt gtgggcaacc ccctgatcaa gaagtacccc 3120
aagctggaga gcgagttcgt gtacggcgac tacaaggtgt acgactgtgc caagatgatc 3180
gccaaagagc agcaggagat cggcaaggcc accgccaagt acttttcta cagcaacatc 3240
atgaactctc tcaagaccga gatcacccctg gccaacggcg agatccgcaa gcgccccctg 3300
atcgagacca accgagagac cggcgagatc gtgtgggaca agggcccgca cttcgccacc 3360
gtgcgcaagg tgcgagcatc gccccaggtg aacatcgtga agaagaccga ggtgcagacc 3420
ggcggcttca gcaaggagag catcctgccc aagcgcaaca gcgacaagct gatcgcccgc 3480
aagaaggact gggaccccaa gaagtacggc ggtctcgaca gcccccactg ggcctacagc 3540
gtgctggtgt tggccaaggt ggagaagggc aagagcaaga agctgaagag cgtgaaggag 3600
ctgctgggca tcacctcat ggagcgacg agcttcgaga agaaccctc cgacttctctg 3660
gagggcaagg gctacaagga ggtgaagaag gacctgatca tcaagctgcc caagtacagc 3720
ctgctcagc tggagaacgg ccgcaagcgc atgctggcca gcgcccggca gctgcagaag 3780
ggcaacgagc tggccctgccc cagcaagtac gtgaacttcc tgtacctggc cagccactac 3840
gagaagctga agggcagccc cgaggacaac gacgagaagc agctgttctg ggagcagc 3900
aagcactacc tggacgagat ctcagcagc atcagcagc tcaagcagc cgtgatcctg 3960
gccgacgcca acctggacaa ggtgctgagc gctacaaca agcaccgca caagcccatc 4020
cgcgagcagg ccgagaacat catcccactg ttcaccctga ccaacctggg cgccccgcc 4080
gcttcaagt acttcgacac caccatcgac cgcaagcgtc acaccagcac caaggaggtg 4140
ctggacgcca cctgatcca ccagagcacc accggtctgt acgagaccgc catcgacctg 4200
agccagctgg gcgcgacta a 4221

```

```

SEQ ID NO: 111          moltype = AA length = 1406
FEATURE                Location/Qualifiers
REGION                 1..1406
                        note = Description of Artificial Sequence: Synthetic
                        polypeptide
REGION                 1..1406
                        note = Amino acid sequence of Cas9 (pET-Cas9N3T)
source                 1..1406
                        mol_type = protein
                        organism = synthetic construct
SEQUENCE: 111
MGSSHHHHHH VYPYDVPDYA ELPPKKRKY GIENLYFQGD KKYSIGLDIG TNSVGWAVIT 60

```


-continued

misc_feature 1..35
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..35
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 117
caatctatga catcgagacc ctgccaaaa atcaa 35

SEQ ID NO: 118 moltype = DNA length = 31
FEATURE Location/Qualifiers

misc_feature 1..31
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..31
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 118
caatctatga catgccctgc caaaaaatca a 31

SEQ ID NO: 119 moltype = DNA length = 30
FEATURE Location/Qualifiers

misc_feature 1..30
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..30
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 119
caatctatga catcaattat tataaatcaa 30

SEQ ID NO: 120 moltype = DNA length = 25
FEATURE Location/Qualifiers

misc_feature 1..25
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..25
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 120
caatctatga catccaaaaa atcaa 25

SEQ ID NO: 121 moltype = DNA length = 19
FEATURE Location/Qualifiers

misc_feature 1..19
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..19
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 121
caatctatga caaaatcaa 19

SEQ ID NO: 122 moltype = DNA length = 46
FEATURE Location/Qualifiers

source 1..46
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 122
tatgtgcaat gaccactaca tcttcaaggg cagcaatcgg agccag 46

SEQ ID NO: 123 moltype = DNA length = 47
FEATURE Location/Qualifiers

misc_feature 1..47
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..47
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 123
tatgtgcaat gaccactaca tcttcaagg gcagcaatcg gagccag 47

SEQ ID NO: 124 moltype = DNA length = 48
FEATURE Location/Qualifiers

misc_feature 1..48
note = Description of Artificial Sequence: Synthetic
oligonucleotide

source 1..48
mol_type = other DNA

-continued

SEQUENCE: 124	organism = synthetic construct	
tatgtgcaat gaccactaca tcctctcaag ggcagcaatc ggagccag		48
SEQ ID NO: 125	moltype = DNA length = 18	
FEATURE	Location/Qualifiers	
misc_feature	1..18	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..18	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 125		
tatgtgcaat ggagccag		18
SEQ ID NO: 126	moltype = DNA length = 13	
FEATURE	Location/Qualifiers	
misc_feature	1..13	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..13	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 126		
tatgtgcaat gac		13
SEQ ID NO: 127	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 127		
tgacatcaat tattatacat cgg		23
SEQ ID NO: 128	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 128		
tgacatcaat tattatagat gga		23
SEQ ID NO: 129	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 129		
tgacatcact tattatgcat ggg		23
SEQ ID NO: 130	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 130		
tgacataaat tattctacat ggg		23
SEQ ID NO: 131	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 131		
tgaaatcaat tatcatagat cgg		23
SEQ ID NO: 132	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
source	1..23	
	mol_type = unassigned DNA	
	organism = Homo sapiens	
SEQUENCE: 132		
ccaggcteca cccgactgga ggg		23
SEQ ID NO: 133	moltype = RNA length = 106	
FEATURE	Location/Qualifiers	
misc_feature	1..106	
	note = Description of Artificial Sequence: Synthetic polynucleotide	

-continued

```

source                1..106
                      mol_type = other RNA
                      organism = synthetic construct

SEQUENCE: 133
ggccaggctc caccgcgactg gagttttaga gctagaaata gcaagttaaa ataaggctag 60
tccgttatca acttgaaaaa gtggcaccga gtcggtgctt tttttt 106

SEQ ID NO: 134        moltype = DNA length = 53
FEATURE              Location/Qualifiers
source               1..53
                      mol_type = unassigned DNA
                      organism = Homo sapiens

SEQUENCE: 134
acttccaggc tccaccggac tggagggcga accccaaggg gacctcatgc agg 53

SEQ ID NO: 135        moltype = DNA length = 13
FEATURE              Location/Qualifiers
misc_feature         1..13
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..13
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 135
acttccaggc tcc 13

SEQ ID NO: 136        moltype = DNA length = 30
FEATURE              Location/Qualifiers
misc_feature         1..30
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..30
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 136
acttccaggc tccaccggac ctcatgcagg 30

SEQ ID NO: 137        moltype = DNA length = 36
FEATURE              Location/Qualifiers
misc_feature         1..36
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..36
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 137
acttccaggc tccaccctcaa ggggacctca tgcagg 36

SEQ ID NO: 138        moltype = DNA length = 54
FEATURE              Location/Qualifiers
misc_feature         1..54
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..54
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 138
acttccaggc tccaccggac ttggagggcg aaccctcaagg ggacctcatg cagg 54

SEQ ID NO: 139        moltype = DNA length = 43
FEATURE              Location/Qualifiers
misc_feature         1..43
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..43
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 139
acttccaggc tccaccggaa cccaagggg acctcatgca ggg 43

SEQ ID NO: 140        moltype = DNA length = 47
FEATURE              Location/Qualifiers
misc_feature         1..47
                      note = Description of Artificial Sequence: Synthetic
                      oligonucleotide
source               1..47
                      mol_type = other DNA
                      organism = synthetic construct

SEQUENCE: 140

```

-continued

acttcaggc tccaccgac tcactatctt ctgggtcct ccatgct 47

SEQ ID NO: 141 moltype = DNA length = 45
 FEATURE Location/Qualifiers
 misc_feature 1..45
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..45
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 141
 acttcaggc tccaccgac gaacccaag gggacctcat gcagg 45

SEQ ID NO: 142 moltype = AA length = 18
 FEATURE Location/Qualifiers
 source 1..18
 mol_type = protein
 organism = Homo sapiens

SEQUENCE: 142
 LPGSTRLEGE PQGDLMQA 18

SEQ ID NO: 143 moltype = DNA length = 57
 FEATURE Location/Qualifiers
 source 1..57
 mol_type = unassigned DNA
 organism = Homo sapiens

CDS
 SEQUENCE: 143
 acttcaggc tccaccgac tggagggcga accccaaggg gacctcatgc aggctcc 57

SEQ ID NO: 144 moltype = DNA length = 46
 FEATURE Location/Qualifiers
 misc_feature 1..46
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..46
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 144
 acttcaggc tccaccgaa cccaagggg acctcatgca ggctcc 46

SEQ ID NO: 145 moltype = DNA length = 43
 FEATURE Location/Qualifiers
 misc_feature 1..43
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..43
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 145
 acttcaggc tccaccgaa cccaagggg acctcatgca ggc 43

SEQ ID NO: 146 moltype = DNA length = 20
 FEATURE Location/Qualifiers
 misc_feature 1..20
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..20
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 146
 acttcaggc tccaccgac 20

SEQ ID NO: 147 moltype = DNA length = 40
 FEATURE Location/Qualifiers
 misc_feature 1..40
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..40
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 147
 acttcaggc tccaccgaa ggggacctca tgcaggctcc 40

SEQ ID NO: 148 moltype = DNA length = 58
 FEATURE Location/Qualifiers
 misc_feature 1..58
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

-continued

source 1..58
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 148
 acttccaggc tccacccgac ttggagggcg aacccaagg ggacctcatg caggctcc 58

SEQ ID NO: 149 moltype = DNA length = 46
 FEATURE Location/Qualifiers
 misc_feature 1..46
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..46
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 149
 acttccaggc tccaggcgaa cccaagggg acctcatgca ggctcc 46

SEQ ID NO: 150 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 150
 ggcaacccc aaggggacct catgcaggct cc 32

SEQ ID NO: 151 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 151
 acttccaggc aaggggacct catgcaggct cc 32

SEQ ID NO: 152 moltype = DNA length = 33
 FEATURE Location/Qualifiers
 misc_feature 1..33
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..33
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 152
 acttccaggc taaggggacc tcctgcaggc tcc 33

SEQ ID NO: 153 moltype = DNA length = 52
 FEATURE Location/Qualifiers
 source 1..52
 mol_type = unassigned DNA
 organism = Homo sapiens
 SEQUENCE: 153
 acttccaggc tccacccgac ttgagggcga accccaaggg gacctcatgc ag 52

SEQ ID NO: 154 moltype = DNA length = 34
 FEATURE Location/Qualifiers
 misc_feature 1..34
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..34
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 154
 acttccaggc gaacccaag gggacctcat gcag 34

SEQ ID NO: 155 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 155
 acttccaggc tccacaaggg gacctcatgc ag 32

-continued

FEATURE	Location/Qualifiers	
misc_feature	1..46	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..46	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 163		
acttccaggc tccaccggag gcgaacccca aggggacctc atgcag		46
SEQ ID NO: 164	moltype = DNA length = 47	
FEATURE	Location/Qualifiers	
misc_feature	1..47	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..47	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 164		
acttccaggc tccaccggag gggaacccc aaggggacct catgcag		47
SEQ ID NO: 165	moltype = DNA length = 24	
FEATURE	Location/Qualifiers	
misc_feature	1..24	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..24	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 165		
acttccaggc tccacctcat gcag		24
SEQ ID NO: 166	moltype = DNA length = 29	
FEATURE	Location/Qualifiers	
misc_feature	1..29	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..29	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 166		
agggcgaacc ccaaggggac ctcatgcag		29
SEQ ID NO: 167	moltype = DNA length = 45	
FEATURE	Location/Qualifiers	
misc_feature	1..45	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..45	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 167		
caatctatga catcaattat tatcggagcc ctgccaaaaa atcaa		45
SEQ ID NO: 168	moltype = DNA length = 45	
FEATURE	Location/Qualifiers	
misc_feature	1..45	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..45	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 168		
caatctatga catcaattat catcggagcc ctgccaaaaa atcaa		45
SEQ ID NO: 169	moltype = DNA length = 42	
FEATURE	Location/Qualifiers	
misc_feature	1..42	
	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..42	
	mol_type = other DNA	
	organism = synthetic construct	
SEQUENCE: 169		
caatctatga catcaattat cggagccctg ccaaaaaatc aa		42
SEQ ID NO: 170	moltype = DNA length = 48	
FEATURE	Location/Qualifiers	
misc_feature	1..48	

-continued

note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..48
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 170
 caatctatga catcaattat tatcatcgga gccttgccaa aaaatcaa 48

SEQ ID NO: 171 moltype = DNA length = 33
 FEATURE Location/Qualifiers
 misc_feature 1..33
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..33
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 171
 caatctatga caagagccct gccaaaaaat caa 33

SEQ ID NO: 172 moltype = DNA length = 52
 FEATURE Location/Qualifiers
 source 1..52
 mol_type = unassigned DNA
 organism = Homo sapiens
 SEQUENCE: 172
 ttctcaaggc agcatcatac ttccccacg gtgggacagc tgcctcct gg 52

SEQ ID NO: 173 moltype = DNA length = 46
 FEATURE Location/Qualifiers
 misc_feature 1..46
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..46
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 173
 ttctcaaggc agcatcatac ttccctggga cagctgcct cctgg 46

SEQ ID NO: 174 moltype = DNA length = 49
 FEATURE Location/Qualifiers
 misc_feature 1..49
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..49
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 174
 ttctcaaggc agcatcatac ttccacggtg ggacagctgc cctcctgg 49

SEQ ID NO: 175 moltype = DNA length = 25
 FEATURE Location/Qualifiers
 misc_feature 1..25
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..25
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 175
 ttctcaaggc agctgcctc cctgg 25

SEQ ID NO: 176 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 176
 ttctcaaggc agcatcatac ttccctcct gg 32

SEQ ID NO: 177 moltype = DNA length = 264
 FEATURE Location/Qualifiers
 misc_difference 38..227
 note = modified_base - a, c, t, g, unknown or other
 source 1..264
 mol_type = unassigned DNA
 organism = Homo sapiens
 SEQUENCE: 177

-continued

```

acaaagcgat tttgaaagat ggaagcgcgg tggctatnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnngg gtgaaactaa 240
actggtccac acggcggaag attg 264

SEQ ID NO: 178      moltype = DNA length = 257
FEATURE           Location/Qualifiers
misc_feature      1..257
                  note = Description of Artificial Sequence: Synthetic
                  polynucleotide
misc_difference   38..227
                  note = modified_base - a, c, t, g, unknown or other
source            1..257
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 178
acaaagcgat tttgaaagat ggaagcgcgg tggctatnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnngg gtgaaactaa 240
aacacggcgg aagattg 257

SEQ ID NO: 179      moltype = DNA length = 43
FEATURE           Location/Qualifiers
misc_feature      1..43
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..43
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 179
acaaagcgat tttgaaagat ggaagcgaca cggcggaaga ttg 43

SEQ ID NO: 180      moltype = DNA length = 44
FEATURE           Location/Qualifiers
misc_feature      1..44
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..44
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 180
acaaagcgat tttgaaagat ggaagcgcac acggcggaag attg 44

SEQ ID NO: 181      moltype = DNA length = 106
FEATURE           Location/Qualifiers
misc_feature      1..106
                  note = Description of Artificial Sequence: Synthetic
                  polynucleotide
source            1..106
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 181
acaaagcgat tttgaaagat ggaagcgaaa tagcaagtta aaataaggct agtccgttat 60
caactgaaa aagtgaccac gagtcggtgc acacggcgga agattg 106

SEQ ID NO: 182      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 182
gggtgggggg agttgtctcc tgg 23

SEQ ID NO: 183      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 183
ggatggaggg agttgtctcc tgg 23

```


-continued

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 191
 gagtccgagc agaagaagaa ggg 23

SEQ ID NO: 192 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 192
 gagttagagc agaagaagaa agg 23

SEQ ID NO: 193 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 193
 gaacctgagc tgctctgacg cgg 23

SEQ ID NO: 194 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 194
 ttggcagggg gtgggagggg agg 23

SEQ ID NO: 195 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 195
 gggagggaga gcttggcagg ggg 23

SEQ ID NO: 196 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 196
 gatggagcca gagaggatcc tgg 23

SEQ ID NO: 197 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 197
 cctgccaagc tctccctccc agg 23

SEQ ID NO: 198 moltype = DNA length = 23
 FEATURE Location/Qualifiers

misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic

-continued

source	oligonucleotide 1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 198		
ctccctccca ggatcctctc tgg		23
SEQ ID NO: 199	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 199		
cctctaaggt ttgcttacga tgg		23
SEQ ID NO: 200	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 200		
ggttctggca aggagagaga tgg		23
SEQ ID NO: 201	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 201		
tctaaccccc acctcctggtt agg		23
SEQ ID NO: 202	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 202		
ttggcagggg gtgggagga tgg		23
SEQ ID NO: 203	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 203		
ttgtagggg gtgggagga tgg		23
SEQ ID NO: 204	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 204		
gggaagggga gcttggcagg tgg		23
SEQ ID NO: 205	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23	

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 205
ggtagtgaga gcttggcagg tgg 23

SEQ ID NO: 206      moltype = DNA length = 23
FEATURE            Location/Qualifiers
misc_feature        1..23
                    note = Description of Artificial Sequence: Synthetic
                    oligonucleotide
source              1..23
                    mol_type = other DNA
                    organism = synthetic construct
SEQUENCE: 206
ctccctccca ggatcctccc agg 23

SEQ ID NO: 207      moltype = DNA length = 23
FEATURE            Location/Qualifiers
misc_feature        1..23
                    note = Description of Artificial Sequence: Synthetic
                    oligonucleotide
source              1..23
                    mol_type = other DNA
                    organism = synthetic construct
SEQUENCE: 207
gtccatggca ggatcctctc agg 23

SEQ ID NO: 208      moltype = DNA length = 23
FEATURE            Location/Qualifiers
misc_feature        1..23
                    note = Description of Artificial Sequence: Synthetic
                    oligonucleotide
source              1..23
                    mol_type = other DNA
                    organism = synthetic construct
SEQUENCE: 208
ctcccccca gtatcctctc agg 23

SEQ ID NO: 209      moltype = DNA length = 108
FEATURE            Location/Qualifiers
source              1..108
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 209
ccaatctatg acatcaatta ttatacatcg gagccctgcc aaaaaatcaa tgtgaagcaa 60
atcgagccc gcctcctgcc tccgctctac tcactggtgt tcatcttt 108

SEQ ID NO: 210      moltype = DNA length = 138
FEATURE            Location/Qualifiers
source              1..138
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 210
ccaccctata atttgaacc tgcagaagaa tctgaacata aaaacaacaa ttacgaacca 60
aacctattta aaactccaca aaggaaacca tcttataatc agctggcttc aactccaata 120
atattcaag agcaaggg 138

SEQ ID NO: 211      moltype = DNA length = 49
FEATURE            Location/Qualifiers
source              1..49
                    mol_type = unassigned DNA
                    organism = Homo sapiens
SEQUENCE: 211
ggccgggaat caagagtcac ccagagacag tgaccaacca tccctgttt 49

SEQ ID NO: 212      moltype = DNA length = 77
FEATURE            Location/Qualifiers
misc_feature        1..77
                    note = Description of Artificial Sequence: Synthetic
                    oligonucleotide
source              1..77
                    mol_type = other DNA
                    organism = synthetic construct
SEQUENCE: 212
ggccgggaat caagagtcac ccagtgacca accatccctg taagcaaacc ttagaggttc 60
tggcaaggag agagatg 77

SEQ ID NO: 213      moltype = DNA length = 27
FEATURE            Location/Qualifiers

```

-continued

```

misc_feature      1..27
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..27
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 213
ggccgggaat caagagtcac ccaggaa                               27

SEQ ID NO: 214      moltype = DNA length = 79
FEATURE            Location/Qualifiers
misc_feature       1..79
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..79
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 214
ggccgggaat caagagtcac ccagacctct ctggctccat cgtaagcaaa ccttagaggt 60
tctggcaagg agagagatg                                           79

SEQ ID NO: 215      moltype = DNA length = 28
FEATURE            Location/Qualifiers
misc_feature       1..28
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..28
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 215
ggccgggaat caagagtcac cctaacag                               28

SEQ ID NO: 216      moltype = DNA length = 66
FEATURE            Location/Qualifiers
misc_feature       1..66
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..66
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 216
ggccgggaat caagagcctg gctccatcgt aagcaaacct tagaggttct ggcaaggaga 60
gagatg                                                           66

SEQ ID NO: 217      moltype = DNA length = 47
FEATURE            Location/Qualifiers
misc_feature       1..47
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..47
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 217
ggctccatcg taagcaaacc ttagaggttc tggcaaggag agagatg         47

SEQ ID NO: 218      moltype = DNA length = 78
FEATURE            Location/Qualifiers
misc_feature       1..78
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..78
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 218
ggccgggaat caagagtcac ccagactctc tggctccatc gtaagcaaac ccttagaggtt 60
ctggcaaggga gagagatg                                           78

SEQ ID NO: 219      moltype = DNA length = 80
FEATURE            Location/Qualifiers
misc_feature       1..80
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source             1..80
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 219
ggccgggaat caagagtcac ccagagacag tgaccaacca tcgtaagcaa accttagaggt 60
ttctggcaag gagagagatg                                           80

```


-continued

source 1..33
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 234
 tatgtgcaat gaccactaca tcagcaatcg gag 33

SEQ ID NO: 235 moltype = DNA length = 34
 FEATURE Location/Qualifiers
 misc_feature 1..34
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..34
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 235
 tatgtgcaat gaccactaca tccagcaatc ggag 34

SEQ ID NO: 236 moltype = DNA length = 10
 FEATURE Location/Qualifiers
 misc_feature 1..10
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..10
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 236
 cagcaatcgg 10

SEQ ID NO: 237 moltype = DNA length = 41
 FEATURE Location/Qualifiers
 misc_feature 1..41
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..41
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 237
 tatgtgcaat gaccactaca tccttcaagg gcagcaatcg g 41

SEQ ID NO: 238 moltype = DNA length = 41
 FEATURE Location/Qualifiers
 misc_feature 1..41
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..41
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 238
 tatgtgcaat gaccactaca tcctccaagg gcagcaatcg g 41

SEQ ID NO: 239 moltype = DNA length = 41
 FEATURE Location/Qualifiers
 misc_feature 1..41
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..41
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 239
 tatgtgcaat gaccactaca tcctbcaagg gcagcaatcg g 41

SEQ ID NO: 240 moltype = DNA length = 34
 FEATURE Location/Qualifiers
 misc_feature 1..34
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..34
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 240
 tatgtgcaat gaccactaca ttggcagcaa tcgg 34

SEQ ID NO: 241 moltype = DNA length = 21
 FEATURE Location/Qualifiers
 misc_feature 1..21
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide

source 1..21
 mol_type = other DNA

-continued

organism = synthetic construct
 SEQUENCE: 241
 tatgtgcaat gaccactaca t 21

SEQ ID NO: 242 moltype = DNA length = 53
 FEATURE Location/Qualifiers
 source 1..53
 mol_type = unassigned DNA
 organism = Homo sapiens

SEQUENCE: 242
 tcatacagat gatgtctcat catcagagga gcgagaaggt aaagtcaaaa tca 53

SEQ ID NO: 243 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 243
 tcatacagat gatacaggta aagtcaaaaat ca 32

SEQ ID NO: 244 moltype = DNA length = 32
 FEATURE Location/Qualifiers
 misc_feature 1..32
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..32
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 244
 tcatacaggt gatgaaggta aagtcaaaaat ca 32

SEQ ID NO: 245 moltype = DNA length = 50
 FEATURE Location/Qualifiers
 misc_feature 1..50
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..50
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 245
 tcatacagat gatgtctcat catcagagcg agaaggtaaa gtcaaaaatca 50

SEQ ID NO: 246 moltype = DNA length = 48
 FEATURE Location/Qualifiers
 misc_feature 1..48
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..48
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 246
 tcatacagat gatgtctcat catcagcgag aaggtaaagt caaatca 48

SEQ ID NO: 247 moltype = DNA length = 52
 FEATURE Location/Qualifiers
 misc_feature 1..52
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..52
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 247
 tcatacagat gatgtctcat catcagggag cgagaaggta aagtcaaaaat ca 52

SEQ ID NO: 248 moltype = DNA length = 41
 FEATURE Location/Qualifiers
 misc_feature 1..41
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..41
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 248
 tcatacagat gatgtctcgc gagaaggtaa agtcaaaaatc a 41

SEQ ID NO: 249 moltype = DNA length = 32

-continued

FEATURE Location/Qualifiers
misc_feature 1..32
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source 1..32
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 249
tcatacagat gatgaaggta aagtcaaaat ca 32

SEQ ID NO: 250 moltype = DNA length = 29
FEATURE Location/Qualifiers
misc_feature 1..29
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source 1..29
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 250
tcatacagat gaaggtaaag tcaaaatca 29

SEQ ID NO: 251 moltype = DNA length = 42
FEATURE Location/Qualifiers
misc_feature 1..42
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source 1..42
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 251
tcatacagat gatgtctaca gatgaaggta aagtcaaaat ca 42

SEQ ID NO: 252 moltype = DNA length = 51
FEATURE Location/Qualifiers
misc_feature 1..51
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source 1..51
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 252
tcatacagat gatgtctcat catcaggagc gagaaggtaa agtcaaaatc a 51

SEQ ID NO: 253 moltype = DNA length = 34
FEATURE Location/Qualifiers
misc_feature 1..34
note = Description of Artificial Sequence: Synthetic
oligonucleotide
source 1..34
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 253
gtcatcctca tcctgataaa ctgcaaaagg ctga 34

SEQ ID NO: 254 moltype = DNA length = 606
FEATURE Location/Qualifiers
source 1..606
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 254
gctggtgtct gggttctgtg ccccttcccc accccagccc accccagggtg tctgtccat 60
tctcaggctg gtcacatggg tggctctagg gtgtcccatg agagatgcaa agcgcctgaa 120
ttttctgact cttcccatca gaccccccaa agacacatgt gaccaccac cccatctctg 180
accatgaggc caccctgagg tgctgggccc tgggcttcta cctgcgagg atcacactga 240
cctggcagcg ggatggcgag gaccaaactc aggacaccga gcttgtggag accagaccag 300
caggagatag aaccttcag aagtgggcag ctgtgggtgt gccttctgga gaagagcaga 360
gatacacatg ccattgtacag catgaggggc tgccgaagcc cctcacctg agatggggta 420
aggaggggga tgaggggtca tatctgttca tatctgttct cagggaaagc aggagccctt 480
ctggagccct tcagcagggt cagggcccct catcttcccc tcctttccca gagccatctt 540
cccagtcac catcccacatc gtgggcattg ttgtggcct ggctgtccta gcagttgtgg 600
tcatcg 606

SEQ ID NO: 255 moltype = DNA length = 26
FEATURE Location/Qualifiers
source 1..26
mol_type = unassigned DNA
organism = Homo sapiens

SEQUENCE: 255
actaccacag ctcttctct gagtgg 26

-continued

```

misc_feature      1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 264
ccatacatta aagatagtc tct                               23

SEQ ID NO: 265      moltype = DNA length = 23
FEATURE           Location/Qualifiers
source           1..23
                  mol_type = unassigned DNA
                  organism = Homo sapiens

SEQUENCE: 265
agatgactat ctttaatgtc tgg                               23

SEQ ID NO: 266      moltype = DNA length = 23
FEATURE           Location/Qualifiers
source           1..23
                  mol_type = unassigned DNA
                  organism = Homo sapiens

SEQUENCE: 266
agatgactat ctttaatgta tgg                               23

SEQ ID NO: 267      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 267
gtagttggag ctgatggcgt agg                               23

SEQ ID NO: 268      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 268
gtagttggag ctggtagcgt agg                               23

SEQ ID NO: 269      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 269
gtagttggag ctggtgacgt agg                               23

SEQ ID NO: 270      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 270
gtagttggag ctaatggcgt agg                               23

SEQ ID NO: 271      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..23
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 271
gtagttggag ctagtagcgt agg                               23

```


-continued

source	oligonucleotide 1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 279		
caaatgagtg atgcacatca tgg		23
SEQ ID NO: 280	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 280		
caaaagaatg atgcacatca tgg		23
SEQ ID NO: 281	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 281		
cgaatgaatg atgcacatca tgg		23
SEQ ID NO: 282	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 282		
caaatgaatg atgcatgtca tgg		23
SEQ ID NO: 283	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 283		
caaatgaatg atgcgctca tgg		23
SEQ ID NO: 284	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 284		
caaatgaatg atgtacgtca tgg		23
SEQ ID NO: 285	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 285		
caaatgaatg gtgcacgtca tgg		23
SEQ ID NO: 286	moltype = DNA length = 23	
FEATURE	Location/Qualifiers	
misc_feature	1..23 note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..23	

-continued

mol_type = other DNA
 organism = synthetic construct
 SEQUENCE: 286
 caaatgagtg atgcacgtca tgg 23

SEQ ID NO: 287 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 287
 caaaagaatg atgcacgtca tgg 23

SEQ ID NO: 288 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 288
 cgaatgaatg atgcacgtca tgg 23

SEQ ID NO: 289 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 source 1..23
 mol_type = unassigned DNA
 organism = Homo sapiens

SEQUENCE: 289
 atcataggtc gtcacgtta tgg 23

SEQ ID NO: 290 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 290
 atcataggtt gtcacgtta tgg 23

SEQ ID NO: 291 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 source 1..23
 mol_type = unassigned DNA
 organism = Homo sapiens

SEQUENCE: 291
 atcataggtc gtcctgctta tgg 23

SEQ ID NO: 292 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..23
 mol_type = other DNA
 organism = synthetic construct

SEQUENCE: 292
 atcataggtt gtcctgctta tgg 23

SEQ ID NO: 293 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 source 1..23
 mol_type = unassigned DNA
 organism = Homo sapiens

SEQUENCE: 293
 ctggacaaga agagtacagt gcc 23

SEQ ID NO: 294 moltype = DNA length = 23
 FEATURE Location/Qualifiers
 misc_feature 1..23
 note = Description of Artificial Sequence: Synthetic
 oligonucleotide
 source 1..23

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 294
ctggaaaaga agagtacagt gcc 23

SEQ ID NO: 295      moltype = DNA length = 23
FEATURE           Location/Qualifiers
source           1..23
                 mol_type = unassigned DNA
                 organism = Homo sapiens
SEQUENCE: 295
actccatcga gatttcactg tag 23

SEQ ID NO: 296      moltype = DNA length = 23
FEATURE           Location/Qualifiers
misc_feature     1..23
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..23
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 296
actccatcga gattttctctg tag 23

SEQ ID NO: 297      moltype = DNA length = 34
FEATURE           Location/Qualifiers
misc_feature     1..34
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..34
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 297
acttccaggc aaccccaagg ggacctcatg cagg 34

SEQ ID NO: 298      moltype = DNA length = 35
FEATURE           Location/Qualifiers
misc_feature     1..35
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..35
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 298
acttccaggc gaacccaag gggacctcat gcagg 35

SEQ ID NO: 299      moltype = DNA length = 13
FEATURE           Location/Qualifiers
misc_feature     1..13
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..13
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 299
acttccaggc tcc 13

SEQ ID NO: 300      moltype = DNA length = 13
FEATURE           Location/Qualifiers
misc_feature     1..13
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..13
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 300
acttccaggc tcc 13

SEQ ID NO: 301      moltype = DNA length = 32
FEATURE           Location/Qualifiers
misc_feature     1..32
                 note = Description of Artificial Sequence: Synthetic
                 oligonucleotide
source          1..32
                 mol_type = other DNA
                 organism = synthetic construct
SEQUENCE: 301
acttccaggc tccaagggg acctcatgca gg 32

```


-continued

```

misc_feature      1..40
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source            1..40
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 309
acttccaggc tccacccacc ccaaggggac ctcatgcagg          40

SEQ ID NO: 310    moltype = DNA length = 45
FEATURE          Location/Qualifiers
misc_feature     1..45
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..45
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 310
acttccaggc tccacccggc gaacccaag gggacctcat gcagg          45

SEQ ID NO: 311    moltype = DNA length = 34
FEATURE          Location/Qualifiers
misc_feature     1..34
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..34
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 311
acttccaggc tccaccctgg ggacctcatg cagg          34

SEQ ID NO: 312    moltype = DNA length = 42
FEATURE          Location/Qualifiers
misc_feature     1..42
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..42
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 312
acttccaggc tccacccgaa cccaagggg acctcatgca gg          42

SEQ ID NO: 313    moltype = DNA length = 31
FEATURE          Location/Qualifiers
misc_feature     1..31
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..31
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 313
acttccaggc tccacccgaa cctcatgcag g          31

SEQ ID NO: 314    moltype = DNA length = 35
FEATURE          Location/Qualifiers
misc_feature     1..35
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..35
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 314
acttccaggc tccacccgag gggacctcat gcagg          35

SEQ ID NO: 315    moltype = DNA length = 36
FEATURE          Location/Qualifiers
misc_feature     1..36
                  note = Description of Artificial Sequence: Synthetic
                  oligonucleotide
source           1..36
                  mol_type = other DNA
                  organism = synthetic construct

SEQUENCE: 315
acttccaggc tccaccccaa ggggacctca tgcagg          36

SEQ ID NO: 316    moltype = DNA length = 42
FEATURE          Location/Qualifiers
misc_feature     1..42
                  note = Description of Artificial Sequence: Synthetic

```

-continued

source	oligonucleotide 1..42 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 316		
acttccaggc tccacccgaa cccaaggagg acctcatgca gg		42
SEQ ID NO: 317	moltype = DNA length = 43 Location/Qualifiers	
FEATURE	1..43	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..43 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 317		
acttccaggc tccacccgag accccaaggagg gacctcatgc agg		43
SEQ ID NO: 318	moltype = DNA length = 48 Location/Qualifiers	
FEATURE	1..48	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..48 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 318		
acttccaggc tccacccgag ggcgaacccc aaggggacct catgcagg		48
SEQ ID NO: 319	moltype = DNA length = 30 Location/Qualifiers	
FEATURE	1..30	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..30 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 319		
acttccaggc tccacccgac ctcatgcagg		30
SEQ ID NO: 320	moltype = DNA length = 42 Location/Qualifiers	
FEATURE	1..42	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..42 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 320		
acttccaggc tccacccgac cccaaggagg acctcatgca gg		42
SEQ ID NO: 321	moltype = DNA length = 42 Location/Qualifiers	
FEATURE	1..42	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..42 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 321		
acttccaggc tccacccgac gaagggccccc aaggggacct ca		42
SEQ ID NO: 322	moltype = DNA length = 45 Location/Qualifiers	
FEATURE	1..45	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..45 mol_type = other DNA organism = synthetic construct	
SEQUENCE: 322		
acttccaggc tccacccgac gaaccccaag gggacctcat gcagg		45
SEQ ID NO: 323	moltype = DNA length = 48 Location/Qualifiers	
FEATURE	1..48	
misc_feature	note = Description of Artificial Sequence: Synthetic oligonucleotide	
source	1..48	

-continued

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 323
acttccaggc tccaccggac ggcgaacccc aaggggacct catgcagg          48

SEQ ID NO: 324      moltype = DNA length = 51
FEATURE            Location/Qualifiers
misc_feature       1..51
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..51
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 324
acttccaggc tccaccggac gtgcttgagg gcgaacccca aggggacctc a      51

SEQ ID NO: 325      moltype = DNA length = 47
FEATURE            Location/Qualifiers
misc_feature       1..47
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..47
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 325
acttccaggc tccaccggac tcactatctt ctgggctcct ccatgct          47

SEQ ID NO: 326      moltype = DNA length = 49
FEATURE            Location/Qualifiers
misc_feature       1..49
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..49
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 326
acttccaggc tccaccggac ttggcgaacc ccaaggggac ctcatgcag        49

SEQ ID NO: 327      moltype = DNA length = 51
FEATURE            Location/Qualifiers
misc_feature       1..51
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..51
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 327
acttccaggc tccaccggac ttgcagggcg aacccaagg ggacctcatg c      51

SEQ ID NO: 328      moltype = DNA length = 53
FEATURE            Location/Qualifiers
misc_feature       1..53
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..53
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 328
acttccaggc tccaccggac ttggagggcg aacccaagg ggacctcatg cag    53

SEQ ID NO: 329      moltype = DNA length = 53
FEATURE            Location/Qualifiers
misc_feature       1..53
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..53
                   mol_type = other DNA
                   organism = synthetic construct
SEQUENCE: 329
acttccaggc tccaccggac tttgagggc gaacccaag gggacctcat gca     53

SEQ ID NO: 330      moltype = DNA length = 53
FEATURE            Location/Qualifiers
misc_feature       1..53
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
source             1..53
                   mol_type = other DNA
                   organism = synthetic construct

```

-continued

```

SEQUENCE: 330
acttcaggc tccaccgac tgttgaggc cgaacccaa ggggacctca tgc      53

SEQ ID NO: 331      moltype = DNA length = 502
FEATURE            Location/Qualifiers
misc_feature       1..502
                   note = Description of Artificial Sequence: Synthetic
                   oligonucleotide
misc_difference    26..480
                   note = misc_feature - a, c, t, g, unknown or other
source            1..502
                   mol_type = other DNA
                   organism = synthetic construct

SEQUENCE: 331
acttcaggc tccaccgac tggagnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 60
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 120
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 480
ggcgaacccc aaggggacct cc      502

SEQ ID NO: 332      moltype = DNA length = 72
FEATURE            Location/Qualifiers
source            1..72
                   mol_type = unassigned DNA
                   organism = Homo sapiens

SEQUENCE: 332
agctctcct cccaggatcc tctctggctc catcgtaagc aaaccttaga ggttctggca 60
aggagagaga tg      72
    
```

The invention claimed is:

1. A method of modifying a target endogenous nucleic acid sequence in a nucleus of a eukaryotic cell, comprising: providing a composition comprising an amount of a recombinant Cas9 protein and an amount of a single-guide RNA (sgRNA) to an in vitro environment, wherein the recombinant Cas9 protein comprises a nuclear localization signal (NLS); wherein the amount of the sgRNA is present in the composition in at least a two-fold molar excess over the amount of the recombinant Cas9 protein, and wherein a Cas9/sgRNA complex forms in the in vitro environment prior to the composition being introduced into the eukaryotic cell; and introducing the composition comprising the Cas9/sgRNA complex into the eukaryotic cell in vitro or ex vivo, wherein the Cas9/sgRNA complex produces a modification of the target endogenous nucleic acid sequence in the nucleus of the eukaryotic cell.
2. The method of claim 1, wherein the composition comprising the Cas9/sgRNA complex is introduced into the eukaryotic cell by transfection, wherein the transfection is performed by the method selected from the group consisting of electroporation, DEAE-dextran treatment, lipofection, nanoparticle-mediated transfection, protein transduction domain mediated transduction, and polyethylene glycol (PEG)-mediated transfection of protoplasts.
3. The method of claim 2, wherein the composition comprising the Cas9/sgRNA complex is introduced into the eukaryotic cell by electroporation.
4. The method of claim 1, wherein the target endogenous nucleic acid comprises a trinucleotide protospacer adjacent motif (PAM) recognized by the recombinant Cas9 protein, wherein the PAM consists of trinucleotide 5'-NGG-3'.
5. The method of claim 1, wherein the NLS is at the C-terminus of the recombinant Cas9 protein.

30

6. The method of claim 1, wherein the modification includes any one of a deletion, insertion, or substitution of at least one nucleotide.
7. The method of claim 1, further comprising inducing divisions of the eukaryotic cell to be divided into a plurality of cells which include the modified nucleic acid sequence.
8. The method of claim 1, wherein the sgRNA and recombinant Cas9 protein are present in a molar ratio from 29:1.4 to 29:14.
9. The method of claim 1, wherein the eukaryotic cell is a mammalian cell.
10. The method of claim 1, wherein the eukaryotic cell is a human cell.
11. The method of claim 1, wherein the recombinant Cas9 protein is derived from a *Streptococcus pyogenes* Cas9 protein.
12. The method of claim 1, wherein the recombinant Cas9 protein is a recombinant Cas9 protein expressed in *E. coli*.
13. The method claim 1, wherein the single-guide RNA comprises a CRISPR RNA (crRNA) fused to a transactivating crRNA (tracrRNA), and wherein the crRNA comprises a sequence complementary the target endogenous nucleic acid sequence.
14. The method of claim 1, wherein the sgRNA and recombinant Cas9 protein are present in a molar ratio from 2:1 to 29:4.5.
15. The method of claim 1, wherein the sgRNA and recombinant Cas9 protein are present in a weight ratio of 8:15 to 20:15.
16. A method of modifying a target endogenous nucleic acid sequence in a nucleus of a human cell, comprising: providing a composition comprising an amount of a recombinant Cas9 protein and an amount of a single-guide RNA (sgRNA) to an in vitro environment, wherein the recombinant Cas9 protein comprises a nuclear localization signal (NLS);

65

wherein the amount of the sgRNA is present in the composition in at least a two-fold molar excess over the amount of the recombinant Cas9 protein, and wherein a Cas9/sgRNA complex forms in the in vitro environment prior to the composition being introduced into the human cell; and

introducing the composition comprising the Cas9/sgRNA complex into the human cell in vitro or ex vivo via electroporation, wherein the Cas9/sgRNA complex produces a modification of the target endogenous nucleic acid sequence in the nucleus of the human cell.

17. The method of claim **16**, wherein the sgRNA and the recombinant Cas9 protein are present in a molar ratio from 2:1 to 29:4.5.

18. The method of claim **16**, wherein the sgRNA and the recombinant Cas9 protein are present in a weight ratio of 8:15 to 20:15.

* * * * *