(12) STANDARD PATENT (11) Application No. **AU 2013335451 C1**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Composition for cleaving a target DNA comprising a guide RNA specific for the target DNA and Cas protein-encoding nucleic acid or Cas protein, and use thereof

(51) International Patent Classification(s)
C12N 15/11 (2006.01) ***C12Q 1/68*** (2006.01)

(21) Application No: **2013335451** (22) Date of Filing: **2013.10.23**

(87) WIPO No: **WO14/065596**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/803,599	**2013.03.20**	**US**
61/837,481	**2013.06.20**	**US**
61/717,324	**2012.10.23**	**US**

(43) Publication Date: **2014.05.01**
(44) Accepted Journal Date: **2016.09.15**
(44) Amended Journal Date: **2024.07.04**

(71) Applicant(s)
ToolGen Incorporated

(72) Inventor(s)
Kim, Jin-Soo;Cho, Seung Woo;Kim, Sojung;Kim, Jong Min;Kim, Seokjoong

(74) Agent / Attorney
Jones Day, Level 41 88 Phillip Street, Sydney, NSW, 2000, AU

(56) Related Art
US 2010/0076057
WO 2013/176772
JINEK, M. et al., Science, 2012, vol. 337, pages 816-821

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

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
1 May 2014 (01.05.2014)

(10) **International Publication Number**
WO 2014/065596 A1

(51) **International Patent Classification**:
C12N 15/11 (2006.01) *C12Q 1/68* (2006.01)

(21) **International Application Number**:
PCT/KR2013/009488

(22) **International Filing Date**:
23 October 2013 (23.10.2013)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:

61/717,324	23 October 2012 (23.10.2012)	US
61/803,599	20 March 2013 (20.03.2013)	US
61/837,481	20 June 2013 (20.06.2013)	US

(71) **Applicant**: **TOOLGEN INCORPORATED** [KR/KR]; 1208, Byeoksan Gyeongin Digital Valley 2 Cha, 184, Gasan Digital 2-ro, Gasan-dong, Geumcheon-gu, Seoul 153-783 (KR).

(72) **Inventors**: **KIM, Jin-Soo**; National Creative Research Initiatives Center for Genome Engineering and Department of Chemistry, Seoul National University, Gwanak-gu, Seoul 151-747 (KR). **CHO, Seung Woo**; National Creative Research Initiatives Center for Genome Engineering and Department of Chemistry, Seoul National University, Gwanak-gu, Seoul 151-747 (KR). **KIM, Sojung**; National Creative Research Initiatives Center for Genome Engineering and Department of Chemistry, Seoul National University, Gwanak-gu, Seoul 151-747 (KR). **KIM, Jong Min**; 1334-15, Mora-dong, Sasang-gu, Busan 617-818 (KR). **KIM, Seokjoong**; #1301, Seocho-Paragon, Seocho 2-dong, Seoho-gu, Seoul 137-858 (KR).

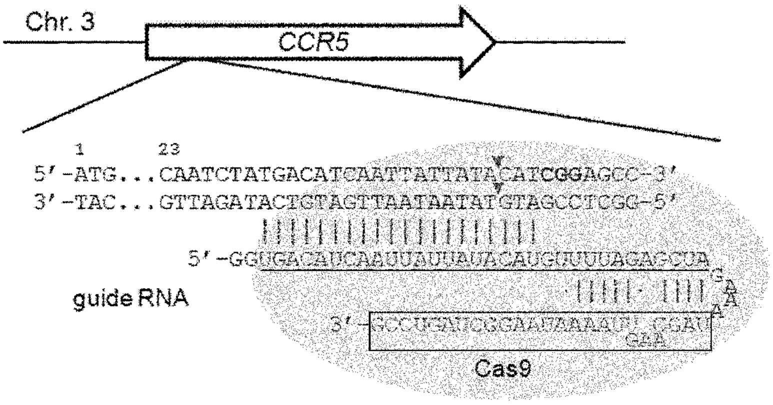
(74) **Agent**: **SON, Min**; Hanol Intellectual Property & Law, STX R&D Center, 6th Floor, 163, YangJaeCheon-ro, Gangnam-gu, Seoul 135-855 (KR).

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

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: COMPOSITION FOR CLEAVING A TARGET DNA COMPRISING A GUIDE RNA SPECIFIC FOR THE TARGET DNA AND CAS PROTEIN-ENCODING NUCLEIC ACID OR CAS PROTEIN, AND USE THEREOF



(57) **Abstract**: 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.

WO 2014/065596 A1

Description

Title of Invention: COMPOSITION FOR CLEAVING A TARGET DNA COMPRISING A GUIDE RNA SPECIFIC FOR THE TARGET DNA AND CAS PROTEIN-ENCODING NUCLEIC ACID OR CAS PROTEIN, AND USE THEREOF

Technical Field

[1] 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.

[2]

Background Art

[3] 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.

[4] 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.

[5] CRISPR/Cas systems offer an advantage to zinc finger and transcription activator-like effector DNA-binding proteins, 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.

[6] However, until now, a genome editing method using the RNA-guided en-

donuclease(RGEN) based on CRISPR/Cas system has not been developed.

[7]

[8] 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.

[9] 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, theses 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.

[10]

Disclosure of Invention

Technical Problem

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

[12] 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.

[13] 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.

[14]

Solution to Problem

[15] 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.

It an object of the present invention to provide a composition comprising a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system for use in introducing a site-specific, double stranded break at a target nucleic acid sequence in a eukaryotic cell, said CRISPR/Cas system comprising (i) a nucleic acid encoding a Cas9 polypeptide comprising a nuclear localization sequence, and (ii) a nucleic acid encoding a guide RNA that hybridizes to a target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion.

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 an objection of the present invention to provide a method of introducing a site-specific, double-stranded break at a target nucleic acid sequence in a eukaryotic cell, the method comprising introducing into the eukaryotic cell a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system, wherein the CRISPR/Cas system comprises:

a) a nucleic acid encoding a Cas9 polypeptide comprising a nuclear localization signal, wherein the nucleic acid is codon-optimized for expression in eukaryotic cells, and

b) a nucleic acid encoding a guide RNA that hybridizes to the target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion, wherein the target nucleic acid sequence comprises a first strand that binds to the crRNA portion and a second strand having a trinucleotide protospacer adjacent motif (PAM),

and wherein the Cas9 polypeptide and the guide RNA form a Cas9/RNA complex in the eukaryotic cell, whereby a site-specific, double stranded break at the target nucleic acid sequence is introduced.

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.

[31] 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.

[32]

[33] 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.

[34]

[35] 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.

[36]

[37] It is still another object of the present invention to provide a

[38] 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.

[39]

[40] 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.

[41]

[42] 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.

[43]

[44] 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.

[45]

[46] 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.

[47]

[48] 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.

[49]

[50] 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.

[51]

[52] 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.

[53]

[54] 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.

[55]

[56] 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.

[57]

[58] 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.

[59]

[60] 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.

[61]

[62] 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.

[63]

[64] 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.

[65]

[66] 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 compostion, specifically comprising a RNA-guided endonuclease (RGEN), wherein the RGEN comprises a guide RNA specific for target DNA and Cas protein.

[67]

[68] 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

[69] 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 compostion and method of the present invention may be used in detection and cleaving naturally-occurring variations and mutations.

[70]

Brief Description of Drawings

[71] Fig. 1 shows Cas9-catalyzed cleavage of plasmid DNA *in vitro*. (a) Schematic representation of target DNA and chimeric RNA sequences. Red triangles indicate cleavage

sites. The PAM sequence recognized by Cas9 is shown in bold. The sequences in the guide RNA derived from crRNA and tracrRNA are shown in box and underlined, respectively. (b) *In vitro* cleavage of plasmid DNA by Cas9. An intact circular plasmid or ApaLI-digested plasmid was incubated with Cas9 and guide RNA.

[72]

[73] Fig. 2 shows Cas9-induced mutagenesis at an episomal target site. (a) 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. (b) Flow cytometry of cells transfected with Cas9. The percentage of cells that express the RFP-GFP fusion protein is indicated.

[74]

[75] Fig. 3 shows RGEN-driven mutations at endogenous chromosomal sites. (a) CCR5 locus. (b) 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 CCR5 and C4BPB wild-type (WT) and mutant clones. The region of the target sequence complementary to the guide RNA is shown in boc. 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.

[76]

[77] Fig. 4 shows that RGEN-driven off-target mutations are undetectable. (a) On-target and potential off-target sequences. The human genome was searched in silico for potential off-target sites. Four sites were identified, each of which carries 3-base mismatches with the CCR5 on-target site. Mismatched bases are underlined. (b) 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. (c) 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.

[78]

[79] Fig. 5 shows RGEN-induced Foxn1 gene targeting in mice. (a) A schematic diagram depicting a sgRNA specific to exon 2 of the mouse Foxn1 gene. PAM in exon 2 is shown in red and the sequence in the sgRNA that is complementary to exon 2 is underlined. Triangles indicate cleavage sites. (b) Representative T7E1 assays demon-

strating 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. (c) DNA sequences of mutant alleles observed in three Foxn1 mutant founders identified in b. The number of occurrences is shown in parentheses. (d) 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.

[80]

[81] Fig. 6 shows Foxn1 gene targeting in mouse embryos by intra-cytoplasmic injection of Cas9 mRNA and Foxn1-sgRNA. (a) A representative result of a T7E1 assay monitoring the mutation rate after injecting the highest dose. Arrows indicate bands cleaved by T7E1. (b) 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. (c) DNA sequences of Foxn1 mutant alleles identified from a subset of T7E1-positive mutant embryos. The target sequence of the wild-type allele is denoted in box.

[82]

[83] Fig. 7 shows Foxn1 gene targeting in mouse embryos using the recombinant Cas9 protein: Foxn1-sgRNA complex. (a) and (b) are representative T7E1 assays results and their summaries. Embryos were cultivated *in vitro* after they underwent pronuclear (a) or intra-cytoplasmic injection (b). Numbers in red indicate T7E1-positive mutant founder mice. (c) DNA sequences of Foxn1 mutant alleles 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.

[84]

[85] Fig. 8 shows Germ-line transmission of the mutant alleles found in Foxn1 mutant founder #12. (a) fPCR analysis. (b) PCR genotyping of wild-type FVB/NTac, the founder mouse, and their F1 progenies.

[86]

[87] Fig. 9 shows Genotypes of embryos generated by crossing Prkdc mutant founders. Prkdc mutant founders ♂25 and ♀15 were crossed and E13.5 embryos were isolated. (a) fPCR analysis of wild-type, founder ♂25, and founder ♀15. Note that, due to the technical limitations of fPCR analysis, these results showed small differences from the precise sequences of the mutant alleles; e.g., from the sequence analysis, $\Delta 269/\Delta 61/WT$ and $\Delta 5+1/+7/+12/WT$ were identified in founders ♂25 and ♀15, respectively. (b) Genotypes of the generated embryos.

[88]

[89] Fig. 10 shows Cas9 protein/sgRNA complex induced targeted mutation.

[90]

[91] Fig. 11 shows recombinant Cas9 protein-induced mutations in Arabidopsis protoplasts.

[92]

[93] Fig. 12 shows recombinant Cas9 protein-induced mutant sequences in the Arabidopsis BRI1 gene.

[94]

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

[96]

[97] Fig. 14 (a, b) shows 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 x 10^6 cells), or (D) co-transfected with 1 μg of Cas9-encoding plasmid and 1 μg of GX_{19} sgRNA expression plasmid (2 x 10^5 cells).

[98]

[99] Fig. 15 (a, b) shows 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).

[100]

[101] Fig. 16 shows that *in vitro* DNA cleavage by Cas9 nickases. (a) Schematic overview of the Cas9 nuclease and the paired Cas9 nickase. The PAM sequences and cleavage sites are shown in box. (b) Target sites in the human AAVS1 locus. The position of each target site is shown in triangle. (c) Schematic overview of DNA cleavage reactions. FAM dyes (shown in box) were linked to both 5’ ends of the DNA substrate. (d) DSBs and SSBs analyzed using fluorescent capillary eletrophoresis. Fluorescently-labeled DNA substrates were incubated with Cas9 nucleases or nickases before electrophoresis.

[102]

[103] Fig. 17 shows comparison of Cas9 nuclease and nickase behavior. (a) On-target mutation frequencies associated with Cas9 nucleases (WT), nickases (D10A), and paired nickases. Paired nickases that would produce 5' overhangs or 3' overhangs are indicated. (b) Analysis of off-target effects of Cas9 nucleases and paired nickases. A

total of seven potential off-target sites for three sgRNAs were analyzed.

[104]

[105] Fig. 18 shows paired Cas9 nickases tested at other endogenous human loci. (a,c) The sgRNA target sites at human CCR5 and BRCA2 loci. PAM sequences are indicated in red. (b,d) Genome editing activities at each target site 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.

[106]

[107] Fig. 19 shows that paired Cas9 nickases mediate homologous recombination. (a) 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. (b) 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.

[108]

[109] Fig. 20 shows DNA splicing induced by paired Cas9 nickases. (a) 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. (b) Genomic deletions detected using PCR. Asterisks indicate deletion-specific PCR products. (c) DNA sequences of deletion-specific PCR products obtained using AS2 and 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. (d) A schematic model of paired Cas9 nickase-mediated chromosomal deletions. Newly-synthesized DNA strands are shown in box.

[110]

[111] Fig. 21 shows that paired Cas9 nickases do not induce translocations. (a) Schematic overview of chromosomal translocations between the on-target and off-target sites. (b) PCR amplification to detect chromosomal translocations. (c) Translocations induced by Cas9 nucleases but not by the nickase pair.

[112]

[113] Fig. 22 shows a conceptual diagram of the T7E1 and RFLP assays. (a) 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. (b) Expected results of T7E1 and RGEN digestion resolved by electrophoresis.

[114]

[115] 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). 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.

[116]

[117] Fig. 24 shows genotyping of mutations induced by engineered nucleases in cells via RGEN-mediated RFLP. (a) Genotype of C4BPB mutant K562 cell clones. (b) 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.

[118]

[119] Fig. 25 shows genotyping of RGEN-induced mutations via the RGEN-RFLP technique. (a) Analysis of C4BPB-disrupted clones using RGEN-RFLP and T7E1 assays. Arrows indicate expected positions of DNA bands cleaved by RGEN or T7E1. (b) 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. (c) Genotyping of RGEN-induced mutations in the HLA-B gene in HeLa cells with RFLP and T7E1 analyses.

[120]

[121] Fig. 26 shows genotyping of mutations induced by engineered nucleases in organisms via RGEN-mediated RFLP. (a) Genotype of Pibf1 mutant founder mice. (b) 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.

[122]

[123] Fig. 27 shows RGEN-mediated genotyping of ZFN-induced mutations. The ZFN target site is shown in box. Black arrows indicate DNA bands cleaved by T7E1.

[124]

[125] Fig. 28 shows polymorphic sites in a region of the human HLA-B gene. 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.

[126]

[127] Fig. 29 shows genotyping of oncogenic mutations via RGEN-RFLP analysis. (a) A recurrent mutation (c.133-135 deletion of TCT) in the human CTNNB1 gene in HCT116 cells was detected by RGENs. HeLa cells were used as a negative control. (b) Genotyping of the KRAS substitution mutation (c.34 G>A) in the A549 cancer cell line with RGENs that contain mismatched guide RNA. 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.

[128]

[129] Fig. 30 shows genotyping of the CCR5 delta32 allele in HEK293T cells via RGEN-RFLP analysis. (a) RGEN-RFLP assays of cell lines. K562, SKBR3, and HeLa cells were used as wild-type controls. Arrows indicate DNA bands cleaved by RGENs. (b) DNA sequence of wild-type and delta32 CCR5 alleles. 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. (c) *In vitro* cleavage of plasmids harboring WT or del32 CCR5 alleles using the wild-type-specific RGEN. (d) 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 or off-target sequences using various amounts of the del32-specific RGEN.

[130]

[131] Fig. 31 shows genotyping of a KRAS point mutation (c.34 G>A). (a) 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 or the mutant sequence. KRAS genotypes in these cells were confirmed by Sanger sequencing. (b) Plasmids harboring either the wild-type or mutant KRAS sequences were digested using RGENs with perfectly matched crRNAs or attenuated, one-base mismatched crRNAs. Attenuated crRNAs that were chosen for genotyping are labeled in box above the gels.

[132]

[133] Fig. 32 shows genotyping of a PIK3CA point mutation (c.3140 A>G). (a) 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 or the mutant sequence. PIK3CA genotypes in these cells were confirmed by Sanger sequencing. (b) Plasmids harboring either the wild-type or mutant PIK3CA sequences were digested using RGENs with perfectly matched crRNAs or attenuated, one-base mismatched crRNAs. Attenuated crRNAs that were chosen for genotyping are labeled in box above the gels.

[134]

[135] Fig. 33 shows genotyping of recurrent point mutations in cancer cell lines. (a) RGEN-RFLP assays of recurrent oncogenic point mutations in IDH (c.394c>T), (b) PIK3CA (c.3140A>G), (c) NRAS (c.181C>A), (d) and BRAF genes (c.1799T>A). 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.

[136]

Best Mode for Carrying out the Invention

[137] 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.

[138]

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

[140]

[141] 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.

[142]

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

[144]

[145] 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.

[146]

[147] 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 tools for genome engineering in eukaryotic cells and organisms. In addtion, 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 recognize 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.

[148]

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

[150]

[151] 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 trans-activating crRNA (tracrRNA).

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

[153] 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, Apern, and Mtube).

[154] 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.

[155]

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

[157]

[158] 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.

[159] Preferably, Cas protein is Cas9 protein or variants thereof.

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

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

[162] The Cas protein derived from Streptococcus pyogens may recognizes NGG trinucleotide. The Cas protein may comprise an amino acid sequence of SEQ ID NO: 109, but it is not limited thereto.

[163]

[164] 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.

[165]

[166] 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.

[167]

[168] 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.

[169] 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.

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

[171] 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.

[172] The crRNA may hybridize with a target DNA.

[173] 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.

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

[175] 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.

[176] 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.

[177] 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.

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

[179] 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.

[180]

[181] 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.

[182]

[183] As used herein, the term "cleavage" refers to the breakage of the covalent backbone of a nucleotide molecule.

[184]

[185] 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.

[186]

[187] 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.

[188]

[189] 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.

[190]

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

[192]

[193] 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.

[194]

[195] 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.

[196]

[197] 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.

[198] 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.

[199] Using Cas protein rather than a nucleic acid encoding Cas protein to induce a targeted mutagenesis is advantageous because exogeneous 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.

[200]

[201] 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 addtion, the present invenion 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.

[202]

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

[204]

[205] 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.

[206]

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

[208]

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

[210]

[211] 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.

[212]

[213] 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.

[214]

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

[216]

[217] 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.

[218]

[219] Preferably, the eukarotic cell or organisms is co-transfected or serial-transfected with a Cas9 protein and a guide RNA.

[220] 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.

[221]

[222] 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.

[223]

[224] 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.

[225]

[226] 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.

[227]

[228] 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.

[229]

[230] 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.

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

[232]

[233] 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.

[234]

[235] 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 Choirinic 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 oviduts.

[236]

[237] 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.

[238]

[239] 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 offsprings.

[240]

[241] 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.

[242]

[243] 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.

[244]

[245] 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.

[246]

[247] 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.

[248]

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

[250]

[251] 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.

[252]

[253] 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.

[254]

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

[256]

[257] 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 addrion,

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.

[258]

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

[260]

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

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

[263]

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

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

[266]

[267] As used herein the term "biological sample" includes samples for analysis, such as tissues, cells, whole blood, semm, plasma, saliva, sputum, cerbrospinal fluid or urine, but is not limited thereto

[268]

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

[270]

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

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

[273]

[274] 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 bialleic mutant clones from wild-type

cells nor heterozygous biallelic mutants from heterozygous monoalleic 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.

[275]

[276] 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 addtion, 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.

[277]

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

[279]

[280] 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 addtion, 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.

[281]

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

[283]

Mode for the Invention

[284] 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.

[285]

[286] **Example 1: Genome editing assay**

[287]

[288] **1-1. DNA cleavage activity of Cas9 protein**

[289]

[290] 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.

[291] 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).

[292] 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-GGSGPPKKKRKVYPYDVPDYA-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).

[293] 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 ℃. 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 supercoiled 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 ℃. Digested DNA was analyzed by electrophoresis using 0.8% agarose gels.

[294] 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).

[295]

[296] **1-2. DNA cleavage by Cas9/guide RNA complex in human cells**

[297]

[298] 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.

[299] 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).

[300] 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.

[301] 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).

[302] Meanwhile, the *in vitro* transcribed chimeric RNA had been prepared as follows. RNA was *in vitro* transcribed through run-off reactions using the MEGAshortscript 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.

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

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

[305] 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-experssing cells were obtained by serial-transfection of the Cas9 plasmid and the guide RNA rather than co-transfection.

[306]

[307] Table 1

[Table 1]

Gene		sequence (5' to 3')	SEQ ID NO.
Oligonucleotides used for the construction of the reporter plasmid			
CCR5	F	AATTCATGACATCAATTATTATACATCGGAGGAG	3
	R	GATCCTCCTCCGATGTATAATAATTGATGTCATG	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	GCTGTTCATTGCAATGGAATG	11
Primers used for the amplification of off-target sites			
ADCY5	F1	GCTCCCACCTTAGTGCTCTG	12
	R1	GGTGGCAGGAACCTGTATGT	13
	F2	GTCATTGGCCAGAGATGTGGA	14
	R2	GTCCCATGACAGGCGTGTAT	15
KCNJ6	F	GCCTGGCCAAGTTTCAGTTA	16
	R1	TGGAGCCATTGGTTTGCATC	17
	R2	CCAGAACTAAGCCGTTTCTGAC	18
CNTNAP2	F1	ATCACCGACAACCAGTTTCC	19
	F2	TGCAGTGCAGACTCTTTCCA	20
	R	AAGGACACAGGGCAACTGAA	21
N/A Chr. 5	F1	TGTGGAACGAGTGGTGACAG	22
	R1	GCTGGATTAGGAGGCAGGATTC	23

	F2	GTGCTGAGAACGCTTCATAGAG	24
	R2	GGACCAAACCACATTCTTCTCAC	25
Primers used for the detection of chromosomal deletions			
Deletion	F	CCACATCTCGTTCTCGGTTT	26
	R	TCACAAGCCCACAGATATTT	27

[308]

[309] **1-3. Targeted disruption of endogeneous genes in mammalian cells by RGEN**

[310]

[311] 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).

[312]

[313] To introduce DSBs in mammalian cells using RGENs, $2x10^6$ K562 cells were transfected with 20 μg 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 μg/ml, respectively).

[314] After 24h, 10-40 μg of *in vitro* transcribed chimeric RNA was nucleofected into $1x10^6$ K562 cells. The *in vitro* transcribed chimeric RNA had been prepared as described in the Example 1-2.

[315]

[316] 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.

[317] 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).

[318]

[319] 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.

[320]

[321] 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 ZFN 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.

[322] 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 ZFN and RGEN specific to CCR5. Whereas the ZFN induced deletions, the RGEN did not (Fig. 4C).

[323]

[324] 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.

[325]

[326] **Example 2: Proteinaceous RGEN-mediated genome editing**

[327]

[328] 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 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.

[329] 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. Purifed 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: $1x10^6$ K562 cells were transfected with 22.5-225 (1.4-14 μM) of Cas9 protein mixed with 100ug (29 μM) of *in vitro* transcribed sgRNA (or crRNA 40ug and tracrRNA 80ug) in 100μl 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 48hr. When $2x10^5$ K562 cells were transfected with 1/5 scale-downed protocol, 4.5-45 μg of Cas9 protein mixed with 6-60ug of *in vitro* transcribed sgRNA (or crRNA 8 μg and tracrRNA 16 μg) were used and nucleofected in 20μl solution. Nucleofected cell were then placed in growth media in 48-well plates. After 48hr, 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.

[330] As shown in Fig. 10, Cas9 protien/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 2x10^{5} cells were transfected with 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%).

[331] 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.

[332]

[333]

[334] Table 2

[Table 2]

Sequences of guide RNA

Target	RNA type	RNA sequence (5' to 3')	Length	SEQ ID NO
CCR5	sgRNA	GGUGACAUCAAUUAUUAUACAUGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	104bp	28
	crRNA	GGUGACAUCAAUUAUUAUACAUGUUUUAGAGCUAUGCUGUUUUG	44bp	29
	tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU	86bp	30

[335]

[336] **Example 3: RNA-guided genome editing in mice**

[337]

[338] 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.

[339] To evaluate the genome-editing activity of the Foxn1-RGEN, we injected Cas9 mRNA (10-ng/μl 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).

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

[341]

[342] Specifically, Cas9 mRNA and sgRNAs were synthesized *in vitro* from linear DNA templates using the mMESSAGE 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.

[343]

[344] Table 3

[Table 3]

RNA Name	Direction	Sequence (5' to 3')	SEQ ID NO
Foxn1 #1 sgRNA	F	GAAATTAATACGACTCACTATAGG*CAGTCTGACGTCACACTTCC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	31
Foxn1 #2 sgRNA	F	GAAATTAATACGACTCACTATAGG*ACTTCCAGGCTCCACCCGAC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	32
Foxn1 #3 sgRNA	F	GAAATTAATACGACTCACTATAGG*CCAGGCTCCACCCGACTGGA*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	33
Foxn1 #4 sgRNA	F	GAAATTAATACGACTCACTATAGG*ACTGGAGGGCGAACCCCAAG*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	34
Foxn1 #5 sgRNA	F	GAAATTAATACGACTCACTATAGG*ACCCCAAGGGGACCTCATGC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	35
Prkdc #1 sgRNA	F	GAAATTAATACGACTCACTATAGG*TTAGTTTTTTCCAGAGACTT*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	36
Prkdc #2 sgRNA	F	GAAATTAATACGACTCACTATAGG*TTGGTTTGCTTGTGTTTATC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	37
Prkdc #3 sgRNA	F	GAAATTAATACGACTCACTATAGG*CACAAGCAAACCAAAGTCTC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	38

Prkdc #4 sgRNA	F	GAAATTAATACGACTCACTATAGG*CCTCAATGCTAAGCGACTTC*GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	39

[345]

[346] 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.

[347] 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).

[348] 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 FemtoJet microinjector (Eppendorf).

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

[350]

[351] To screen F0 mice and *in vitro* cultivated mouse embryos with RGEN-induced mutations, T7E1 assays were performed as previously described using genomic DNA samples from tail biopsies and lysates of whole embryos (Cho et al., 2013).

[352] 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.

[353]

[354] Table 4

[Table 4]

Primers used in the T7E1 assay

Gene	Direction	Sequence(5' to 3')	SEQ ID NO
Foxn1	F1	GTCTGTCTATCATCTCTTCCCTTCTCTCC	40
	F2	TCCCTAATCCGATGGCTAGCTCCAG	41
	R1	ACGAGCAGCTGAAGTTAGCATGC	42
	R2	CTACTCAATGCTCTTAGAGCTACCAGGCTTGC	43
Prkdc	F	GACTGTTGTGGGGAGGGCCG	44
	F2	GGGAGGGCCGAAAGTCTTATTTTG	45
	R1	CCTGAAGACTGAAGTTGGCAGAAGTGAG	46
	R2	CTTTAGGGCTTCTTCTCTACAATCACG	47

[355]

[356] Table 5

[Table 5]

Primers used for amplification of off-target sites

Gene	Notation	Direction	Sequence(5' to 3')	SEQ ID NO
Foxn1	off 1	F	CTCGGTGTGTAGCCCTGAC	48
		R	AGACTGGCCTGGAACTCACAG	49
	off 2	F	CACTAAAGCCTGTCAGGAAGCCG	50
		R	CTGTGGAGAGCACACAGCAGC	51
	off 3	F	GCTGCGACCTGAGACCATG	52
		R	CTTCAATGGCTTCCTGCTTAGGCTAC	53
	off 4	F	GGTTCAGATGAGGCCATCCTTTC	54
		R	CCTGATCTGCAGGCTTAACCCTTG	55

Prkdc	off 1	F	CTCACCTGCACATCACATGTGG	56
		R	GGCATCCACCCTATGGGGTC	57
	off 2	F	GCCTTGACCTAGAGCTTAAAGAGCC	58
		R	GGTCTTGTTAGCAGGAAGGACACTG	59
	off 3	F	AAAACTCTGCTTGATGGGATATGTGGG	60
		R	CTCTCACTGGTTATCTGTGCTCCTTC	61
	off 4	F	GGATCAATAGGTGGTGGGGGATG	62
		R	GTGAATGACACAATGTGACAGCTTCAG	63
	off 5	F	CACAAGACAGACCTCTCAACATTCAGTC	64
		R	GTGCATGCATATAATCCATTCTGATTGCTCTC	65
	off 6	F1	GGGAGGCAGAGGCAGGT	66
		F2	GGATCTCTGTGAGTTTGAGGCCA	67
		R1	GCTCCAGAACTCACTCTTAGGCTC	68

[357]

[358] 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:
5'-CTACTCCCTCCGCAGTCTGA-3' (SEQ ID NO: 69) and
5'-CCAGGCCTAGGTTCCAGGTA-3' (SEQ ID NO: 70) for the Foxn1 gene,
5'-CCCCAGCATTGCAGATTTCC-3' (SEQ ID NO: 71) and
5'-AGGGCTTCTTCTCTACAATCACG-3' (SEQ ID NO: 72) for Prkdc gene.

[359] 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).

[360] 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.

[361]

[362] 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 pseudo-pregnant foster mothers.

[363] 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).

[364]

[365] Table 6

[Table 6]

RGEN-mediated gene-targeting in FVB/NTac mice

Target Gene	Cas9 mRNA + sgRNA (ng/µl)	Injected embryos	Transferred 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)
	Total	280	242 (86)	159 (66)	147 (61)	99 (67)
Prkdc	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)

[366]

[367] 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 93% (100 ng/μl Foxn1-sgRNA) (Tables 6 and 7, Fig. 5b).

[368]

[369] Table 7

[Table 7]

DNA sequences of Foxn1 mutant alleles identified from a subset of T7E1-positive mutant founders

ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGACCTCATGCAGG	del+ins	#	Founder mice
ACTTCCAGGC-------------------AACCCCAAGGGGACCTCATGCAGG	Δ19	1	20
ACTTCCAGGC------------------GAACCCCAAGGGGACCTCATGCAGG	Δ18	1	115
ACTTCCAGGCTCC--	Δ60	1	19
ACTTCCAGGCTCC--	Δ44	1	108
ACTTCCAGGCTCC---------------------CAAGGGGACCTCATGCAGG	Δ21	1	64
ACTTCCAGGCTCC------------TTAGGAGGCGAACCCCAAGGGGACCTCA	Δ12+6	1	126
ACTTCCAGGCTCCACC----------------------------TCATGCAGG	Δ28	1	5
ACTTCCAGGCTCCACCC---------------------CCAAGGGACCTCATG	Δ21+4	1	61
ACTTCCAGGCTCCACCC------------------AAGGGGACCTCATGCAGG	Δ18	2	95, 29
ACTTCCAGGCTCCACCC-----------------CAAGGGGACCTCATGCAGG	Δ17	7	12, 14, 27, 66, 108, 114, 126
ACTTCCAGGCTCCACCC---------------ACCCAAGGGGACCTCATGCAG	Δ15+1	1	32
ACTTCCAGGCTCCACCC---------------CACCCAAGGGGACCTCATGCA	Δ15+2	1	124
ACTTCCAGGCTCCACCC-------------ACCCCAAGGGGACCTCATGCAGG	Δ13	1	32
ACTTCCAGGCTCCACCC--------GGCGAACCCCAAGGGGACCTCATGCAGG	Δ8	1	110
ACTTCCAGGCTCCACCCT-------------------GGGGACCTCATGCAGG	Δ20+1	1	29

ACTTCCAGGCTCCACCCG-----------AACCCCAAGGGGACCTCATGCAGG	Δ11	1	111
ACTTCCAGGCTCCACCCGA----------------------ACCTCATGCAGG	Δ22	1	79
ACTTCCAGGCTCCACCCGA------------------GGGGACCTCATGCAGG	Δ18	2	13, 127
ACTTCCAGGCTCCACCCCA-----------------AGGGGACCTCATGCAGG	Δ17	1	24
ACTTCCAGGCTCCACCCGA-----------ACCCCAAGGGGACCTCATGCAGG	Δ11	5	14, 53, 58, 69, 124
ACTTCCAGGCTCCACCCGA----------GACCCCAAGGGGACCTCATGCAGG	Δ10	1	14
ACTTCCAGGCTCCACCCGA-----GGGCGAACCCCAAGGGGACCTCATGCAGG	Δ5	3	53, 79, 115
ACTTCCAGGCTCCACCCGAC-----------------------CTCATGCAGG	Δ23	1	108
ACTTCCAGGCTCCACCCGAC-----------CCCCAAGGGGACCTCATGCAGG	Δ11	1	3
ACTTCCAGGCTCCACCCGAC-----------GAAGGGCCCCAAGGGGACCTCA	Δ11+6	1	66
ACTTCCAGGCTCCACCCGAC--------GAACCCCAAGGGGACCTCATGCAGG	Δ8	2	3, 66
ACTTCCAGGCTCCACCCGAC-----GGCGAACCCCAAGGGGACCTCATGCAGG	Δ5	1	27
ACTTCCAGGCTCCACCCGAC--GTGCTTGAGGGCGAACCCCAAGGGGACCTCA	Δ2+6	2	5
ACTTCCAGGCTCCACCCGACT------CACTATCTTCTGGGCTCCTCCATGTC	Δ6+25	2	21, 114
ACTTCCAGGCTCCACCCGACT----TGGCGAACCCCAAGGGGACCTCATGCAG	Δ4+1	1	53
ACTTCCAGGCTCCACCCGACT--TGCAGGGCGAACCCCAAGGGGACCTCATGC	Δ2+3	1	126
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCAG	+1	15	3, 5, 12, 19, 29, 55, 56,

			61, 66, 68, 81, 108, 111, 124, 127
ACTTCCAGGCTCCACCCGACTTTGGAGGGCGAACCCCAAGGGGACCTCATGCA	+2	2	79, 120
ACTTCCAGGCTCCACCCGACTGTTGGAGGGCGAACCCCAAGGGGACCTCATGC	+3	1	55
ACTTCCAGGCTCCACCCGACTGGAG(+455)GGCGAACCCCAAGGGGACCTCC	+455	1	13

[370]

[371] 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.

[372] 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.

[373]

[374] Table 8

[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
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

[375] Underlined alleles were sequenced.

[376] Alleles in red, detected by sequencing, but not by fPCR.

[377] *only one clone sequenced.

[378] **Not determined by fPCR.

[379]

[380] 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.

[381]

[382] Example 4: RNA-guided genome editing in plants

[383]

[384] 4-1. Production of Cas9 protein

[385]

[386] 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 0.2mM 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).

[387]

[388] 4-2. Production of guide RNA

[389]

[390] 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 vitor 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}x20, 72°C 5 min.

[391]

[392] Table 9

[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)	GAAATTAATACGACTCACTATAGGTTTGAAAGATGGAAGCGCGGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	73
BRI1 target 2 (Forward)	GAAATTAATACGACTCACTATAGGTGAAACTAAACTGGTCCACAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	74
Universal (Reverse)	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGC	75

[393]

[394] 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 ul of purified Cas9 protein (12 μg/μl) and 4 ul 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.

[395]

[396] **4-3. Transfection of Cas9/sgRNA complex to protoplast**

[397]

[398] 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, 20mM 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 X g for 3~5 min. Protoplasts were re-suspended in CPW solution after counting cells under the microscope (X100) using a hemacytometer. Finally, protoplasts were re-suspended at 1X10^6 /ml in MMG solution (4mM 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 uL

of Cas9/sgRNA complex [Cas9 protein (6 μg/μL) and two sgRNAs (2.2 μg/μL each)] and 200 ul 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 X g, washed with 1 ml of W5 solution, centrifuged for another 5 min at 100 X g. The density of protoplasts was adjusted to 1 X 10^5 /ml and they were cultured in modified KM 8p liquid medium with 400 mM glucose.

[399]

[400] **4-4. Detection of mutations in Arabidopsis protoplasts and plants**

[401]

[402] 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 Figure 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(Figure 12). The other three mutant sequences had deletions of ~220-bp DNA segments between the two RGEN site.

[403]

[404] **Example 5: Cas9 protein transduction using a cell-penutrating peptide or protein transduction domain**

[405]

[406] **5-1. Construction of His-Cas9-encoding plasmid**

[407]

[408] 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, Merk Millipore, Germany) containing His-tag at the N-terminus.

[409]

[410] **5-2. Cell culture**

[411]

[412] 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.

[413]

[414] **5-3. Expression and purification of Cas9 protein**

[415]

[416] To express the Cas9 protein, E. coli BL21 cells were transformed with thepET28-(a) vector encoding Cas9and 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 0.5mM.

[417] The cells were collected by centrifugation at 4000 rpm for 15-20 mins, resuspendedin a lysis buffer (20mM Tris-Cl pH8.0, 300mM NaCl, 20mM imidazole, 1X 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 (20mM Tris-Cl pH8.0, 300mM NaCl, 20mM imidazole, 1X protease inhibitor cocktail) and the bound protein was eluted at the flow rate of 0.5 ml/min with an elution buffer (20mM Tris-Cl pH8.0, 300mM NaCl, 250mM imidazole, 1X 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.

[418]

[419] **5-4. Conjugation of Cas9 to 9R4L**

[420]

[421] 1mg Cas9 protein diluted in PBS at the concentration of 1mg/mL and 50 µg of maleimide-9R4L peptide in 25µL DW (Peptron, Korea) weregently mixed using a rotor at room temperature for 2 hrsand at 4°C overnight. To remove unconjugated maleimide-9R4L, the samples were dialyzed using 50kDamolecular 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.

[422]

[423] **5-5. Preparation of sgRNA-9R4L**

[424]

[425] 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).

[426]

[427] **5-6. Cas9 protein and sgRNA treatments**

[428]

[429] Cas9-9R4L and sgRNA-C9R4LC were treated to the cells as follows: 1μg 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.

[430]

[431] **5-7. Cas9-9R4L and sgRNA-9R4L can edit endogenous genes in cultured mammalian cells without the use of additional delivery tools**

[432]

[433] 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.

[434]

[435] **Example 6: Control of off-target mutation according to guide RNA structure**

[436]

[437] 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'-GGX_{20}, whereas 5'-GX_{19}, in which X_{20} or GX_{19} 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.

[438] Next, we tested whether the addition of two guanine nucleotides at the 5' end of sgRNA could make RGENs more specific by comparing 5'-GGX_{20} (or 5'-GGGX_{19}) sgRNA with 5'-GX_{19} sgRNA. Four GX_{19} sgRNAs complexed with Cas9 induced indels equally efficiently at on-target and off-target sites, tolerating up to four nucleotide mismatches. In sharp contrast, GGX_{20} 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 GGX_{20} sgRNAs (Fig. 15). We noticed, however, that two GGX_{20} sgRNAs (VEGFA sites 1 and 3) were less active at on-target sites than were the corresponding GX_{19} 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.

[439]

[440] These results suggest that three factors-the use of synthetic guide RNA rather than guide RNA-encoding plasmids, dualRNA rather than sgRNA, and GGX_{20} sgRNA rather than GX_{19} sgRNA-have cumulative effects on the discrimination of off-target sites.

[441]

[442] **Example 7: Paired Cas9 nickases**

[443]

[444] In principle, single-strand breaks (SSBs) cannot be repaired by error-prone NHEJ but still trigger high fidelity homology-directed repair (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.

[445] 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).

[446]

[447] 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.

[448]

[449] **Example 8: Chromosomal DNA splicing induced by paired Cas9 nickases**

[450]

[451] Two concurrent DSBs produced by engineered nucleases such as ZFNs and TALENs can promote large deletions of the intervening chromosomal segments has reproted. 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.

[452]

[453] 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.

[454]

[455] 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).

[456]

[457] 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.

[458]

[459] Example 9: Genotyping with CRISPR/Cas-derived RNA-guided endonucleases

[460]

[461] 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 wildtype target sequences efficiently.

[462]

[463] 9-1. RGEN components

[464]

[465] crRNA and tracrRNA were prepared by *in vitro* transcription using MEGAshortcript 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’-GAAATTAATACGACTCACTATAGGX_{20}GTTTTAGAGCTATGCTGTTTTG-3’(

SEQ ID NO: 76), in which X_{20} is the target sequence, and its complementary oligonucleotide. The template for tracrRNA was synthesized by extension of forward and reverse oligonucleotides (5'-GAAATTAATACGACTCACTATAGGAACCATTCAAAACAGCATAGCAAG TTAAAATAAGGCTAGTCCG-3' (SEQ ID NO: 77) and 5'-AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAG CCTTATTTTAACTTGCTATG-3'(SEQ ID NO: 78)) using Phusion polymerase (New England Biolabs).

[466]

[467] **9-2. Recombinant Cas9 protein purification**

[468]

[469] 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 hour 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 immidazole, 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 immidazole, and eluted with buffer containing 20 mM Tris pH 8.0, 500 mM NaCl, and 250 mM immidazole. 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.

[470]

[471] **9-3. T7 endonuclease I assay**

[472]

[473] 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.

[474]

[475] **9-4. RGEN-RFLP assay**

[476]

[477] 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.

[478]

[479] Table 10

[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*
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*

[480]

[481] Table 11

[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	GCACAGGGTGGAACAAGATGGA	82
	R	GCCAGGTACCTATCGATTGTCAGG	83
CCR5(del32)	F	GAGCCAAGCTCTCCATCTAGT	84
	R	ACTCTGACTG GGTCACCAGC	85
C4BPB	*F1*	*TATTTGGCTGGTTGAAAGGG*	*86*
	R1	*AAAGTCATGAAATAAACACACCCA*	*87*
	F2	*CTGCATTGATATGGTAGTACCATG*	*88*
	R2	*GCTGTTCATTGCAATGGAATG*	*89*
CTNNB1	F	ATGGAGTTGGACATGGCCATGG	90
	R	ACTCACTATCCACAGTTCAGCATTTACC	91
KRAS	*F*	*TGGAGATAGCTGTCAGCAACTTT*	*92*
	R	*CAACAA AGCAAAGGTAAAGTTGGTAATAG*	*93*
PIK3CA	*F*	*GGTTTCAGGAGATGTGTTACAAGGC*	*94*
	R	*GATTGTGCAATTCCTATGCAATCGGTC*	*95*
NRAS	*F*	*CACTGGGTACTTAATCTGTAGCCTC*	*96*
	R	*GGTTCCAAGTCATTCCCAGTAGC*	*97*
IDH1	*F*	*CATCACTGCAGTTGTAGGTTATAACTATCC*	*98*
	R	*TTGAAAACCACAGATCTGGTTGAACC*	*99*
BRAF	*F*	*GGAGTGCCAAGAGAATATCTGG*	*100*
	R	*CTGAAACTGGTTTCAAAATATTCGTTTTAAGG*	*101*

PIBF	*F*	*GCTCTGTATGCCCTGTAGTAGG*	*102*
	R	*TTTGCATCTGACCTTACCTTTG*	*103*

[482]

[483] **9-5. Plasmid cleavage assay**

[484]

[485] 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, and100 mM EDTA. Products were resolved with 1% agarose gel electrophoresis and visualized with EtBr staining.

[486]

[487] **9-6. Strategy of RFLP**

[488]

[489] 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).

[490] 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.

[491]

[492] **9-7. Detection of mutations induced by the same RGENs using RGEN-mediated RFLP**

[493]

[494] 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).

[495]

[496] Table 12

[Table 12]

Target sequence of RGENs used in this study

Gene	Target sequence	SEQ ID NO
human C4BPB	AATGACCACTACATCCTCAA GGG	104
mouse Pibf1	AGATGATGTCTCATCATCAG AGG	105

[497]

[498] 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 (-/-).

[499]

[500] 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 -/- cloned 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 ZFNs, TALENs and RGENs.

[501]

[502] **9-8. Quantitative assay for RGEN-RFLP analysis**

[503]

[504] 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.

[505]

[506] **9-9. Analysis of mutant mouse founders using a RGEN-mediated RFLP genotyping**

[507]

[508] 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.

[509]

[510] **9-10. Detection of mutations induced in human cells by a CCR5-specific ZFN using RGENs**

[511]

[512] 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 ZFN and TALEN sites.

[513]

[514] **9-11. Detection of polymorphisms or variations using RGEN**

[515]

[516] 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.

[517]

[518] **9-12. Detection of recurrent mutations found in cancer and naturally-occurring polymorphisms through RGEN-RFLP analysis**

[519]

[520] 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.

[521]

[522] 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.

[523]

[524] 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.

[525]

[526] 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.

[527]

[528] 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.

[529]

[530] [References]

[531] 1. M. Jinek et al., Science 337, 816 (Aug 17, 2012).

[532] 2. H. Kim, E. Um, S. R. Cho, C. Jung, J. S. Kim, Nat Methods 8, 941 (Nov, 2011).

[533] 3. H. J. Kim, H. J. Lee, H. Kim, S. W. Cho, J. S. Kim, Genome Res 19, 1279 (Jul, 2009).

[534] 4. E. E. Perez et al., Nat Biotechnol 26, 808 (Jul, 2008).

5. J. C. Miller et al., Nat Biotechnol 29, 143 (Feb, 2011).

6. C. Mussolino et al., Nucleic Acids Res 39, 9283 (Nov, 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 (Sep, 2011).

9. R. Gabriel et al., Nat Biotechnol 29, 816 (Sep, 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 (Mar, 2012).

12. H. J. Lee, E. Kim, J. S. Kim, Genome Res 20, 81 (Jan, 2010).

13. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotech advance online publication (2013)

Throughout the specification and claims, unless the context requires otherwise, the word “comprise” or variations such as “comprises” or “comprising”, will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

Federal Court of Australia
District Registry: New South Wales
Division: General No: NSD1909/2018

TOOLGEN INCORPORATED
Appellant/Cross-Respondent

GRANT FISHER and another named in the schedule
First Respondent/First Cross-Appellant

ORDER

JUDGE: JUSTICE NICHOLAS

DATE OF ORDER: 23 May 2024

WHERE MADE: Sydney

THE COURT ORDERS THAT:

1. The appeal be dismissed.
2. The cross-appeal be allowed.
3. The decision of a Delegate of the Commissioner of Patents in relation to Australian Patent Application No. 2013335451 ("the Patent Application") given on 18 September 2018 be set aside except insofar as it relates to the costs awarded to the first respondent (the opponent) by the Delegate.
4. The complete specification of the Patent Application be amended pursuant to section 105(1A) of the *Patents Act 1990* (Cth) by substituting for the existing claims the amended claims set out in the Annexure to these orders.
5. The Patent Application proceed to grant as amended in accordance with order 4.
6. The appellant pay the respondents' costs incurred up to and including 16 November 2023 of and incidental to their consideration of the interlocutory application dated 1 September 2023.

7. The appellant serve a copy of these orders on the Commissioner of Patents within 7 days.

Date that entry is stamped: 23 May 2024

Sia Lagos

Registrar

AMENDED CLAIMS

1. A method of introducing a site-specific, double stranded break at a target nucleic acid sequence in a eukaryotic cell, the method comprising introducing into the eukaryotic cell a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system, wherein the CRISPR/Cas system comprises:

a) a nucleic acid encoding a Cas9 polypeptide derived from *Streptococcus pyogenes* that contains a nuclear localization signal and recognizes a 5'-NGG-3' trinucleotide protospacer adjacent motif (PAM), wherein the nucleic acid is codon-optimized for expression in eukaryotic cells; and

b) an *in vitro* transcribed sg RNA (+48) derived from *Streptococcus pyogenes* that hybridizes to the target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion, wherein the target nucleic acid sequence comprises a first strand that binds to the crRNA portion and a second strand having a 5'-NGG-3' PAM,

and wherein the Cas9 polypeptide and the guide RNA form a Cas9/RNA complex in the eukaryotic cell, whereby a site specific, double stranded break at the target nucleic acid sequence is introduced.

2. The method of claim 1 wherein the eukaryotic cell is a mammalian cell.

3. The method of claim 2 wherein the mammalian cell is a human cell.

4. The method of any one of claims 1 - 3 wherein the target nucleic acid sequence is a genomic sequence located at its endogenous site in the genome of the eukaryotic cell.

5. The method of any one of claims 1 - 4 wherein said guide RNA comprises 2 additional guanine nucleotides at the 5' end.

Schedule

No: NSD1909/2018

Federal Court of Australia
District Registry: New South Wales
Division: General

Second Respondent/Second Cross-Appellant	ACN 004 552 363 PTY LTD

AMENDED CLAIMS

1. A method of introducing a site-specific, double stranded break at a target nucleic acid sequence in a eukaryotic cell, the method comprising introducing into the eukaryotic cell a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system, wherein the CRISPR/Cas system comprises:

a) a nucleic acid encoding a Cas9 polypeptide derived from *Streptococcus pyogenes* that contains a nuclear localization signal and recognizes a 5'-NGG-3' trinucleotide protospacer adjacent motif (PAM), wherein the nucleic acid is codon-optimized for expression in eukaryotic cells; and

b) an *in vitro* transcribed sg RNA (+48) derived from *Streptococcus pyogenes* that hybridizes to the target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion, wherein the target nucleic acid sequence comprises a first strand that binds to the crRNA portion and a second strand having a 5'-NGG-3' PAM,

and wherein the Cas9 polypeptide and the guide RNA form a Cas9/RNA complex in the eukaryotic cell, whereby a site specific, double stranded break at the target nucleic acid sequence is introduced.

2. The method of claim 1 wherein the eukaryotic cell is a mammalian cell.

3. The method of claim 2 wherein the mammalian cell is a human cell.

4. The method of any one of claims 1 -3 wherein the target nucleic acid sequence is a genomic sequence located at its endogenous site in the genome of the eukaryotic cell.

5. The method of any one of claims 1-4 wherein said guide RNA comprises 2 additional guanine nucleotides at the 5' end.

CLAIMS

1. A composition comprising a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system for use in introducing a site-specific, double stranded break at a target nucleic acid sequence in a eukaryotic cell, said CRISPR/Cas system comprising (i) a nucleic acid encoding a Cas9 polypeptide comprising a nuclear localization sequence, and (ii) a nucleic acid encoding a guide RNA that hybridizes to a target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion.

2. The composition of claim 1, wherein said Cas9 polypeptide is a *Streptococcus* Cas9 polypeptide.

3. The composition of claim 2, wherein said Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 polypeptide.

4. The composition of any one of claims 1-3, wherein said nucleic acid encoding a Cas9 polypeptide is codon-optimized for expression in eukaryotic cells.

5. The composition of claim 4, wherein said nucleic acid encoding a Cas9 polypeptide is codon-optimized for expression in mammalian cells.

6. The composition of any one of claims 1-5, wherein said nuclear localization sequence is located at the C terminus of the Cas9 polypeptide.

7. The composition of any one of claims 1-5, wherein the target nucleic acid is an endogenous target nucleic acid.

8. The composition of any one of claims 1-5, wherein the guide RNA is in the form of a vector.

9. The composition of any one of claims 1-5, wherein said guide RNA comprises 2 additional guanine nucleotides at the 5’ end.

10. A method of introducing a site-specific, double-stranded break at a target nucleic acid sequence in a eukaryotic cell, the method comprising introducing into the eukaryotic cell a Type II Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas system, wherein the CRISPR/Cas system comprises:

a) a nucleic acid encoding a Cas9 polypeptide comprising a nuclear localization signal, wherein the nucleic acid is codon-optimized for expression in eukaryotic cells, and

b) a nucleic acid encoding a guide RNA that hybridizes to the target nucleic acid, wherein the guide RNA is a chimeric guide RNA comprising a CRISPR RNA (crRNA) portion fused to a trans activating crRNA (tracrRNA) portion, wherein the target nucleic acid sequence comprises a first strand that binds to the crRNA portion and a second strand having a trinucleotide protospacer adjacent motif (PAM),

and wherein the Cas9 polypeptide and the guide RNA form a Cas9/RNA complex in the eukaryotic cell, whereby a site-specific, double stranded break at the target nucleic acid sequence is introduced.

11. The method of claim 10, wherein the Cas9 polypeptide is a *Streptococcus* Cas9 polypeptide.

12. The method of claim 11, wherein the Cas9 polypeptide is a *Streptococcus pyogenes* Cas9 polypeptide.

13. The method of any one of claims 10-12, wherein the nucleic acid encoding the Cas 9 polypeptide is codon-optimized for expression in mammalian cells.

14. The method of any one of claims 10-13, wherein the nuclear localization signal is located at the C terminus of the Cas9 polypeptide.

15. The method of any one of claims 10-14, wherein the eukaryotic cell is a mammalian cell.

16. The method of claim 15, wherein the mammalian cell is a human cell.

17. The method of any one of claims 10-16, wherein the target nucleic acid sequence is a genomic sequence located at its endogenous site in the genome of the eukaryotic cell.

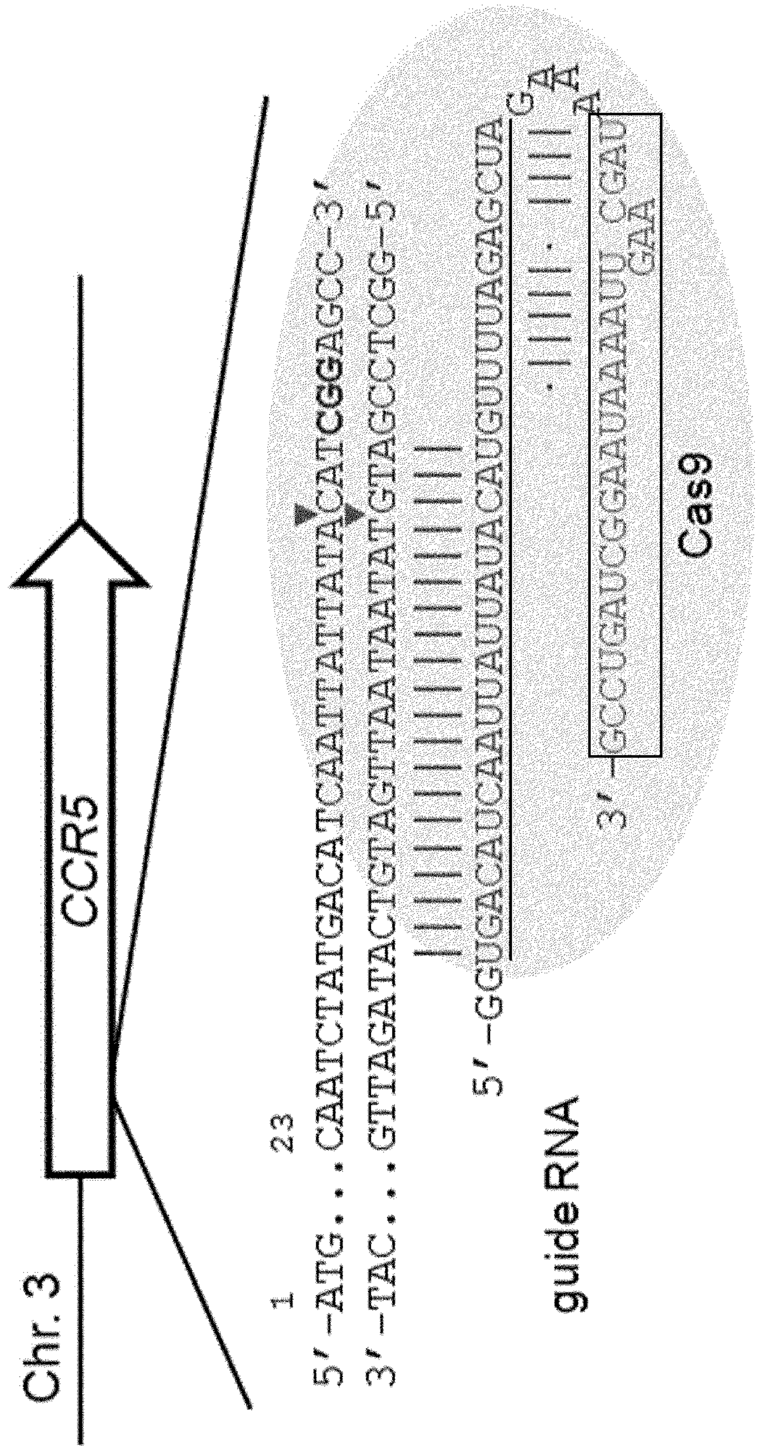
18. The method of any one of claims 10-16, wherein the nucleic acid encoding the guide RNA is a vector.

19. The method of any one of claims 10-16, wherein the nucleic acid encoding the guide RNA is in vitro transcribed RNA.

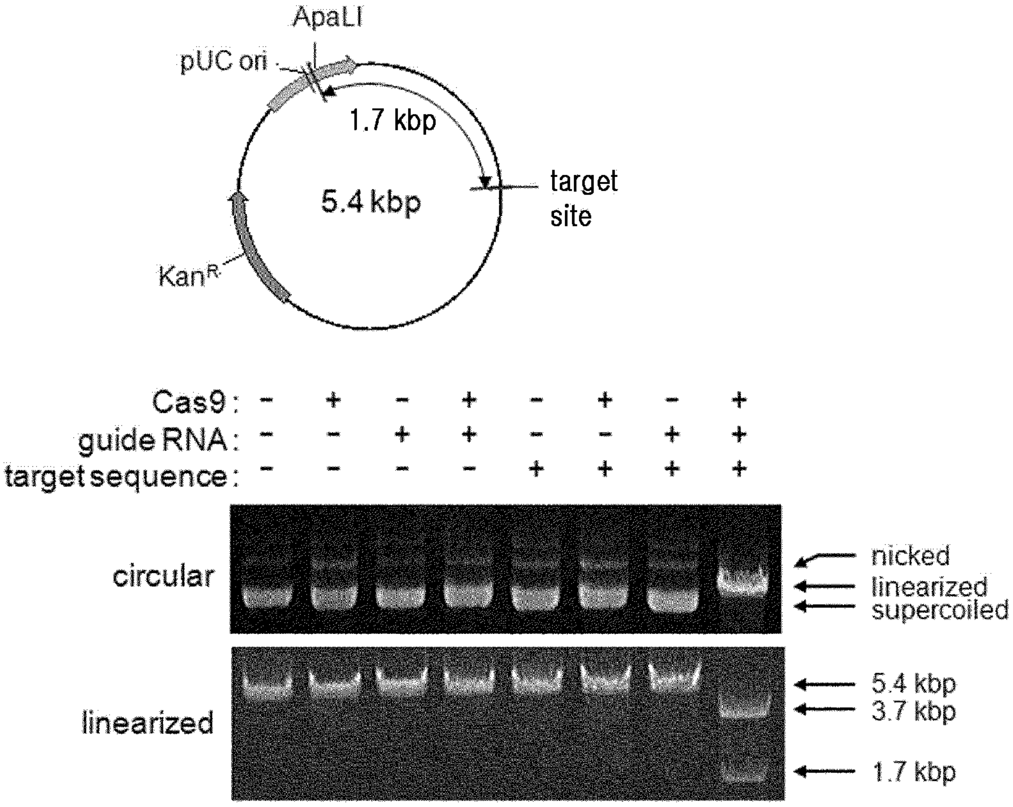
20. The method of any one of claims 10-16, wherein said guide RNA comprises 2 additional guanine nucleotides at the 5′ end.

21. The method of any one of claims 10-16, wherein the nucleic acid encoding the Cas9 polypeptide is introduced into the eukaryotic cell before introducing the nucleic acid encoding the guide RNA into the eukaryotic cell.

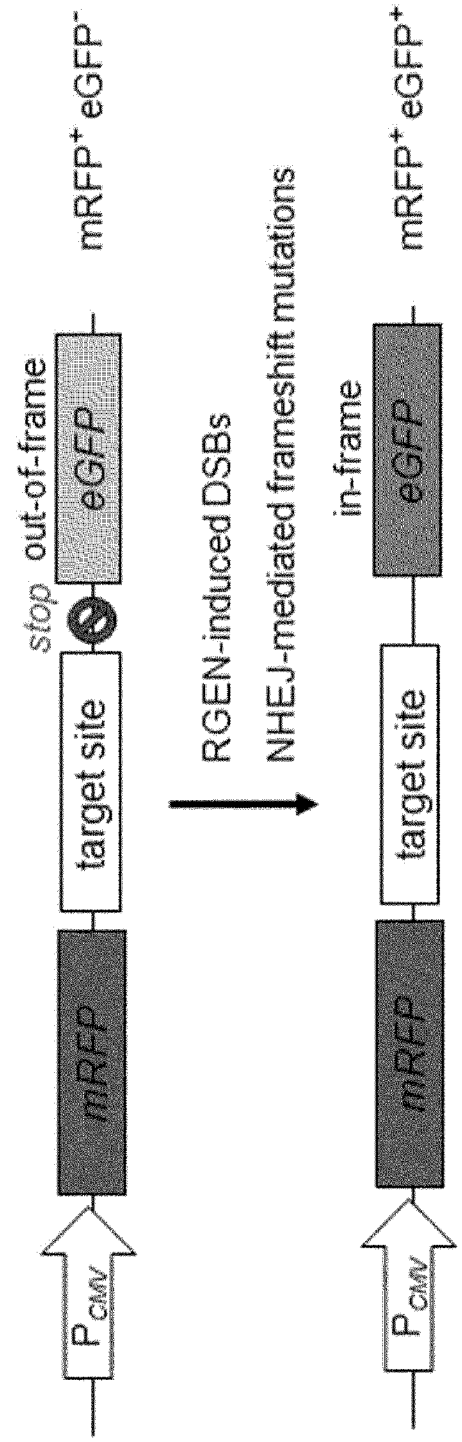
[Fig. 1a]


Chr. 3
CCR5
1
23
5′-ATG...CAATCTATGACATCAATTATTATACATCGGAGCC-3′
3′-TAC...GTTAGATACTGTAGTTAATAATATGTAGCCTCGG-5′
5′-GGUGACAUCAAUUAUUAUACAUGUUUUAGAGCUA
GAAA
3′-GCCUGAUCGGAAUAAAAUUCGAU
GAA
guide RNA
Cas9

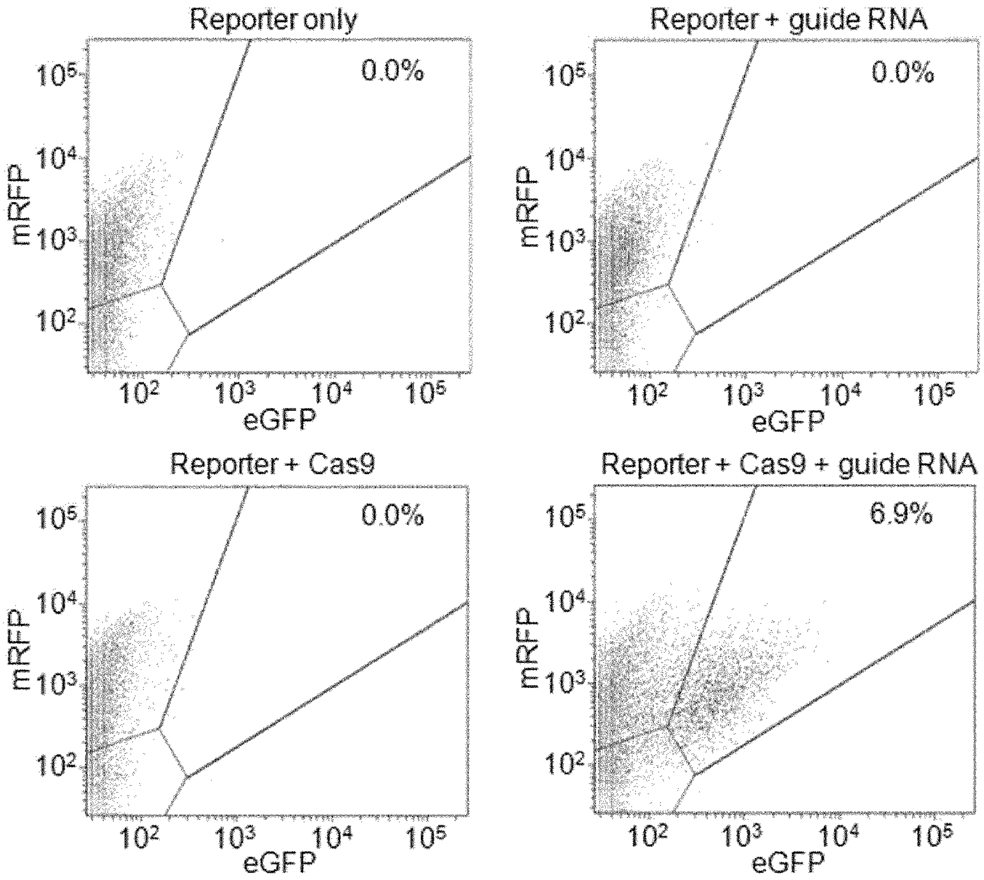
[Fig. 1b]


ApaLI
pUC ori
1.7 kbp
5.4 kbp
target site
KanR
Cas9 : - + - + - + - +
guide RNA : - - + + - - + +
target sequence : - - - - + + + +
circular
nicked
linearized
supercoiled
linearized
5.4 kbp
3.7 kbp
1.7 kbp

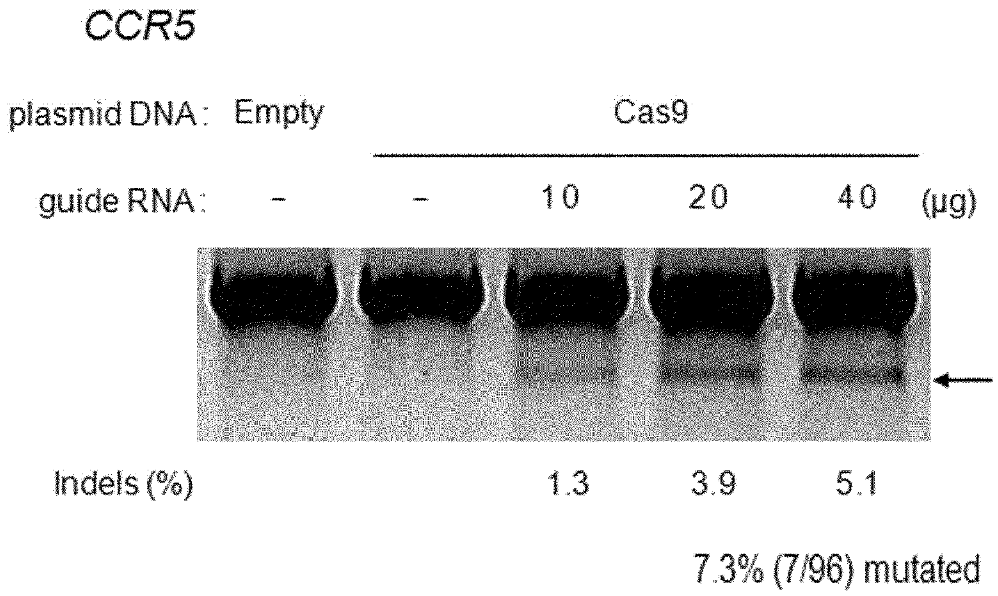
[Fig. 2a]


P_{CMV}
mRFP
target site
stop
out-of-frame
eGFP
mRFP$^+$ eGFP$^-$
RGEN-induced DSBs
NHEJ-mediated frameshift mutations
P_{CMV}
mRFP
target site
in-frame
eGFP
mRFP$^+$ eGFP$^+$

[Fig. 2b]

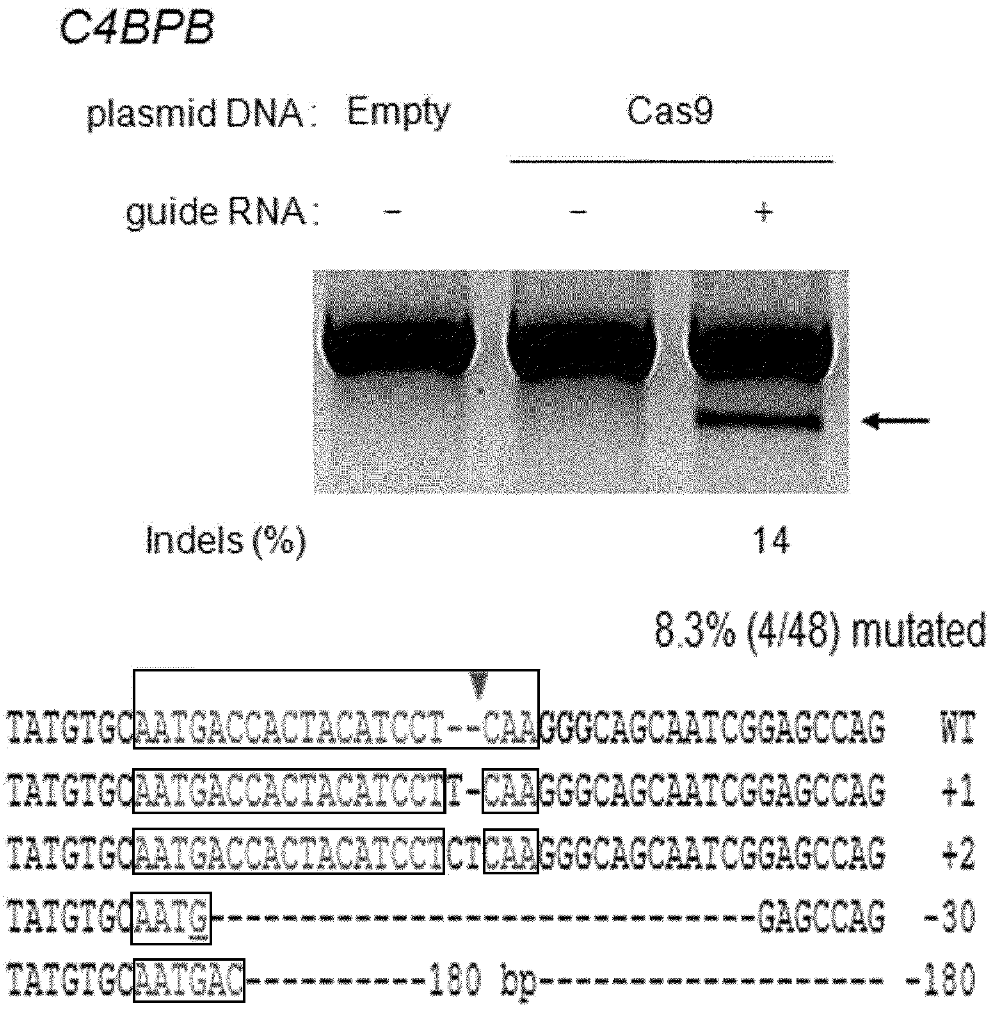


[Fig. 3a]

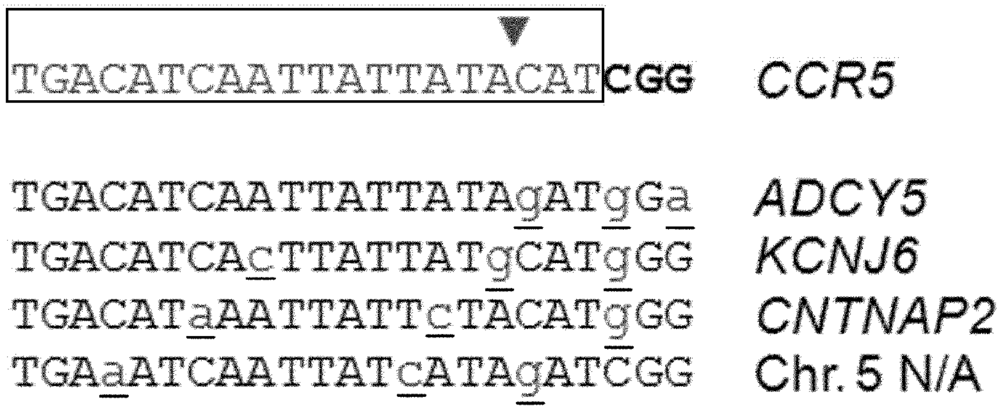


```
CAATCTATGACATCAATTATTATA-CATCGGAGCCCTGCCAAAAAATCAA  WT
CAATCTATGACATCAATTATTATAACATCGGAGCCCTGCCAAAAAATCAA  +1
CAATCTATGACATCAATTATTAT--------------GCCAAAAAATCAA -13
CAATCTATGACATC---------------GGAGCCCTGCCAAAAAATCAA -14
CAATCTATGACAT-------------------GCCCTGCCAAAAAATCAA -18
CAATCTATGACATCAATTATTAT--------------------AAATCAA -19
CAATCTATGACATC-------------------------CAAAAAATCAA -24
CAATCTATGACA-------------------------------AAATCAA -30
```

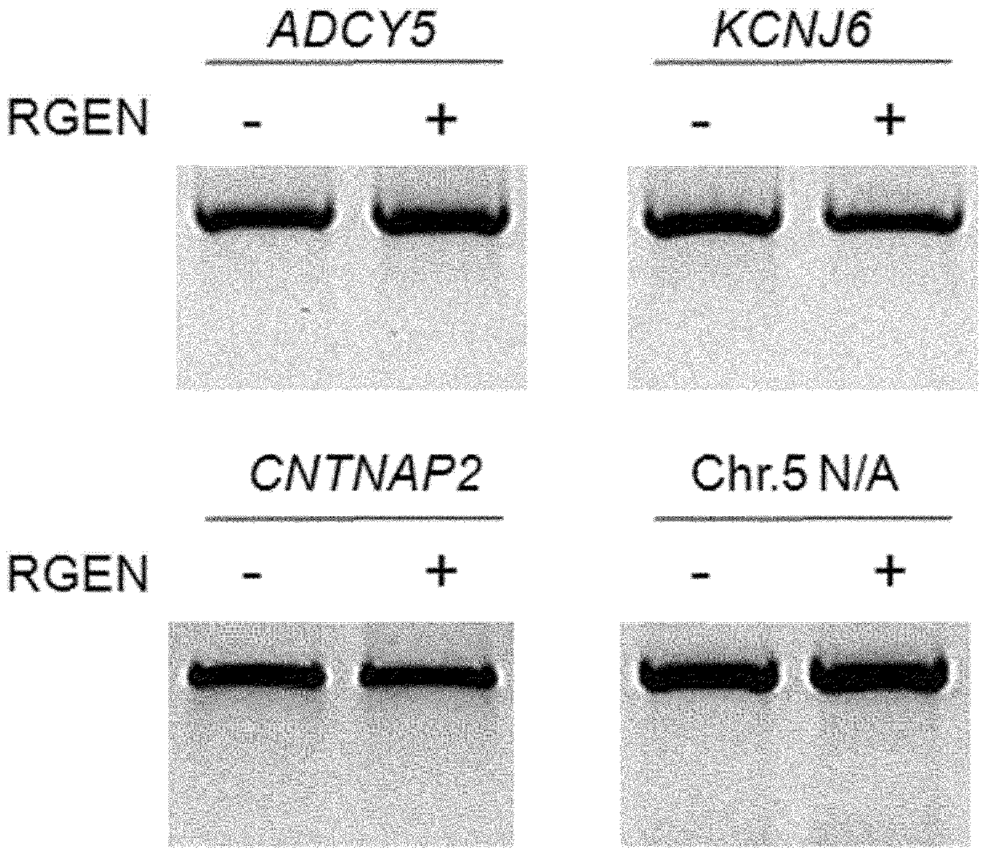
[Fig. 3b]


C4BPB
plasmid DNA : Empty Cas9
guide RNA : - - +
Indels (%) 14
8.3% (4/48) mutated
TATGTGCAATGACCACTACATCCT--CAAGGGCAGCAATCGGAGCCAG WT
TATGTGCAATGACCACTACATCCTT-CAAGGGCAGCAATCGGAGCCAG +1
TATGTGCAATGACCACTACATCCTCTCAAGGGCAGCAATCGGAGCCAG +2
TATGTGCAATG------------------------------GAGCCAG -30
TATGTGCAATGAC---------180 bp----------------- -180

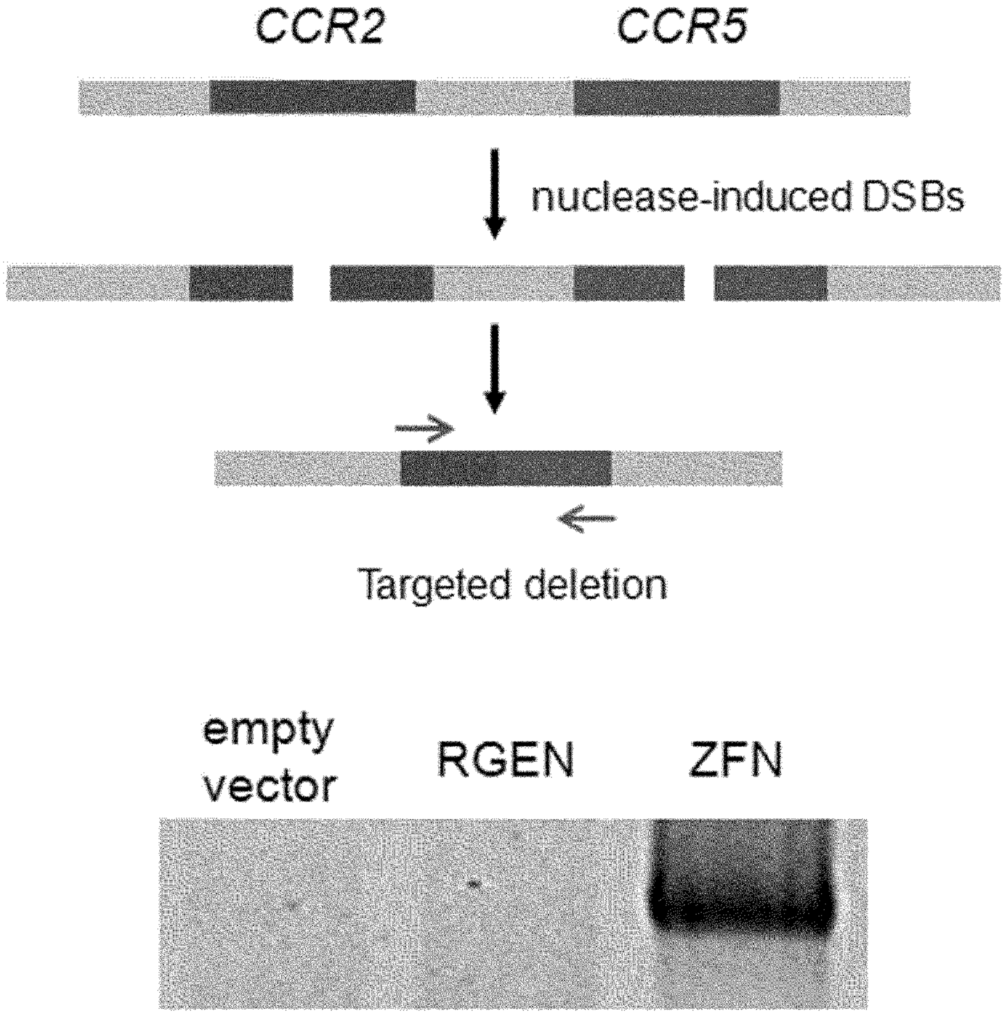
[Fig. 4a]


TGACATCAATTATTATACATCGG CCR5
TGACATCAATTATTATAgATgGa ADCY5
TGACATCAcTTATTATgCATgGG KCNJ6
TGACATaAATTATTcTACATgGG CNTNAP2
TGAaATCAATTATcATAgATCGG Chr. 5 N/A

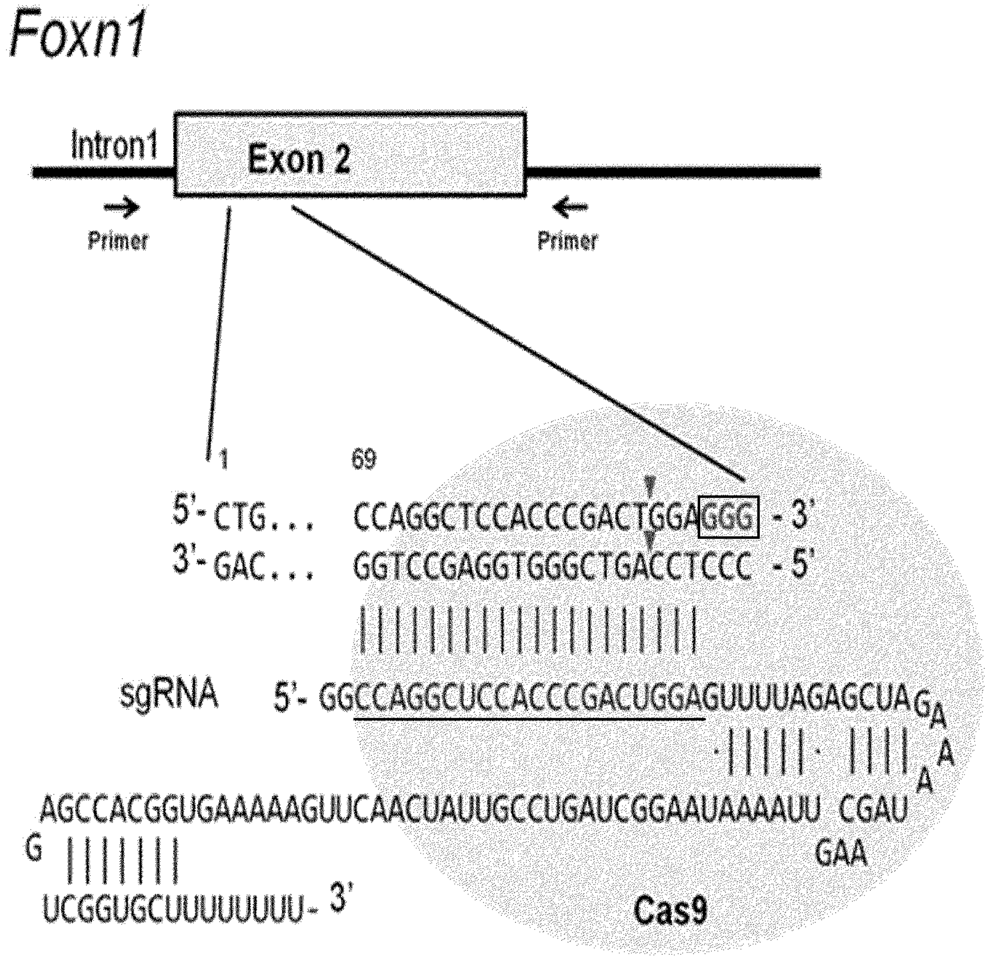
[Fig. 4b]



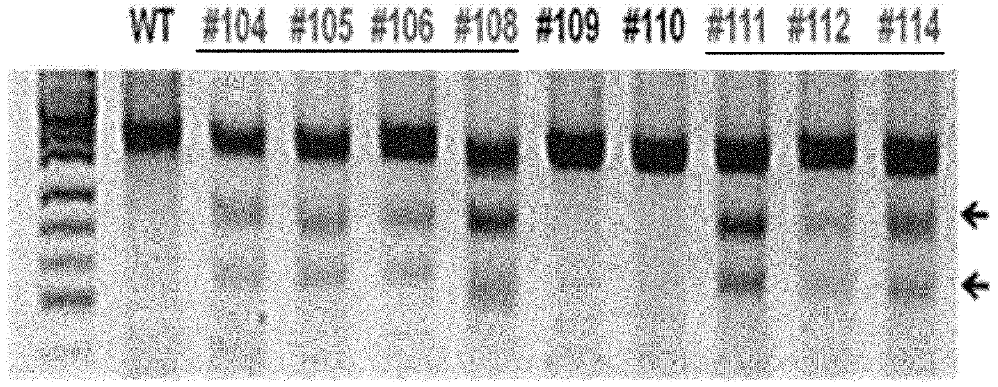
[Fig. 4c]


CCR2
CCR5
nuclease-induced DSBs
Targeted deletion
empty vector
RGEN
ZFN

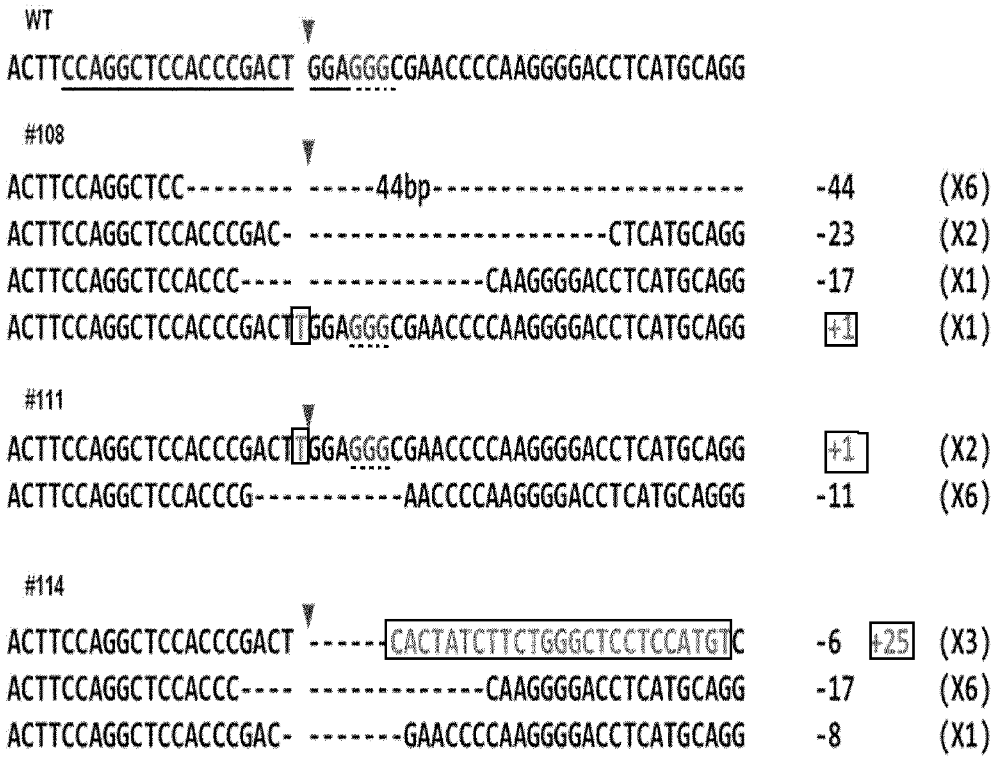
[Fig. 5a]


Foxn1
Intron1
Exon 2
Primer
Primer
1
69
5'- CTG... CCAGGCTCCACCCGACTGGAGGG - 3'
3'- GAC... GGTCCGAGGTGGGCTGACCTCCC - 5'
sgRNA
5'- GGCCAGGCUCCACCCGACUGGAGUUUUAGAGCUAGA
A
A
CGAU
GAA
AGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCGGAAUAAAAUU
G
UCGGUGCUUUUUUUU- 3'
Cas9

[Fig. 5b]

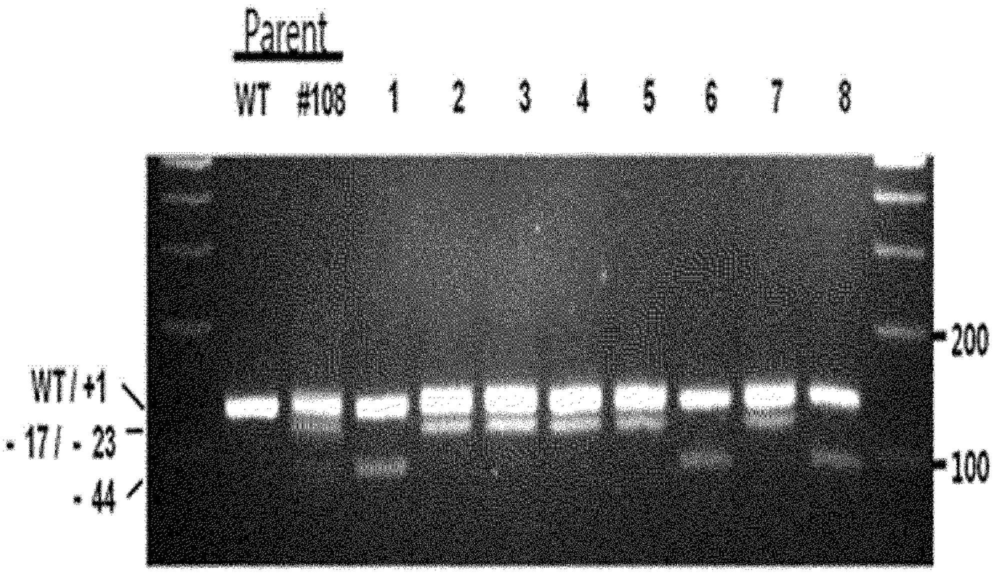

WT
#104
#105
#106
#108
#109
#110
#111
#112
#114

[Fig. 5c]



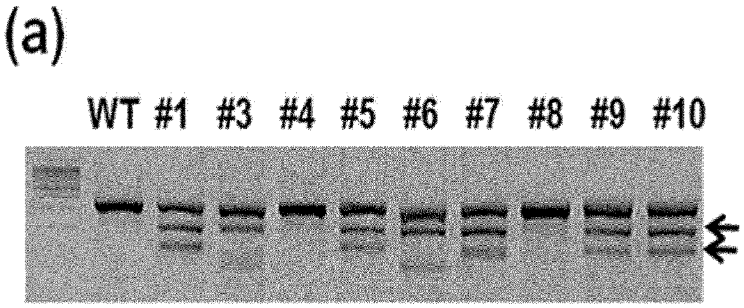
```
WT
ACTTCCAGGCTCCACCCGACT GGAGGGCGAACCCCAAGGGGACCTCATGCAGG

#108
ACTTCCAGGCTCC-------- -----44bp---------------------     -44    (X6)
ACTTCCAGGCTCCACCCGAC- ----------------------CTCATGCAGG   -23    (X2)
ACTTCCAGGCTCCACCC---- -------------CAAGGGGACCTCATGCAGG   -17    (X1)
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCAGG   +1     (X1)

#111
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCAGG   +1     (X2)
ACTTCCAGGCTCCACCCG-----------AACCCCAAGGGGACCTCATGCAGGG   -11    (X6)

#114
ACTTCCAGGCTCCACCCGACT -----CACTATCTTCTGGGCTCCTCCATGTC    -6 +25 (X3)
ACTTCCAGGCTCCACCC---- -------------CAAGGGGACCTCATGCAGG   -17    (X6)
ACTTCCAGGCTCCACCCGAC- -------GAACCCCAAGGGGACCTCATGCAGG   -8     (X1)
```

[Fig. 5d]



[Fig. 6]

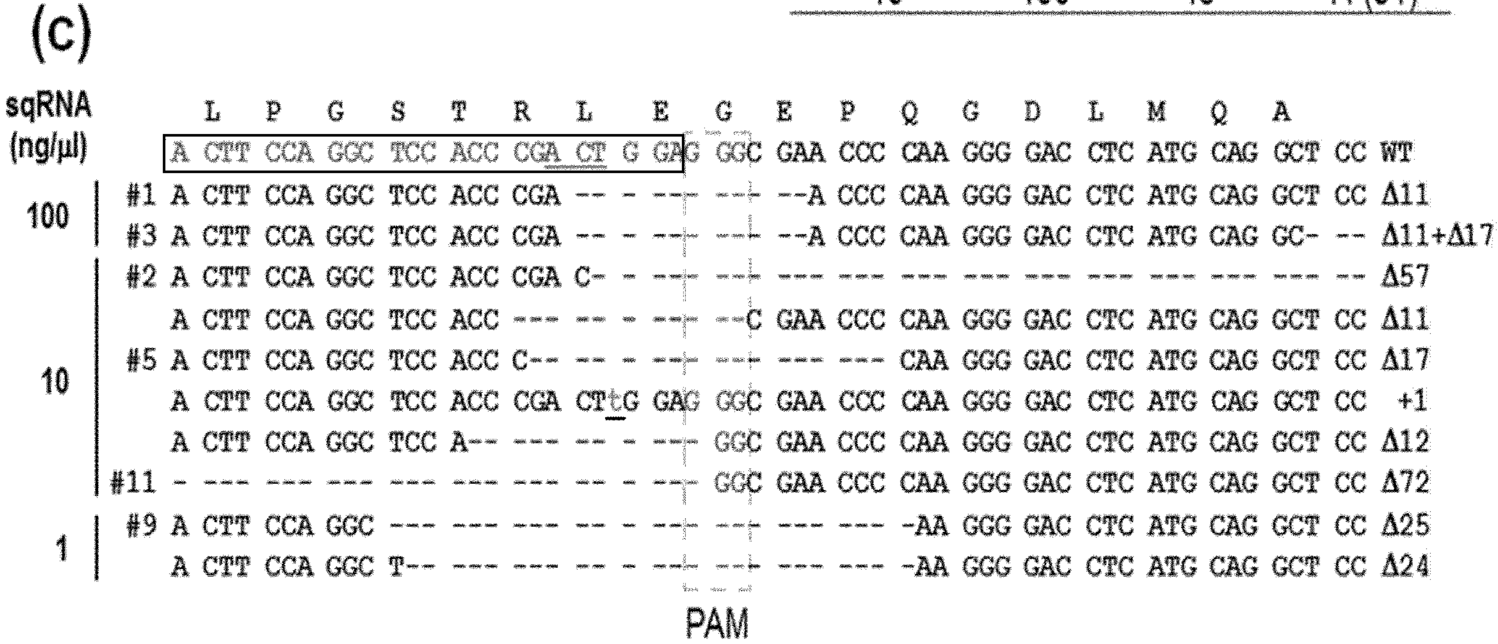
(a)



(b)

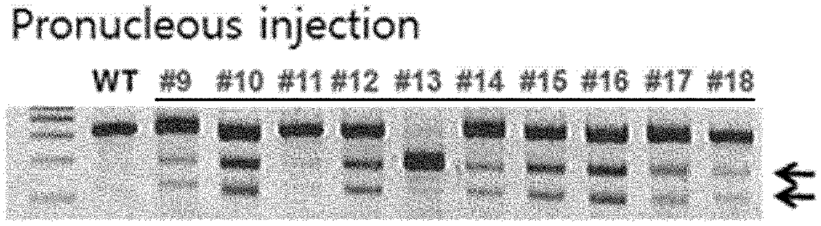
Cas9 mRNA (ng/μl)	sgRNA (ng/μl)	Tested embryos	Mutants (%)
10	1	27	9 (33)
10	10	49	28 (57)
10	100	45	41 (91)

(c)



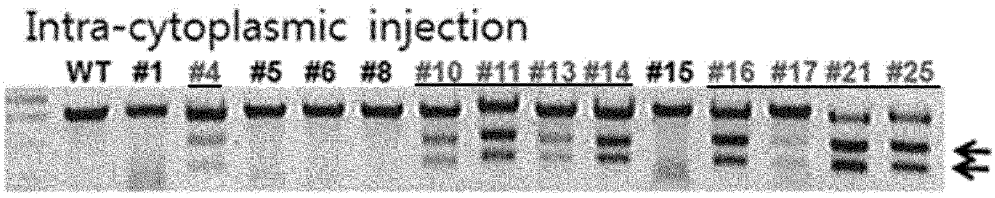
```
sqRNA          L   P   G   S   T   R   L    E   G   E   P   Q   G   D   L   M   Q   A
(ng/μl)      A CTT CCA GGC TCC ACC CGA CT G GAG GGC GAA CCC CAA GGG GAC CTC ATG CAG GCT CC WT
100      #1  A CTT CCA GGC TCC ACC CGA -- - --- --- --A CCC CAA GGG GAC CTC ATG CAG GCT CC Δ11
         #3  A CTT CCA GGC TCC ACC CGA -- - --- --- --A CCC CAA GGG GAC CTC ATG CAG GC- -- Δ11+Δ17
10       #2  A CTT CCA GGC TCC ACC CGA C- - --- --- --- --- --- --- --- --- --- --- --- -- Δ57
             A CTT CCA GGC TCC ACC --- -- - --- --C GAA CCC CAA GGG GAC CTC ATG CAG GCT CC Δ11
         #5  A CTT CCA GGC TCC ACC C-- -- - --- --- --- --- CAA GGG GAC CTC ATG CAG GCT CC Δ17
             A CTT CCA GGC TCC ACC CGA CTtG GAG GGC GAA CCC CAA GGG GAC CTC ATG CAG GCT CC +1
             A CTT CCA GGC TCC A-- --- -- - --- GGC GAA CCC CAA GGG GAC CTC ATG CAG GCT CC Δ12
        #11  - --- --- --- --- --- --- -- - --- GGC GAA CCC CAA GGG GAC CTC ATG CAG GCT CC Δ72
1        #9  A CTT CCA GGC --- --- --- -- - --- --- --- --- -AA GGG GAC CTC ATG CAG GCT CC Δ25
             A CTT CCA GGC T-- --- --- -- - --- --- --- --- -AA GGG GAC CTC ATG CAG GCT CC Δ24
                                              PAM
```

[Fig. 7a]



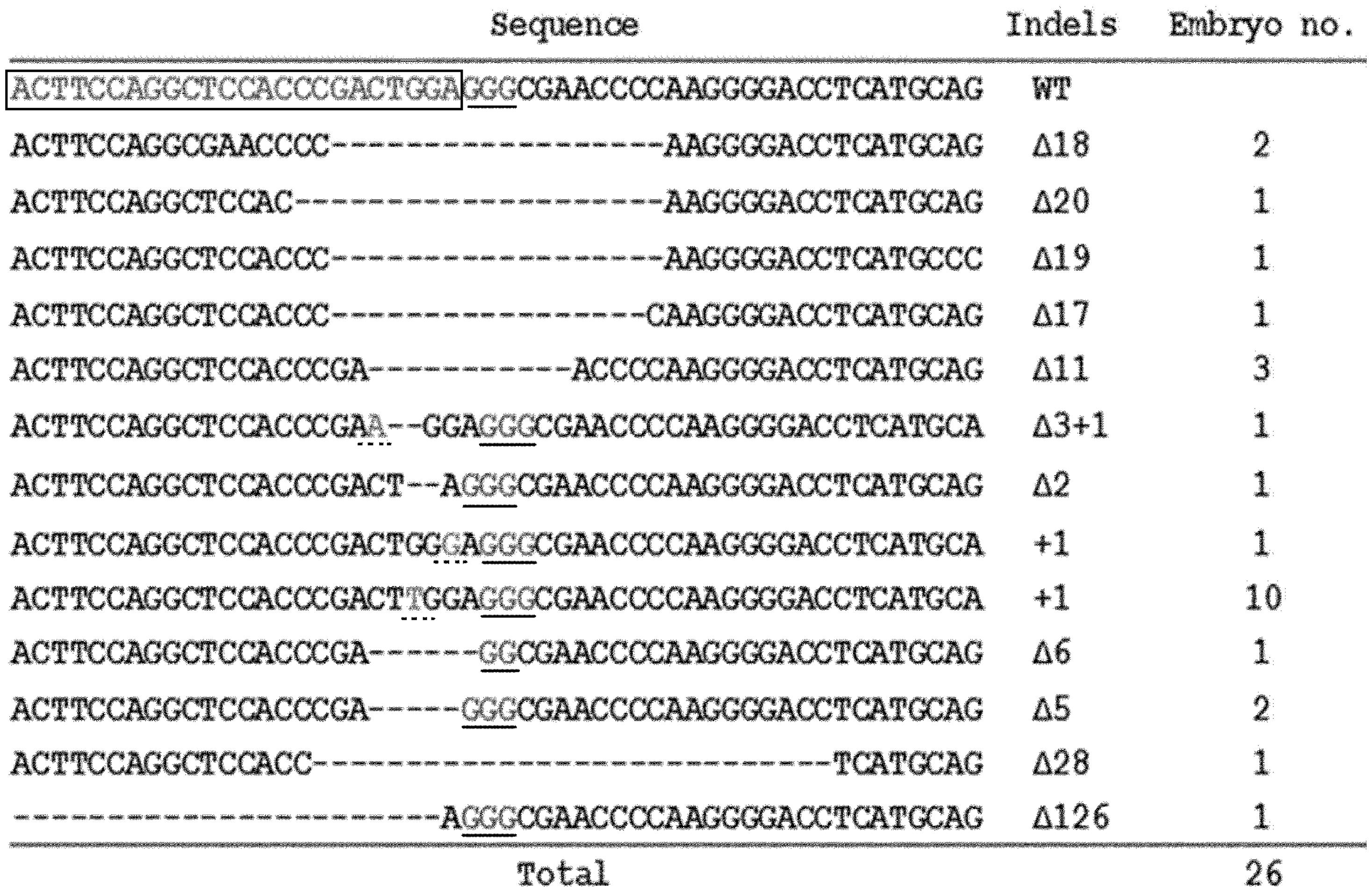
Cas9 protein (nM)	sgRNA (nM)	Tested embryos	Mutants (%)
2	4	20	3 (15%)
20	40	15	5 (33%)
200	400	17	15 (88%)

[Fig. 7b]



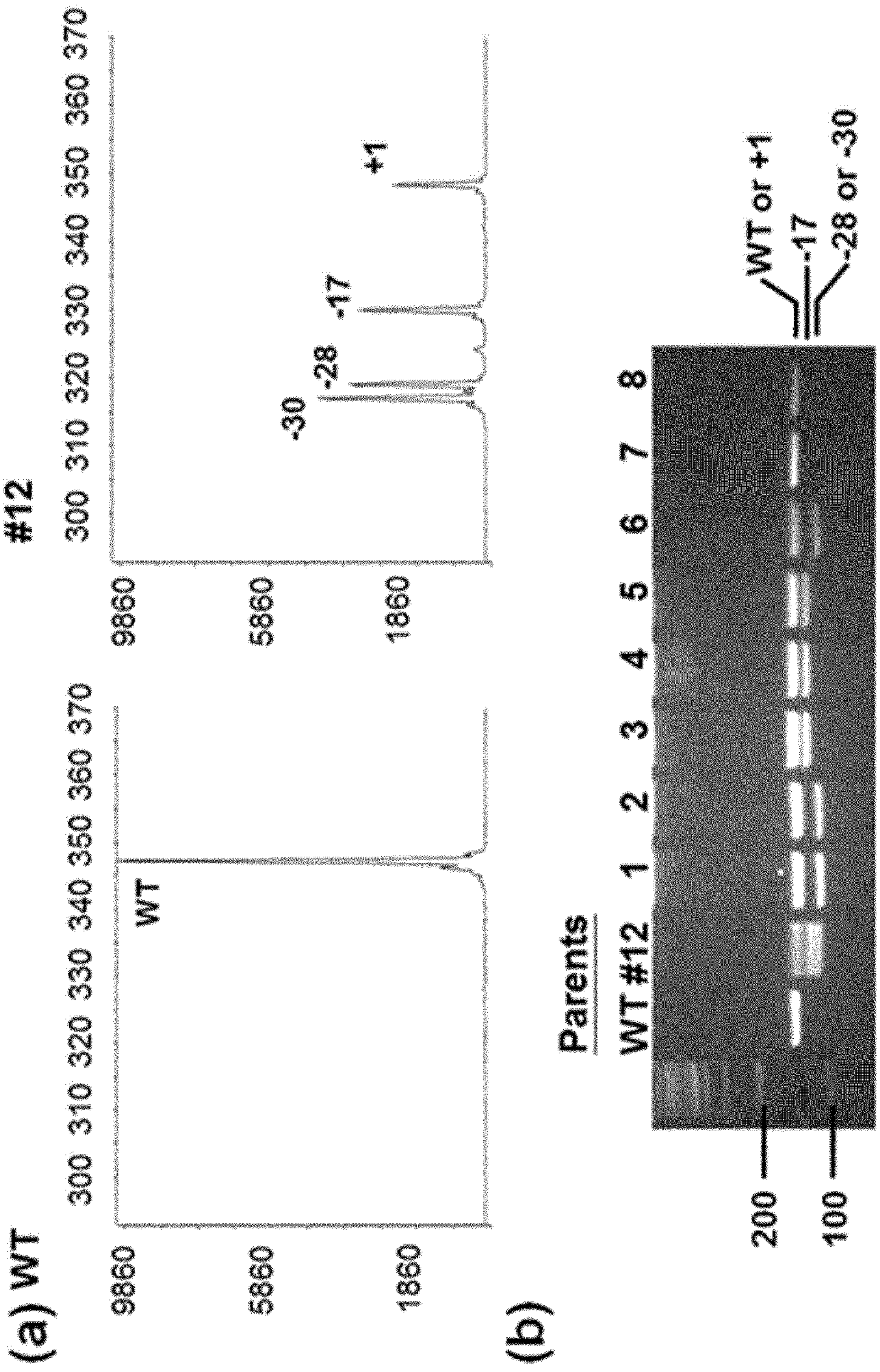
Cas9 protein (nM)	sgRNA (nM)	Tested embryos	Mutants (%)
2	4	18	1 (6%)
20	40	19	6 (32%)
200	400	14	10 (71%)

[Fig. 7c]

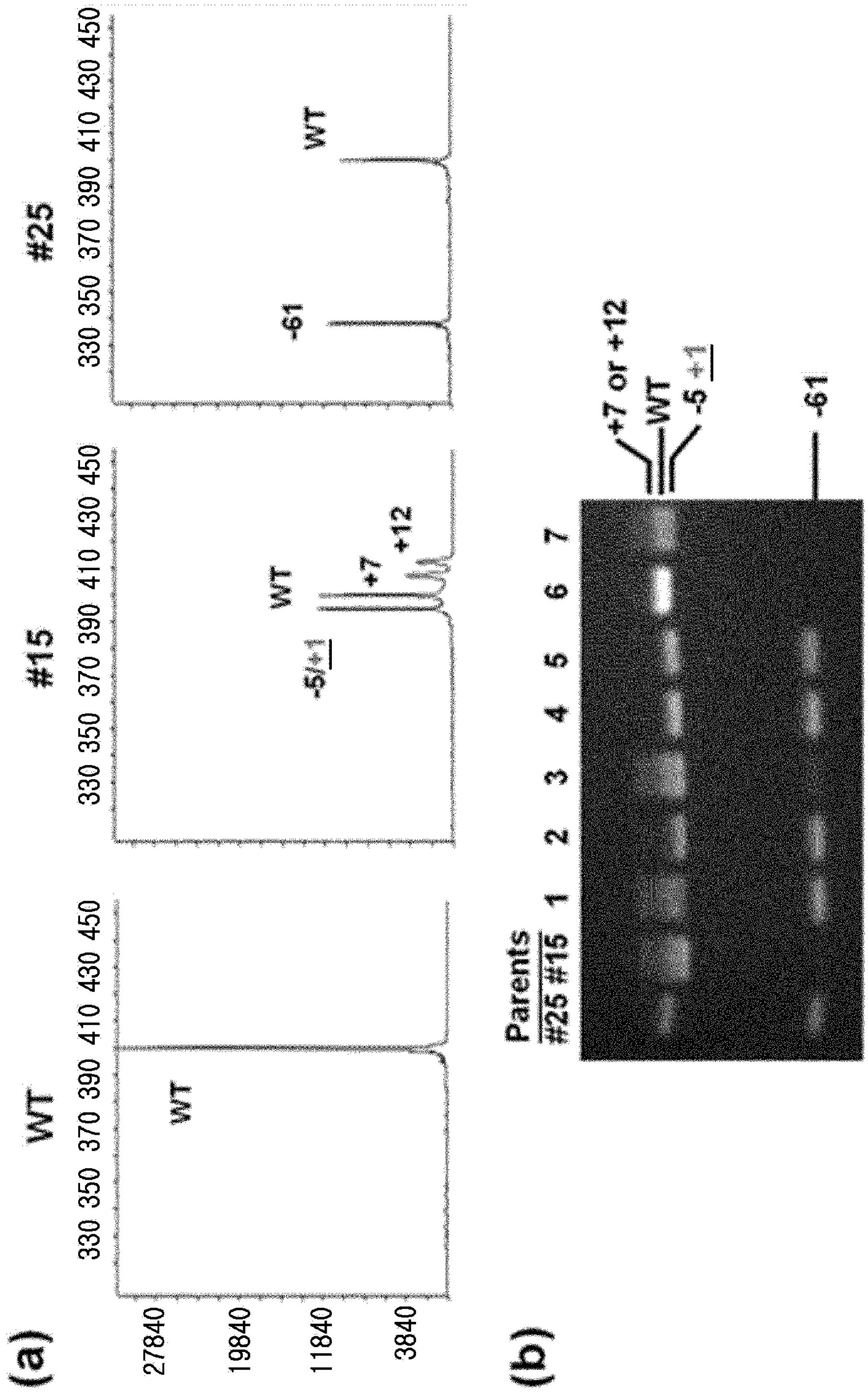


Sequence	Indels	Embryo no.
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGACCTCATGCAG	WT	
ACTTCCAGGCGAACCCC------------------AAGGGGACCTCATGCAG	Δ18	2
ACTTCCAGGCTCCAC--------------------AAGGGGACCTCATGCAG	Δ20	1
ACTTCCAGGCTCCACCC------------------AAGGGGACCTCATGCCC	Δ19	1
ACTTCCAGGCTCCACCC-----------------CAAGGGGACCTCATGCAG	Δ17	1
ACTTCCAGGCTCCACCCGA-----------ACCCCAAGGGGACCTCATGCAG	Δ11	3
ACTTCCAGGCTCCACCCGAA--GGAGGGCGAACCCCAAGGGGACCTCATGCA	Δ3+1	1
ACTTCCAGGCTCCACCCGACT--AGGGCGAACCCCAAGGGGACCTCATGCAG	Δ2	1
ACTTCCAGGCTCCACCCGACTGGGAGGGCGAACCCCAAGGGGACCTCATGCA	+1	1
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCA	+1	10
ACTTCCAGGCTCCACCCGA------GGCGAACCCCAAGGGGACCTCATGCAG	Δ6	1
ACTTCCAGGCTCCACCCGA-----GGGCGAACCCCAAGGGGACCTCATGCAG	Δ5	2
ACTTCCAGGCTCCACC----------------------------TCATGCAG	Δ28	1
-----------------------AGGGCGAACCCCAAGGGGACCTCATGCAG	Δ126	1
Total		26

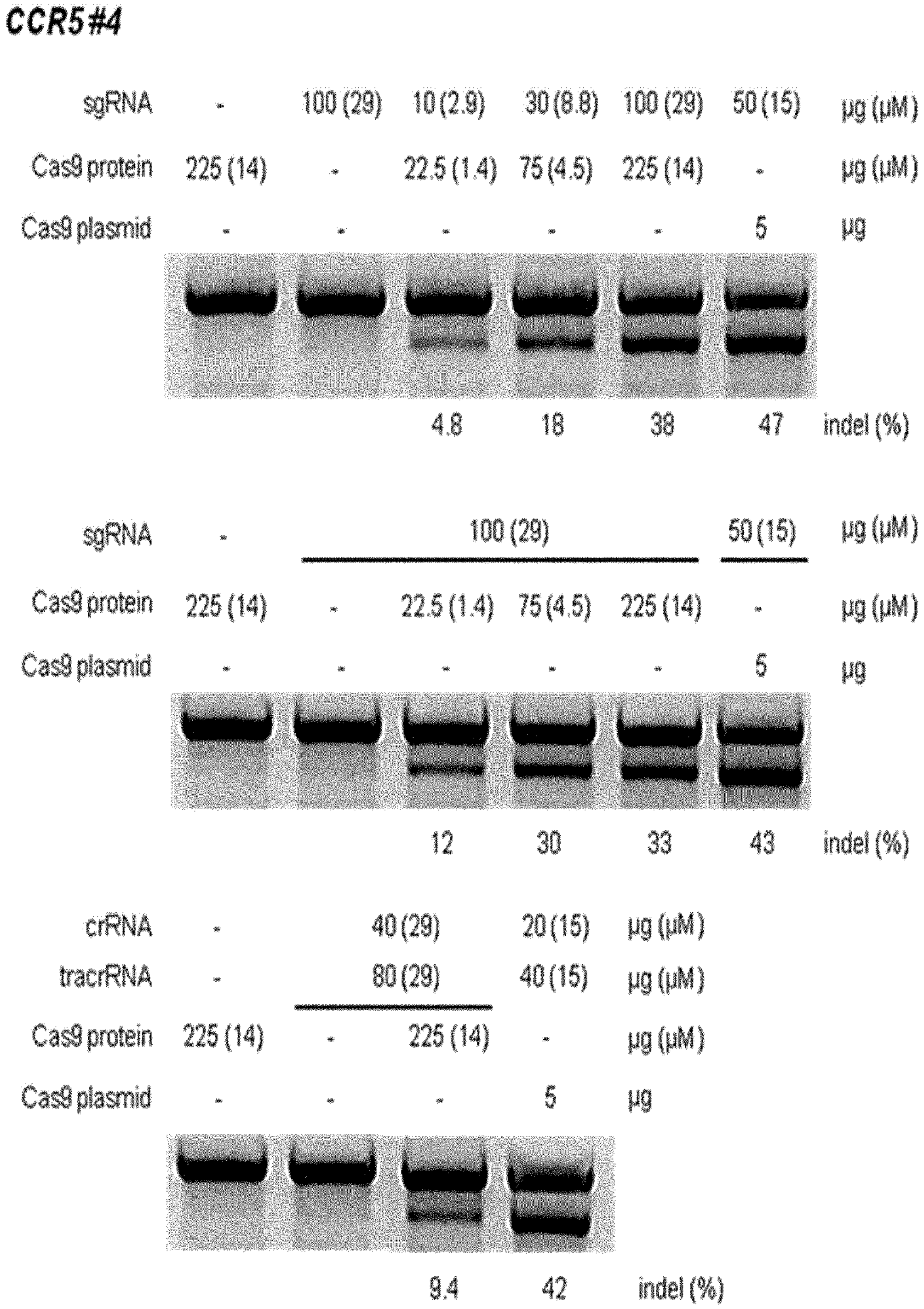
[Fig. 8]


(a) WT
9860
5860
1860
300
310
320
330
340
350
360
370
WT
#12
-30
-28
-17
+1
(b)
Parents
WT
#12
1
2
3
4
5
6
7
8
WT or +1
-17
-28 or -30
200
100

[Fig. 9]

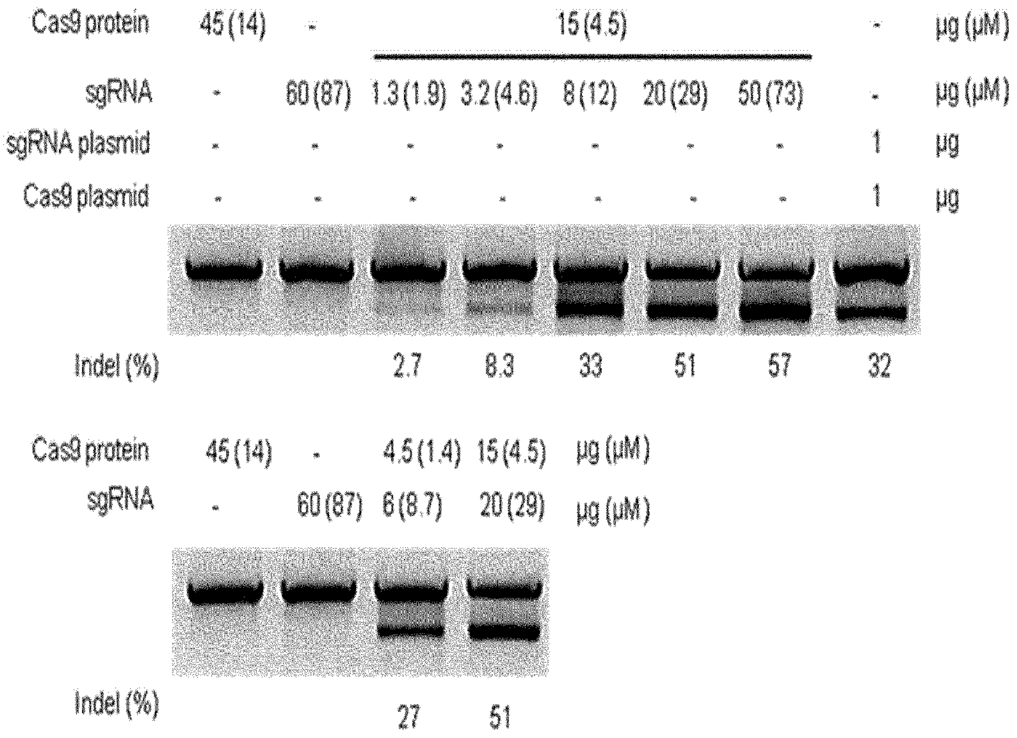

(a)
WT
#15
#25
330 350 370 390 410 430 450
27840
19840
11840
3840
WT
-5/+1
WT
+7
+12
-61
WT
(b)
Parents
#25 #15
1
2
3
4
5
6
7
+7 or +12
WT
-5 +1
-61

[Fig. 10a]


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. 10b]

CCR5#4



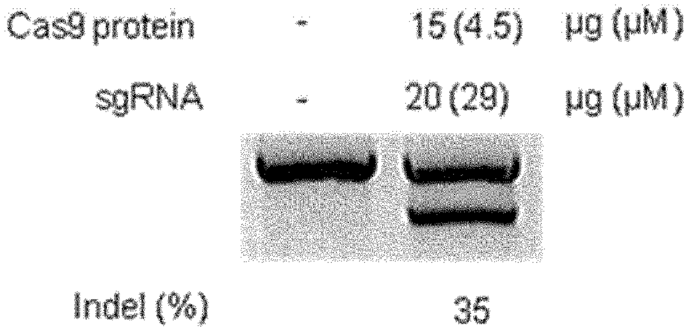
[Fig. 10c]

CCR5

CAATCTATGACATCAATTATTATA-CAT**CGG**AGCCCTGCCAAAAAATCAA WT
CAATCTATGACATCAATTATTAT-----CGGAGCCCTGCCAAAAAATCAA -4
CAATCTATGACATCAATTAT-----CATCGGAGCCCTGCCAAAAAATCAA -4
CAATCTATGACATCAATTAT--------CGGAGCCCTGCCAAAAAATCAA -7
CAATCTATGACATCAATTATTAT--CATCGGAGCCCTGCCAAAAAATCAA -1
CAATCTATGACATCAATTATTATAACATCGGAGCCCTGCCAAAAAATCAA +1
CAATCTATGACAA-----------------GAGCCCTGCCAAAAAATCAA -17,+1

[Fig. 10d]

ABCC11

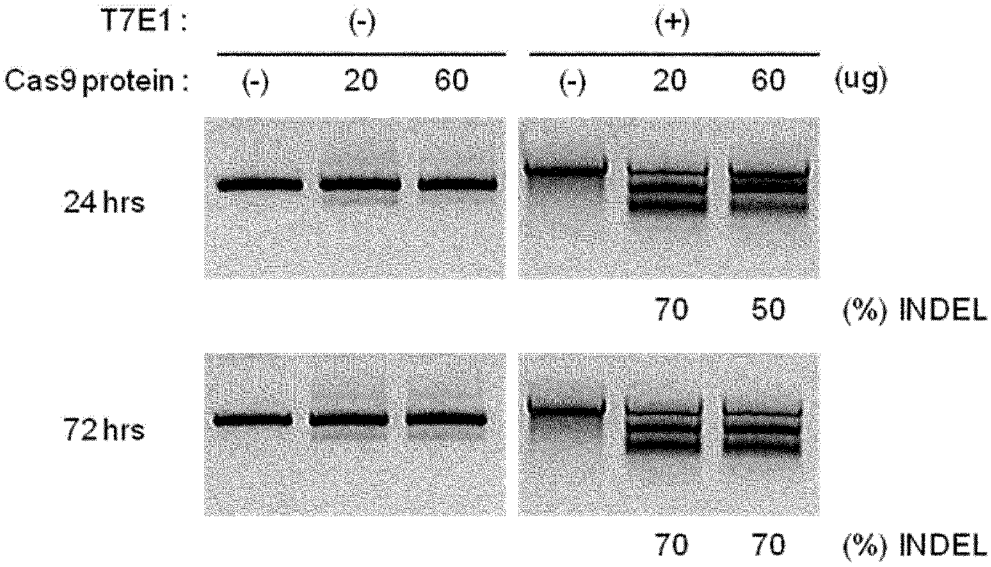


[Fig. 10e]

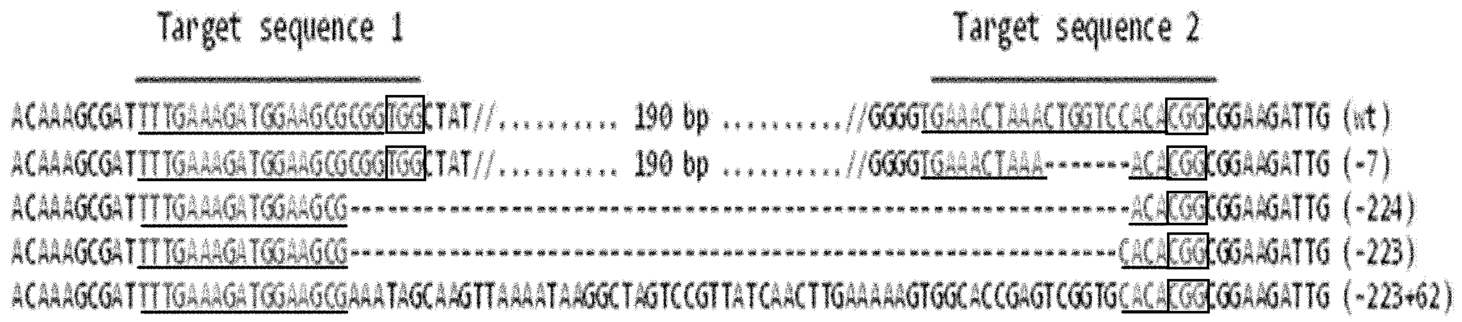
ABCC11

```
TTCTCAAGGCAGCATCATACTTCCCCCACGGTGGGACAGCTGCCCTCCCTGG   WT
TTCTCAAGGCAGCATCATACTTCC------CTGGGACAGCTGCCCTCCCTGG   -6
TTCTCAAGGCAGCATCATACTTC---CACGGTGGGACAGCTGCCCTCCCTGG   -3
TTCTCAAGGCAGC---------------------------TGCCCTCCCTGG  -29
TTCTCAAGGCAGCATCATACTT--------------------CCCTCCCTGG  -20
TTCTCAAGGCAGCATCATACTT--------------------CCCTCCCTGG  -20
TTCTC-----------------------------------------------  -256
```

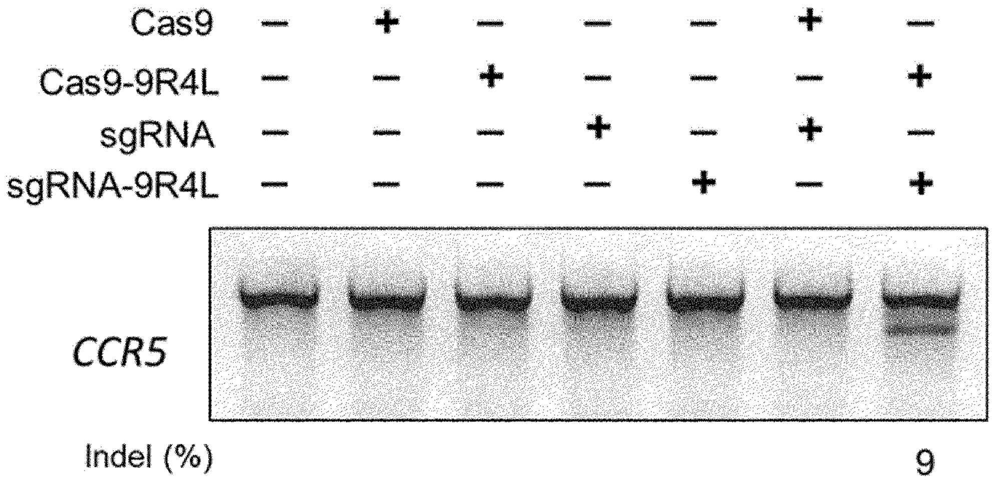
[Fig. 11]



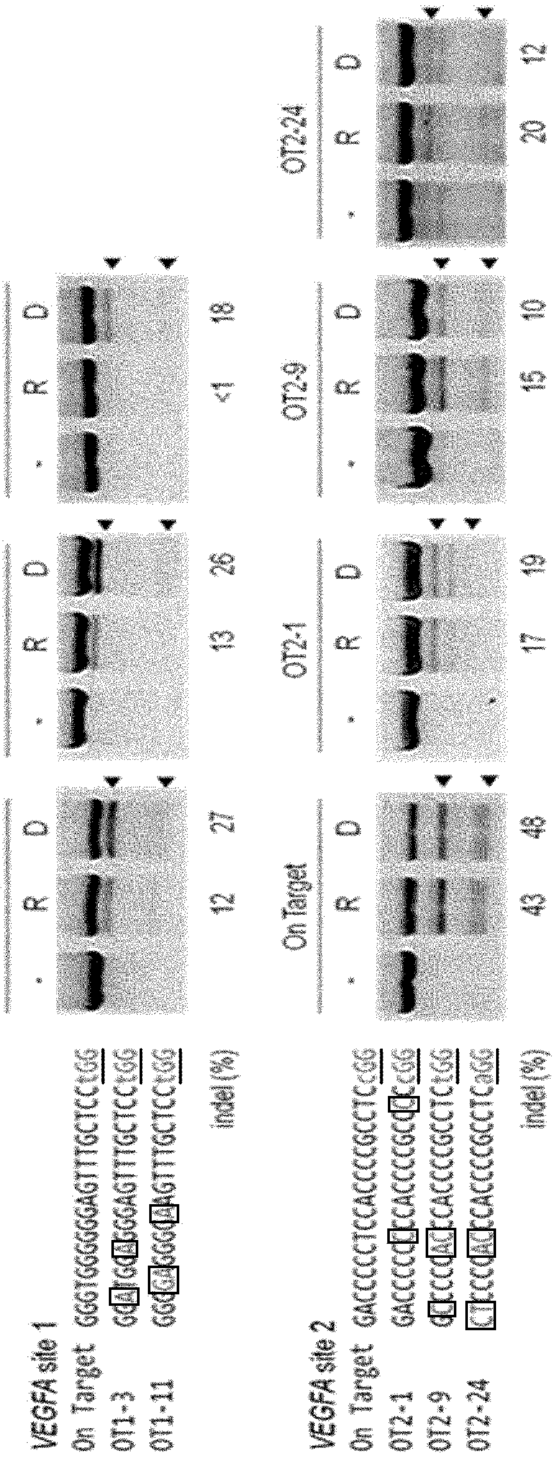
[Fig. 12]


Target sequence 1
Target sequence 2
ACAAAGCGATTTTGAAAGATGGAAGCGCGGTGGCTAT//.......... 190 bp ..../GGGGTGAAACTAAACTGGTCCACACGGCGGAAGATTG (wt)
ACAAAGCGATTTTGAAAGATGGAAGCGCGGTGGCTAT//.......... 190 bp ..../GGGGTGAAACTAAA-------ACACGGCGGAAGATTG (-7)
ACAAAGCGATTTTGAAAGATGGAAGCG--ACACGGCGGAAGATTG (-224)
ACAAAGCGATTTTGAAAGATGGAAGCG---CACACGGCGGAAGATTG (-223)
ACAAAGCGATTTTGAAAGATGGAAGCGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCACACGGCGGAAGATTG (-223+62)

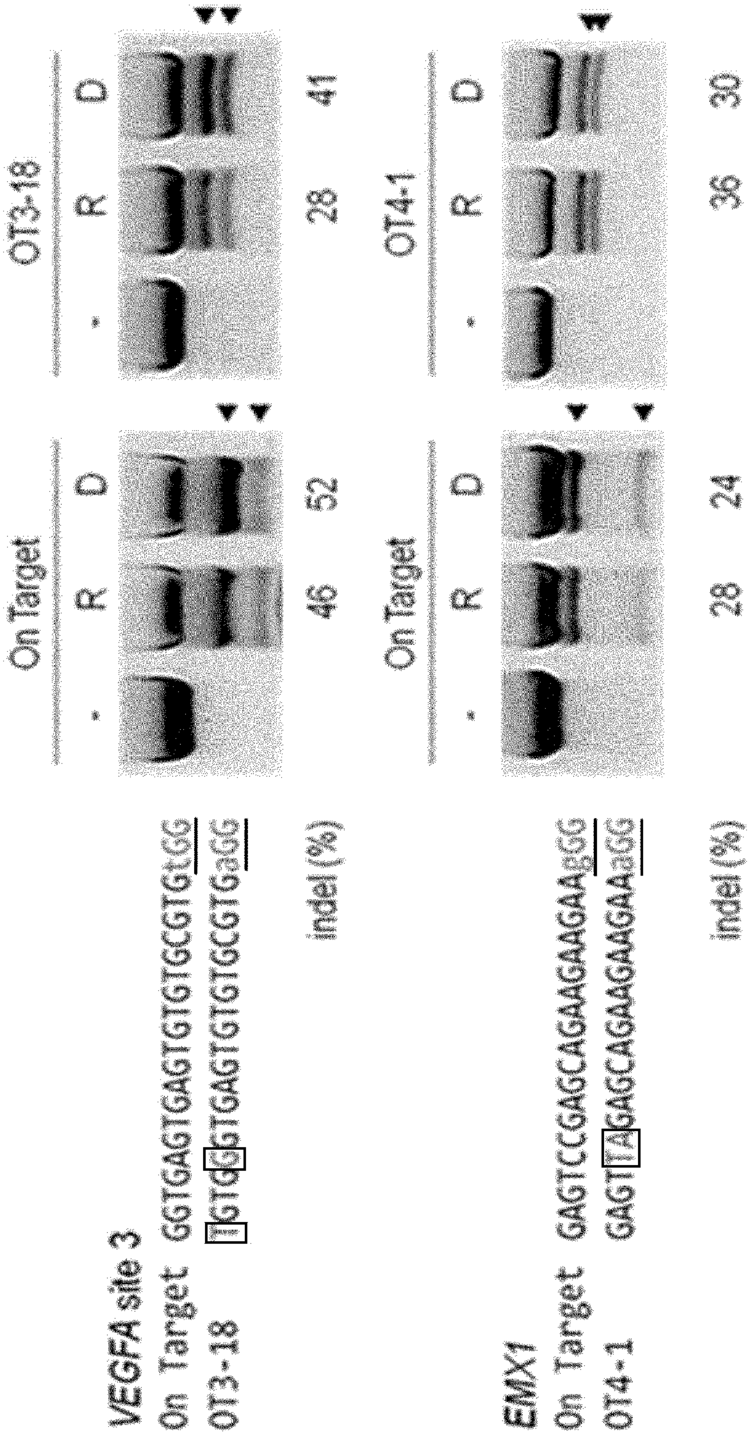
[Fig. 13]



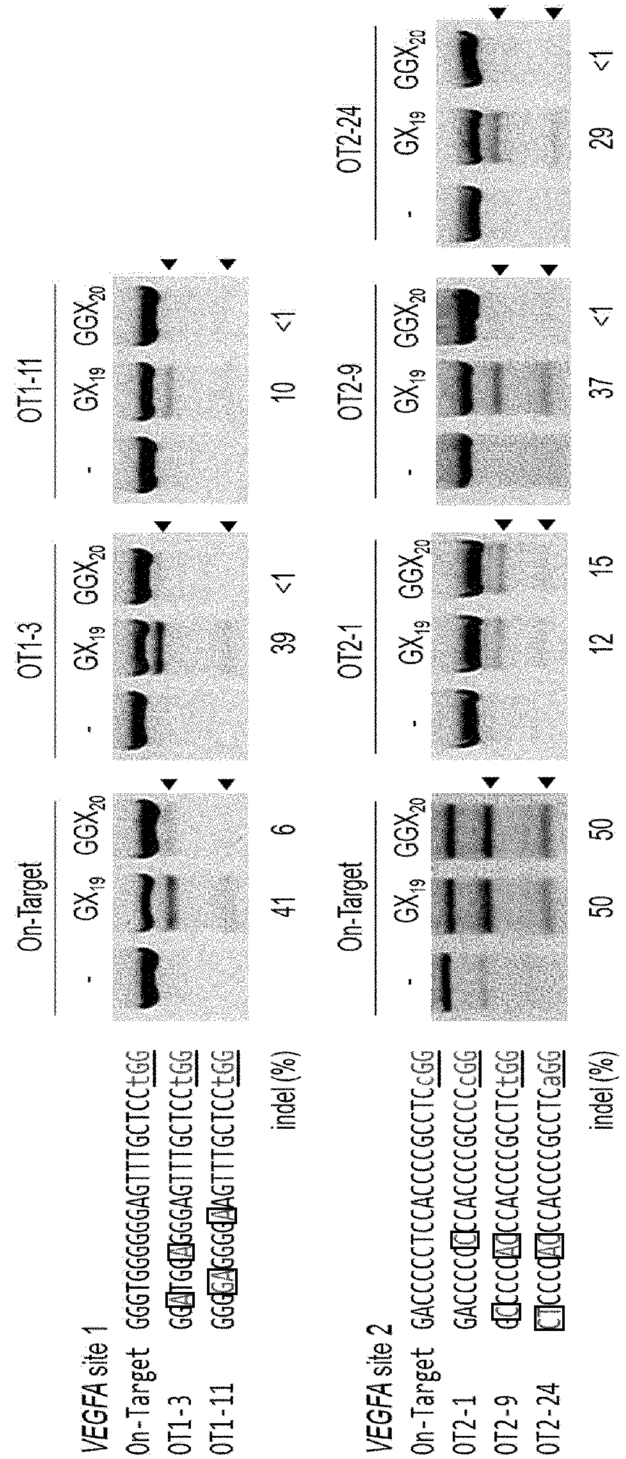
[Fig. 14a]


VEGFA site 1
On Target GGGTGGGGGGAGTTTGCTCCtGG
OT1-3 GGATGGAGGGAGTTTGCTCCtGG
OT1-11 GGGAGGGGGAAGTTTGCTCCtGG
indel (%)
On Target
-
R
D
12
27
OT1-3
-
R
D
13
26
OT1-11
-
R
D
<1
18
VEGFA site 2
On Target GACCCCCTCCACCCCGCCTCcGG
OT2-1 GACCCCCCCCACCCCGCCCCcGG
OT2-9 GCCCCCACCCACCCCGCCTCtGG
OT2-24 CTCCCCACCCACCCCGCCTCaGG
indel (%)
On Target
-
R
D
43
48
OT2-1
-
R
D
17
19
OT2-9
-
R
D
15
10
OT2-24
-
R
D
20
12

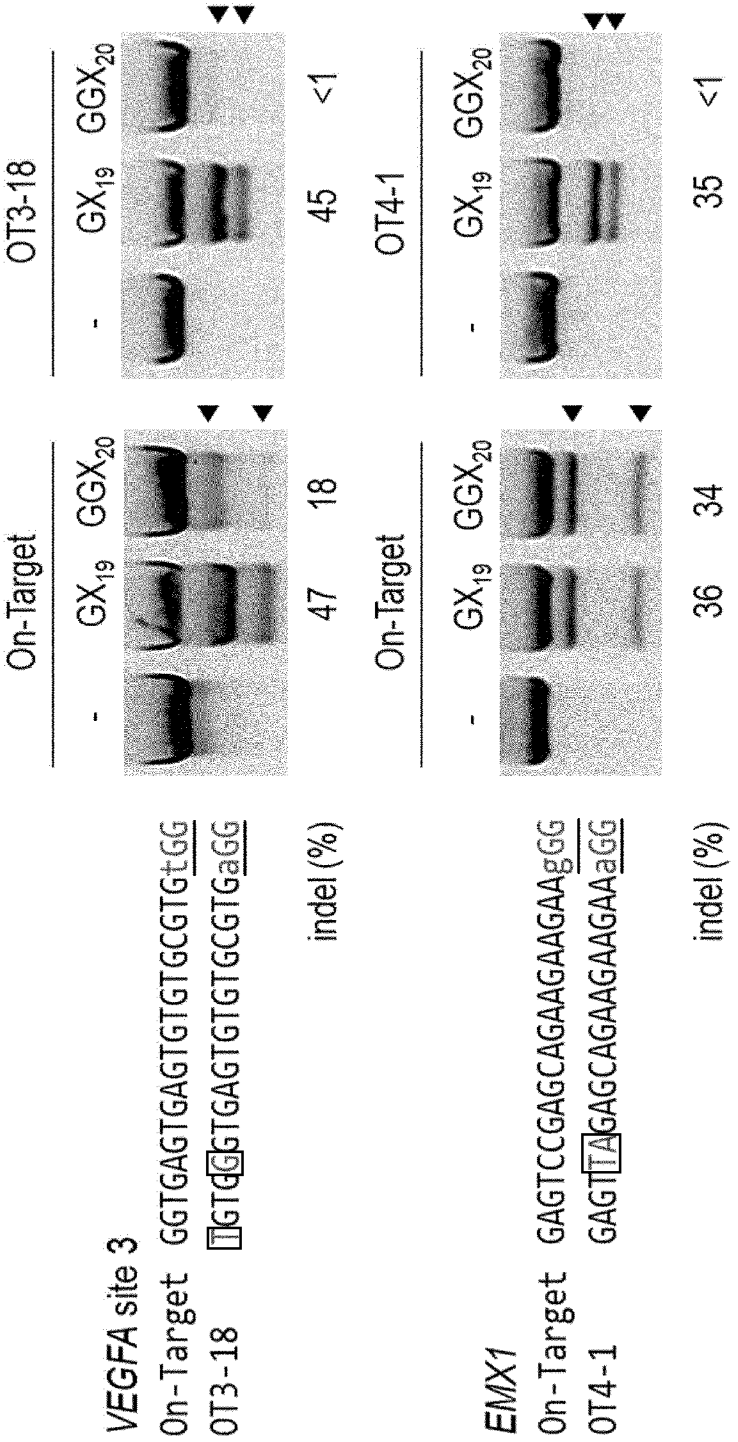
[Fig. 14b]


VEGFA site 3
On Target GGTGAGTGAGTGTGTGCGTGtGG
OT3-18 TGTGGGTGAGTGTGTGCGTGaGG
indel (%)
On Target
-
R
D
46
52
OT3-18
-
R
D
28
41
EMX1
On Target GAGTCCGAGCAGAAGAAGAAgGG
OT4-1 GAGTTAGAGCAGAAGAAGAAaGG
indel (%)
On Target
-
R
D
28
24
OT4-1
-
R
D
36
30

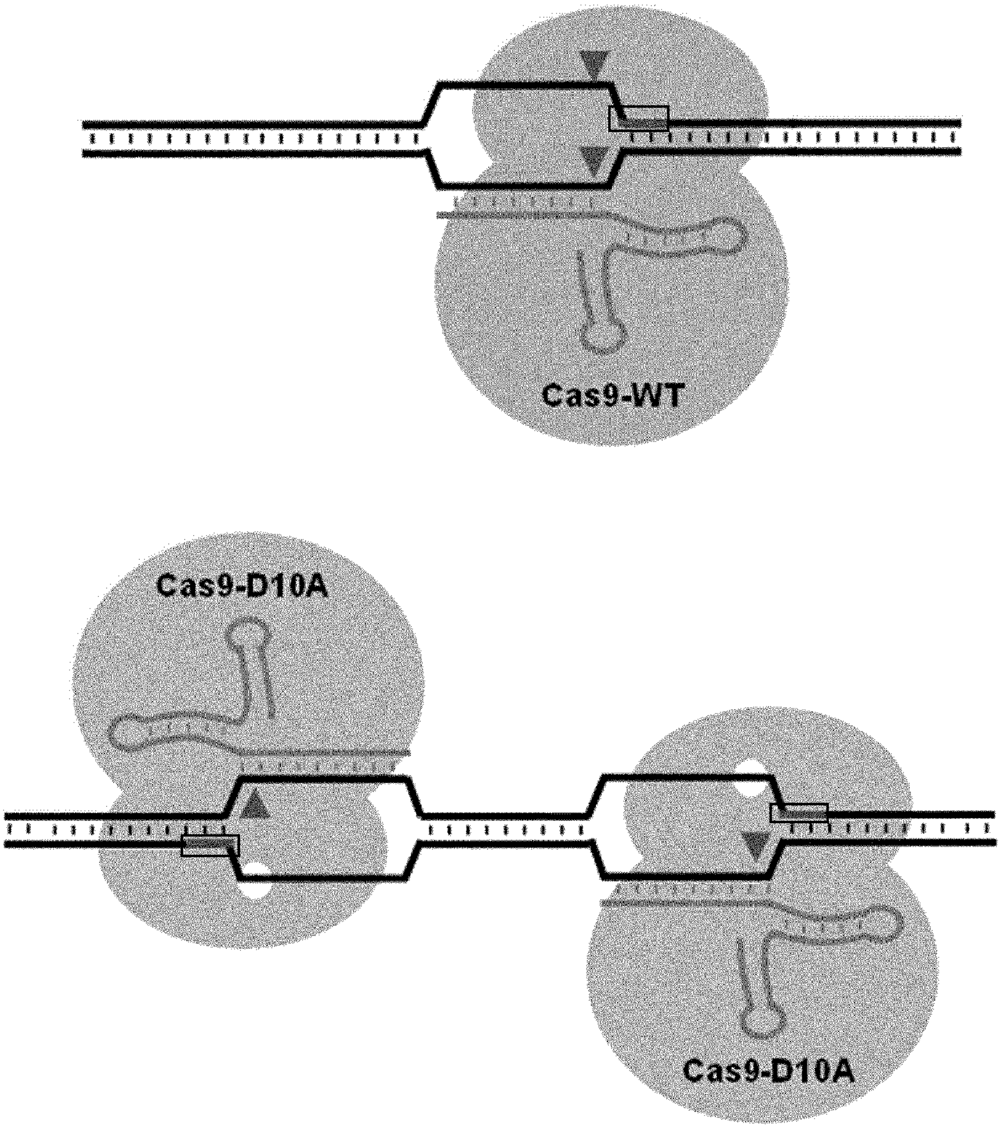
[Fig. 15a]


VEGFA site 1
On-Target GGGTGGGGGGAGTTTGCTCCtGG
OT1-3 GGATGGAGGGAGTTTGCTCCtGG
OT1-11 GGGGAGGGGAAGTTTGCTCCtGG
indel (%)
VEGFA site 2
On-Target GACCCCCTCCACCCCGCCTCcGG
OT2-1 GACCCCCCCCACCCCGCCCCcGG
OT2-9 CCCCCCACCCACCCCGCCTCtGG
OT2-24 CTCCCCACCCACCCCGCCTCaGG
indel (%)
On-Target
-
GX19
GGX20
41
6
OT1-3
39
<1
OT1-11
10
<1
On-Target
50
50
OT2-1
12
15
OT2-9
37
<1
OT2-24
29
<1

[Fig. 15b]

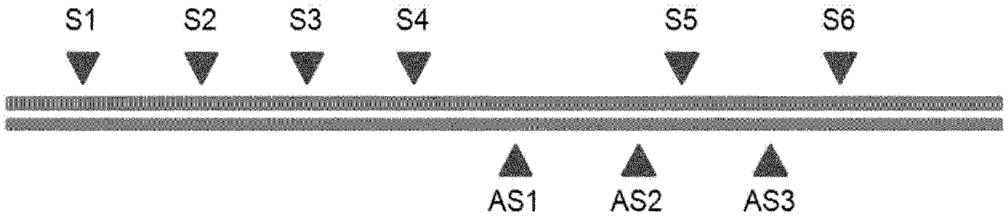

VEGFA site 3
On-Target GGTGAGTGAGTGTGTGCGTGtGG
OT3-18 TGTGGGTGAGTGTGTGCGTGaGG
indel (%)
On-Target
-
GX19
GGX20
47
18
OT3-18
-
GX19
GGX20
45
<1
EMX1
On-Target GAGTCCGAGCAGAAGAAGAAgGG
OT4-1 GAGTTAGAGCAGAAGAAGAAaGG
indel (%)
On-Target
-
GX19
GGX20
36
34
OT4-1
-
GX19
GGX20
35
<1

[Fig. 16a]

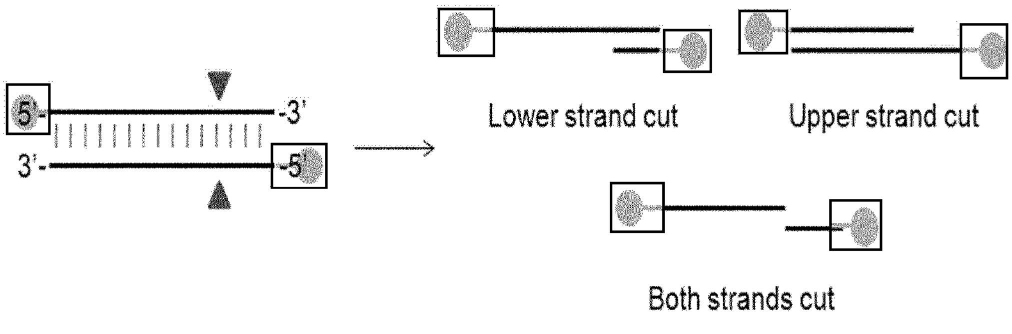


[Fig. 16b]

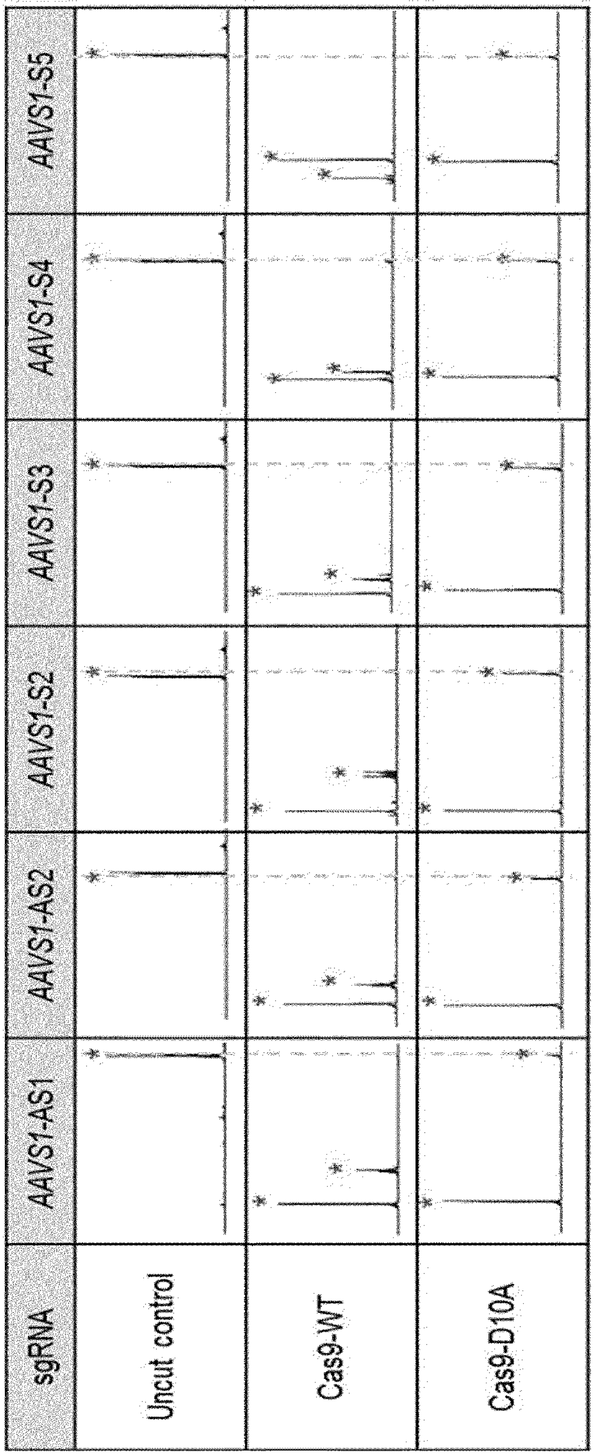
AAVS1 locus



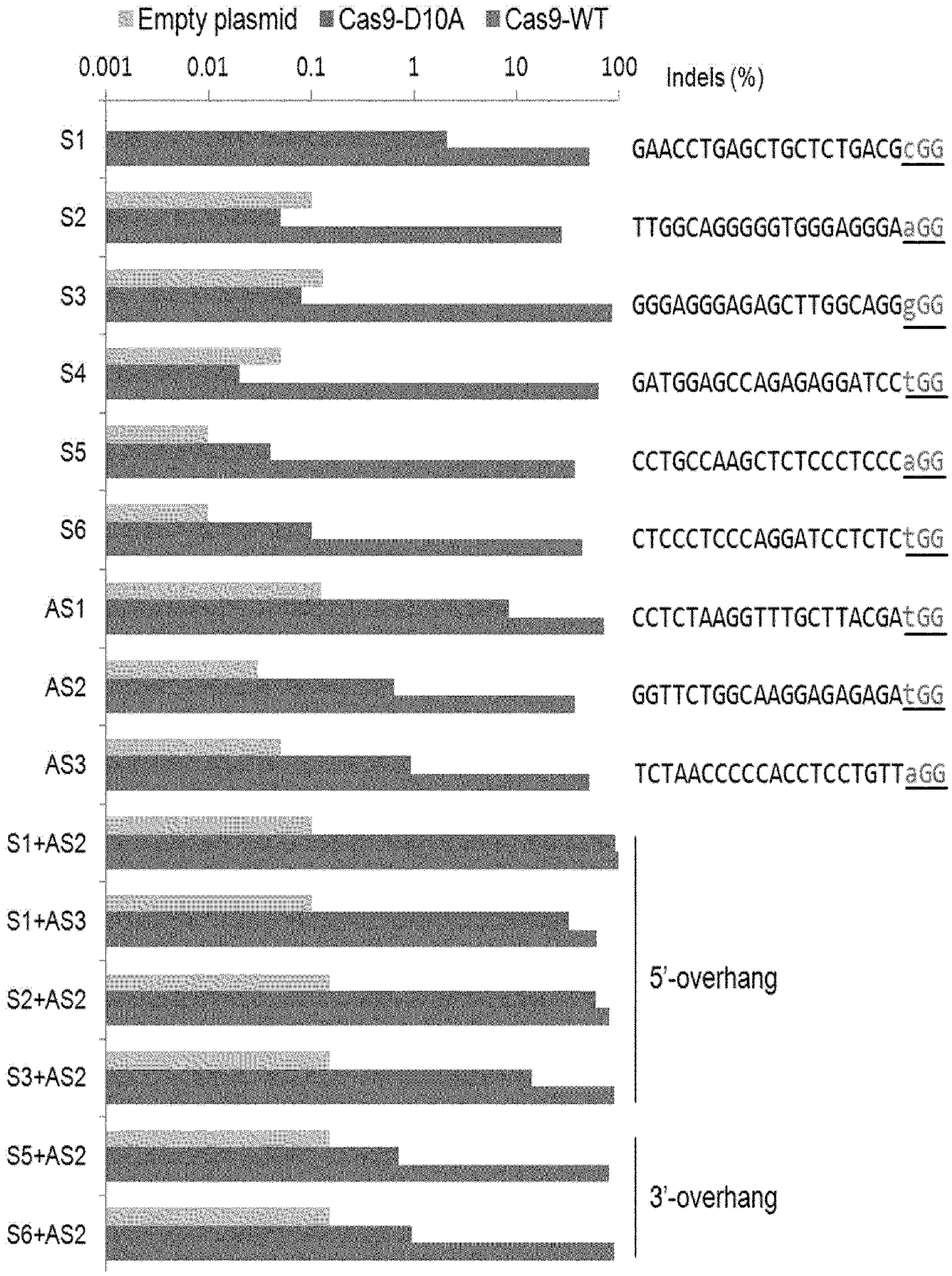
[Fig. 16c]



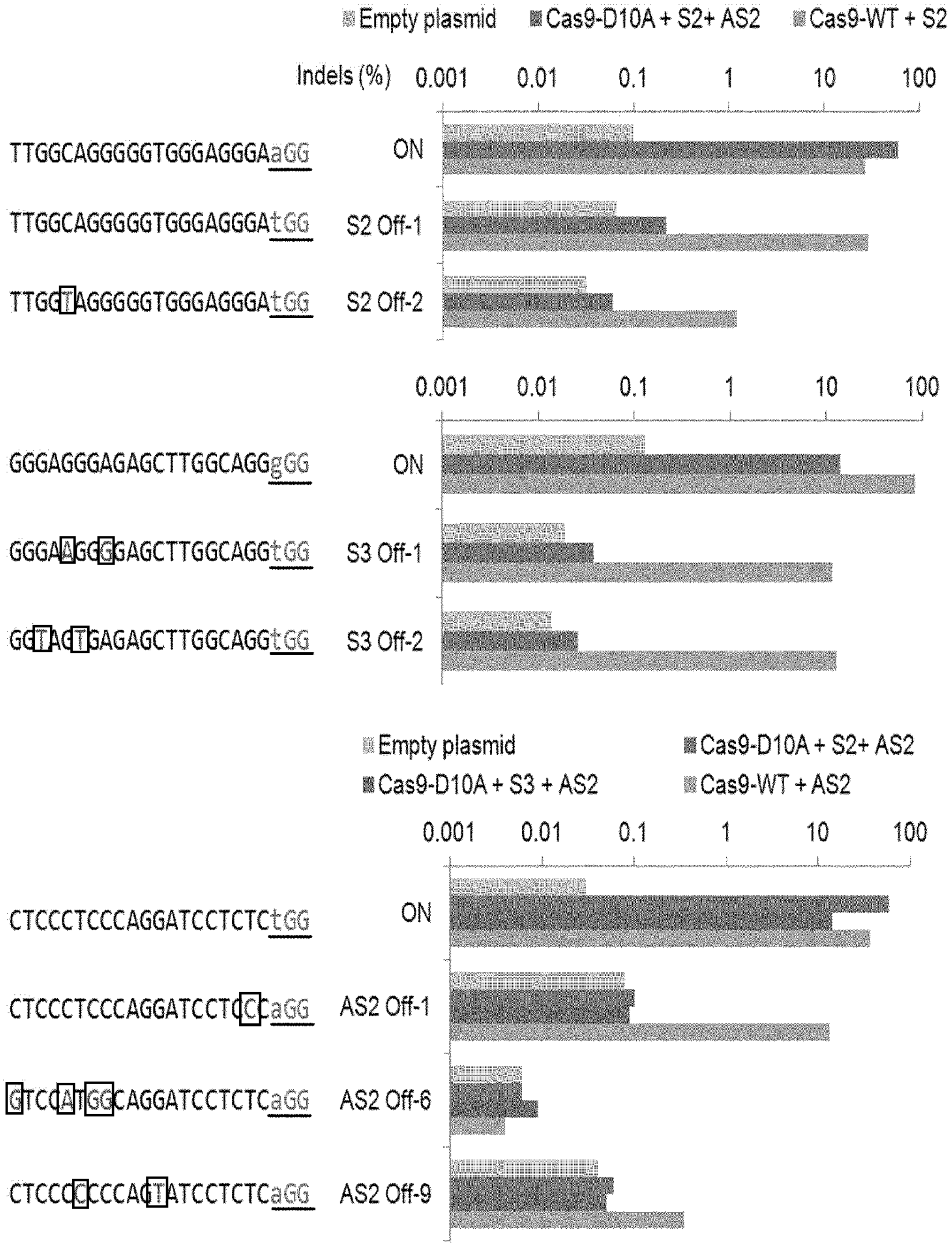
[Fig. 16d]


sgRNA
AAVS1-AS1
AAVS1-AS2
AAVS1-S2
AAVS1-S3
AAVS1-S4
AAVS1-S5
Uncut control
Cas9-WT
Cas9-D10A
*

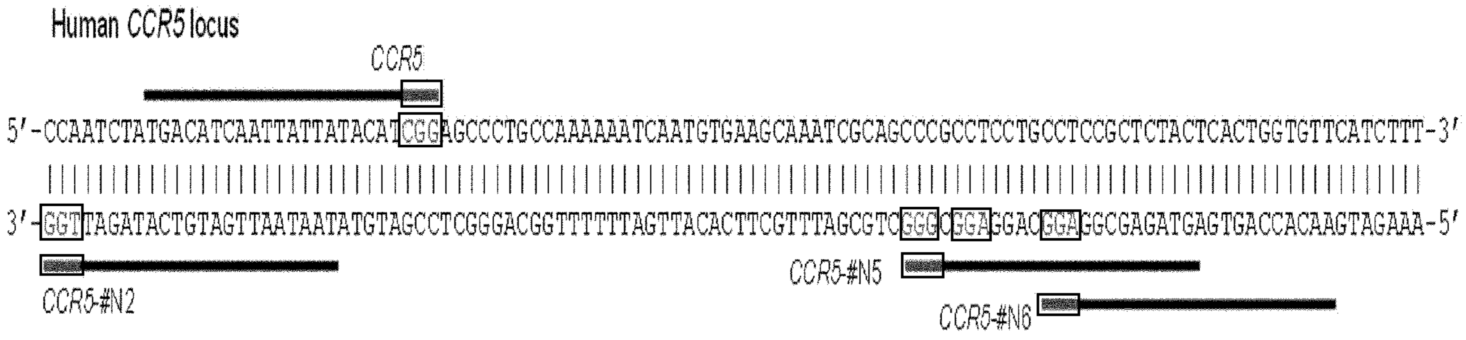
[Fig. 17a]


Empty plasmid
Cas9-D10A
Cas9-WT
0.001
0.01
0.1
1
10
100
Indels (%)
S1
S2
S3
S4
S5
S6
AS1
AS2
AS3
S1+AS2
S1+AS3
S2+AS2
S3+AS2
S5+AS2
S6+AS2
GAACCTGAGCTGCTCTGACGcGG
TTGGCAGGGGGTGGGAGGGAaGG
GGGAGGGAGAGCTTGGCAGGgGG
GATGGAGCCAGAGAGGATCCtGG
CCTGCCAAGCTCTCCCTCCCaGG
CTCCCTCCCAGGATCCTCTCtGG
CCTCTAAGGTTTGCTTACGAtGG
GGTTCTGGCAAGGAGAGAGAtGG
TCTAACCCCCACCTCCTGTTaGG
5'-overhang
3'-overhang

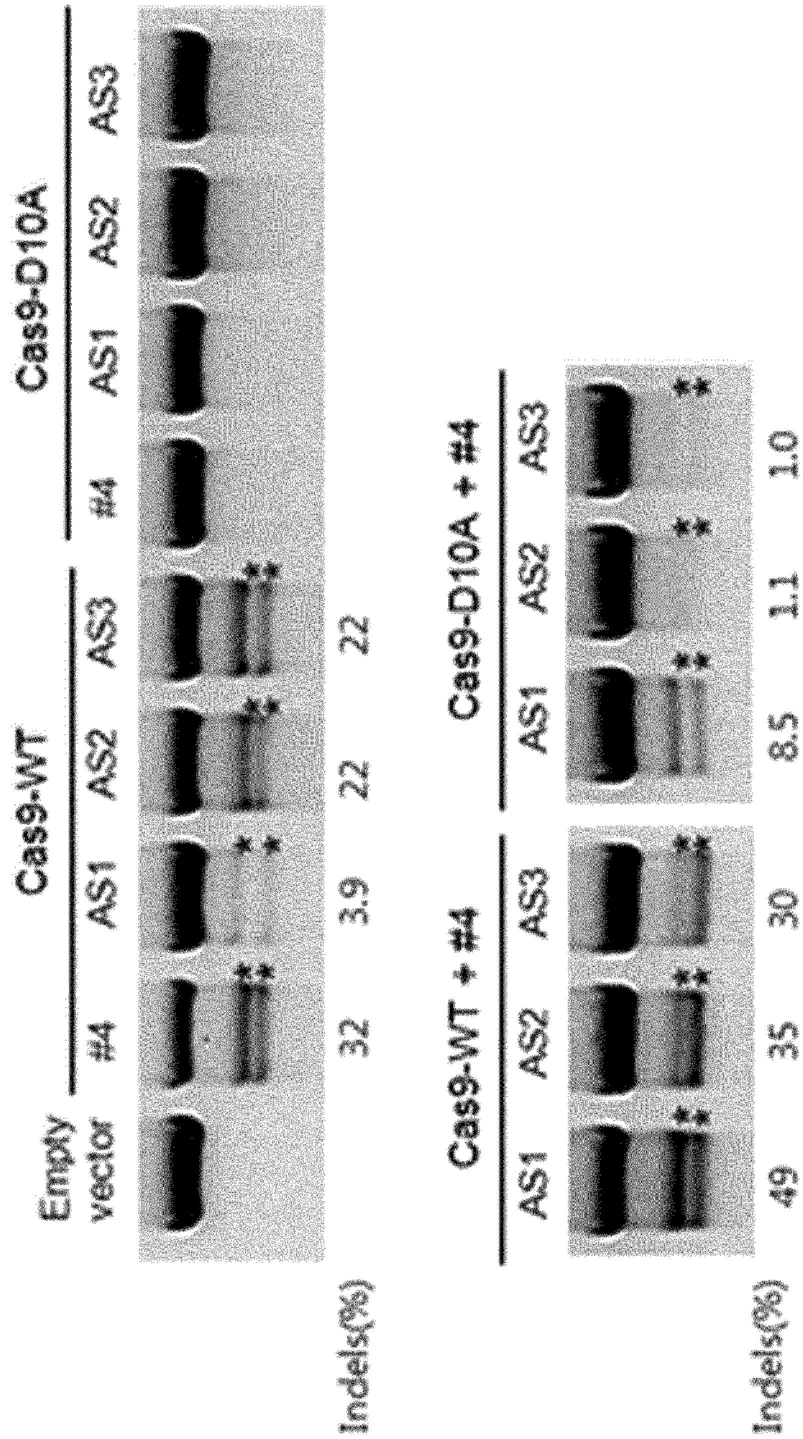
[Fig. 17b]


Empty plasmid
Cas9-D10A + S2+ AS2
Cas9-WT + S2
Indels (%)
0.001
0.01
0.1
1
10
100
TTGGCAGGGGGTGGGAGGGAaGG
ON
TTGGCAGGGGGTGGGAGGGAtGG
S2 Off-1
TTGGTAGGGGGTGGGAGGGAtGG
S2 Off-2
0.001
0.01
0.1
1
10
100
GGGAGGGAGAGCTTGGCAGGgGG
ON
GGGAAGGGGAGCTTGGCAGGtGG
S3 Off-1
GGTAGTGAGAGCTTGGCAGGtGG
S3 Off-2
Empty plasmid
Cas9-D10A + S2+ AS2
Cas9-D10A + S3 + AS2
Cas9-WT + AS2
0.001
0.01
0.1
1
10
100
CTCCCTCCCAGGATCCTCTCtGG
ON
CTCCCTCCCAGGATCCTCCCaGG
AS2 Off-1
GTCCATGGCAGGATCCTCTCaGG
AS2 Off-6
CTCCCCCCCAGTATCCTCTCaGG
AS2 Off-9

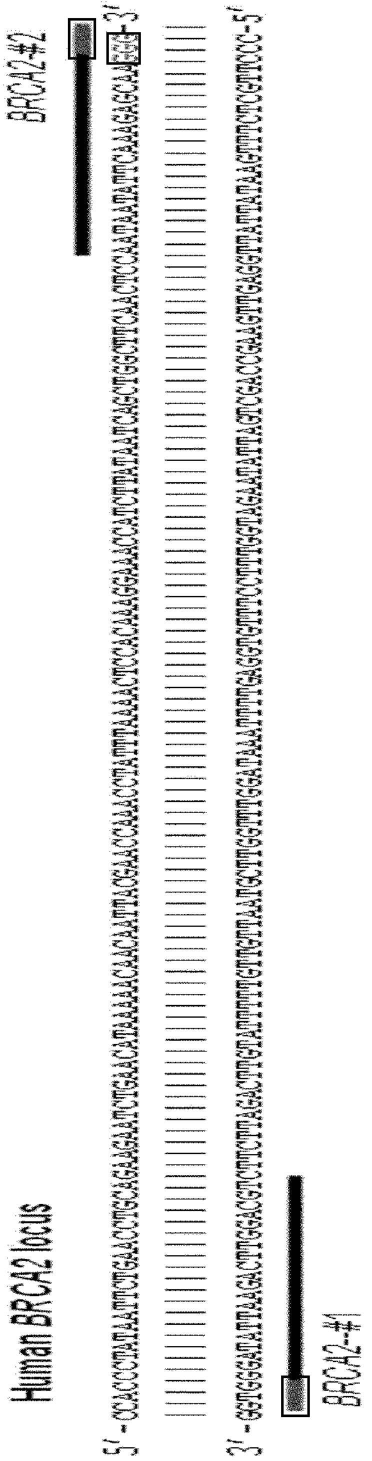
[Fig. 18a]


Human CCR5 locus
CCR5
5′-CCAATCTATGACATCAATTATTATACATCGGAGCCCTGCCAAAAAATCAATGTGAAGCAAATCGCAGCCCGCCTCCTGCCTCCGCTCTACTCACTGGTGTTCATCTTT-3′
3′-GGTTAGATACTGTAGTTAATAATATGTAGCCTCGGGACGGTTTTTTAGTTACACTTCGTTTAGCGTCGGGCGGAGGACGGAGGCGAGATGAGTGACCACAAGTAGAAA-5′
CCR5-#N2
CCR5-#N5
CCR5-#N6

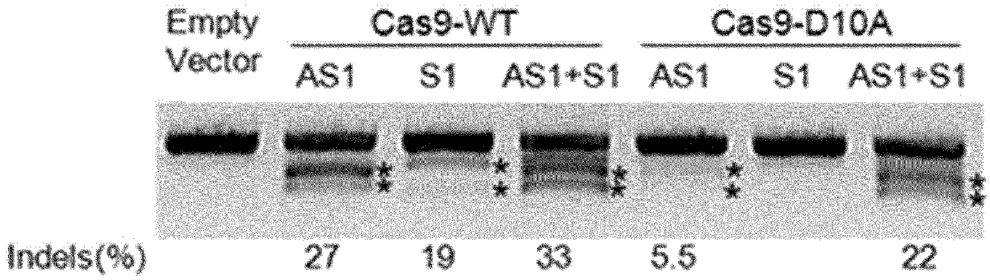
[Fig. 18b]


Empty vector
Cas9-WT
#4
AS1
AS2
AS3
Cas9-D10A
#4
AS1
AS2
AS3
Indels(%)
32
3.9
22
22
Cas9-WT + #4
AS1
AS2
AS3
Cas9-D10A + #4
AS1
AS2
AS3
Indels(%)
49
35
30
8.5
1.1
1.0

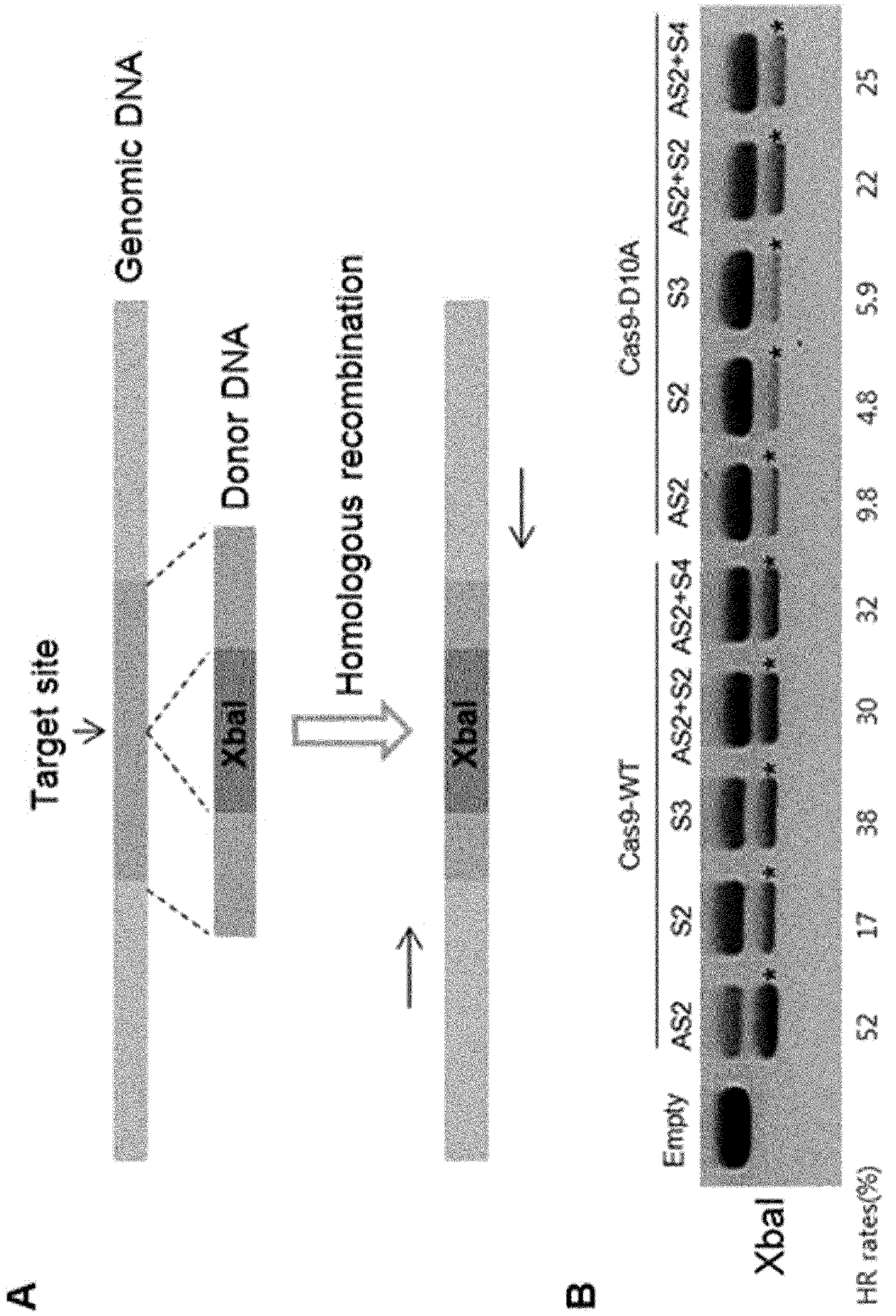
[Fig. 18c]



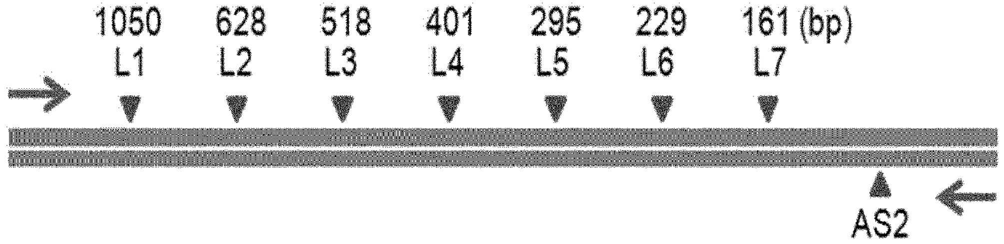
[Fig. 18d]



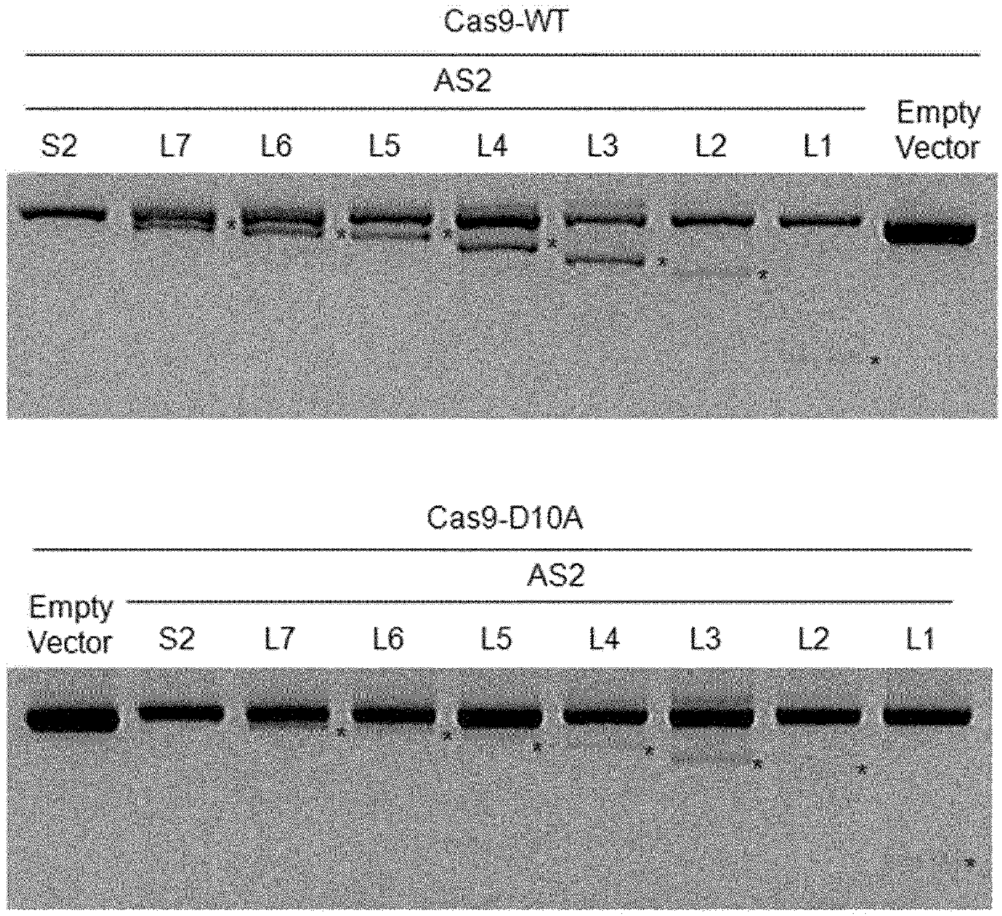
[Fig. 19]



[Fig. 20a]



[Fig. 20b]


Cas9-WT
AS2
S2
L7
L6
L5
L4
L3
L2
L1
Empty Vector
*
Cas9-D10A
AS2
Empty Vector
S2
L7
L6
L5
L4
L3
L2
L1
*

[Fig. 20c]

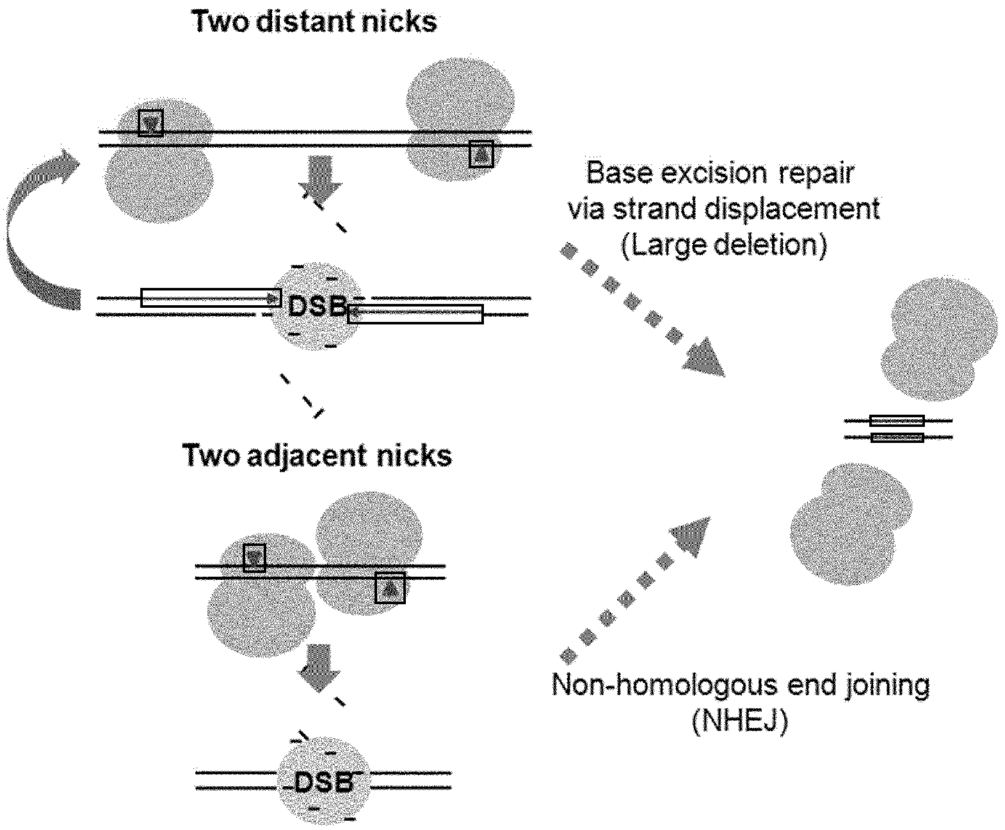
Cas9-WT AS2 + L1 (~1050bp deletion)

```
ggccgggaatcaagagtcaCCCAGAGACAGTGACCAACCATCcctgttt…//…agctCTCCCTCCCAGGATCCTCTCTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg WT
ggccgggaatcaagagtcaCCCAG------TGACCAACCATCcct------------------------------------------gtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCCAG--------------------------------------------------------------------------------------------------gaa
ggccgggaatcaagagtcaCCCAGA----------------------------------------------CCTCTCTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCC-------------------------------------------------------------------------------------------------taacag
ggccgggaatcaaga------------------------------------------------------------cgTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggc------------------------------------------------------------------------------tccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCCAGA-----------------------------------------------CTCTCTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg
```

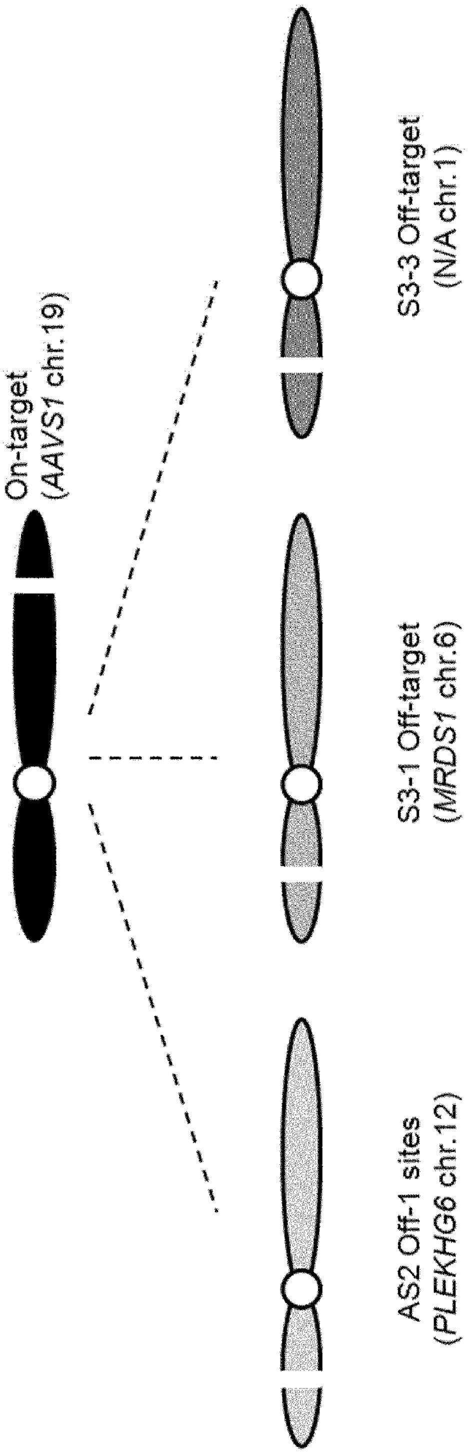
Cas9-D10A AS2 + L1 (~1050bp deletion)

```
ggccgggaatcaagagtcaCCCAGAGACAGTGACCAACCATCcctgttt…//…agctCTCCCTCCCAGGATCCTCTCTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg WT
ggccgggaatcaagagtcgCCCAGAGACAGTGACCAACCATC---------------------------------------------gtaagcaaaccttagaggttctggcaaggagagagatg
gg-------------------------------------------------------------------------------tccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCCA--------------------------------------------------------tctccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCCAGA-------------------------------------------------CTCTGGctccatcgtaagcaaaccttagaggttctggcaaggagagagatg
gg-------------------------------------------------------------------------------tccatcgtaagcaaaccttagaggttctggcaaggagagagatg
ggccgggaatcaagagtcaCCCAGAGACAGTGACCAACCATCcc---------------------------------------------------------------------------atatca
ggccgggaatcaagagtca------------------------------------------------------------------tcgtaagcaaaccttagaggttctggcaaggagagagatg x2
```

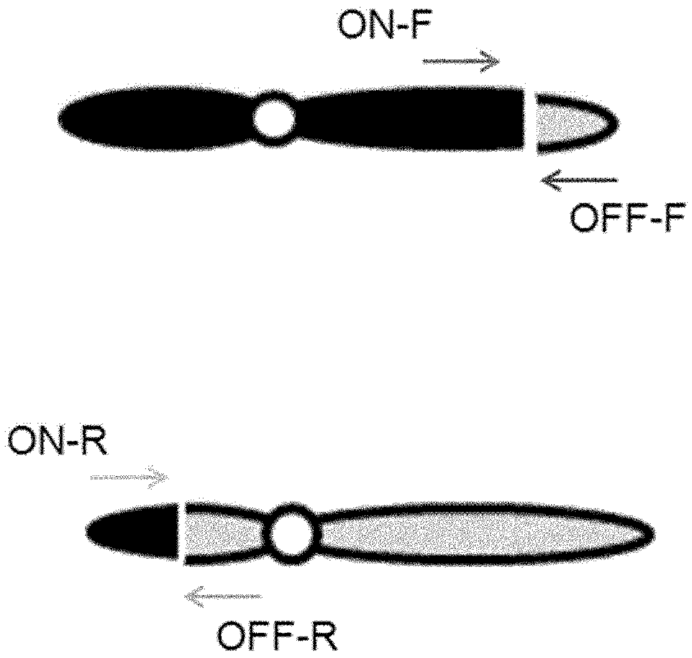
[Fig. 20d]


Two distant nicks
DSB
Base excision repair
via strand displacement
(Large deletion)
Two adjacent nicks
DSB
Non-homologous end joining
(NHEJ)

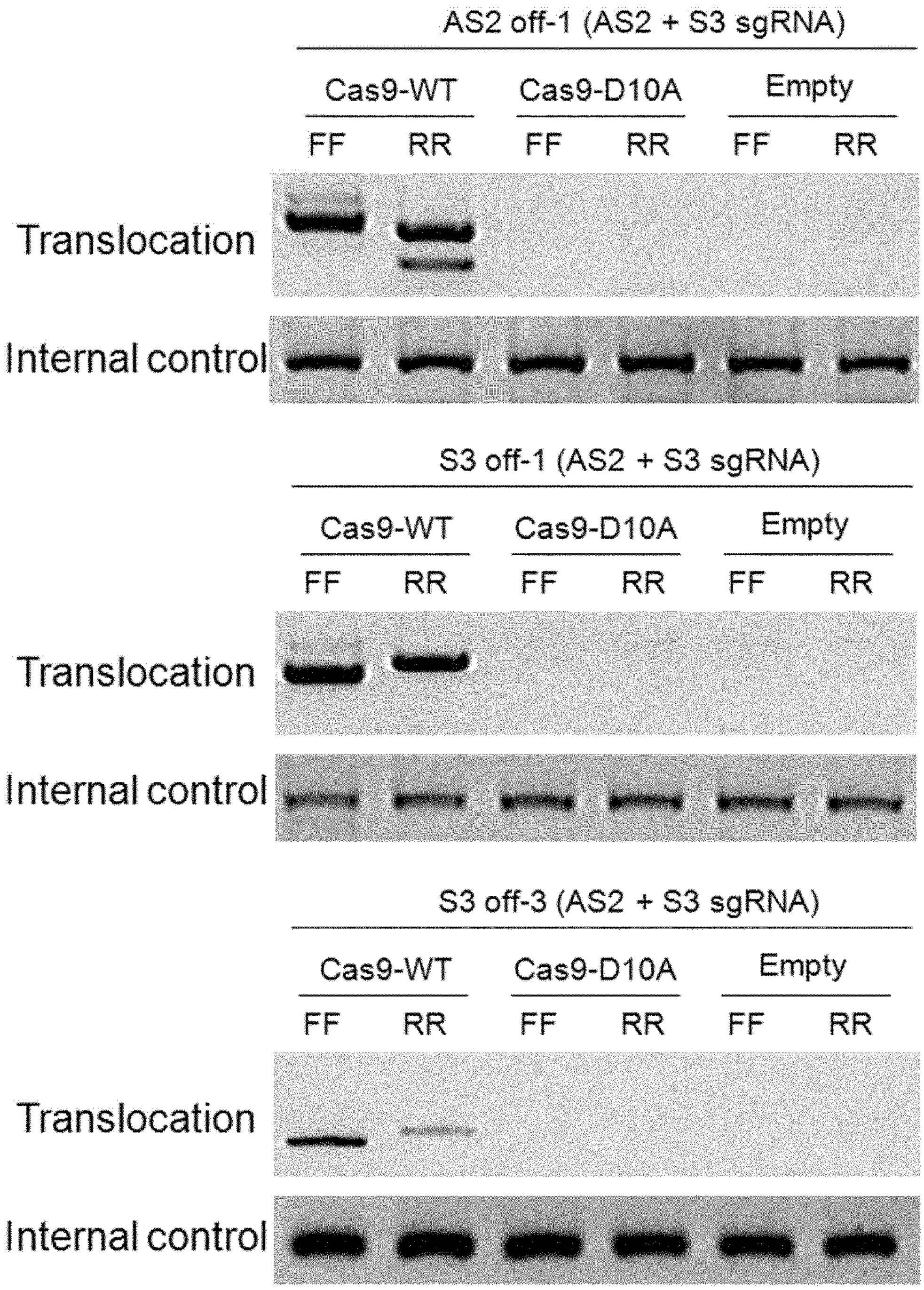
[Fig. 21a]


On-target
(*AAVS1* chr.19)
S3-3 Off-target
(N/A chr.1)
S3-1 Off-target
(*MRDS1* chr.6)
AS2 Off-1 sites
(*PLEKHG6* chr.12)

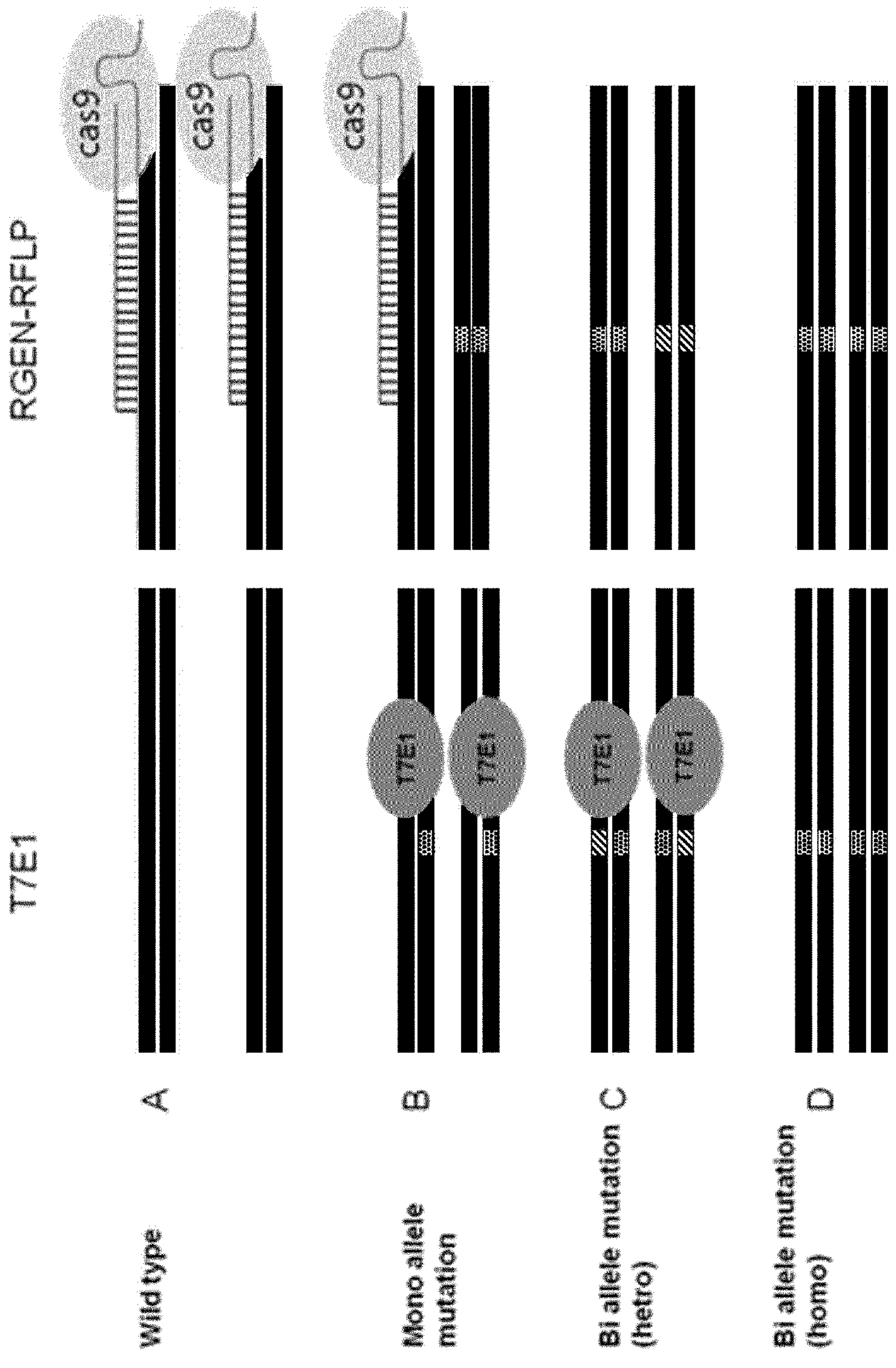
[Fig. 21b]



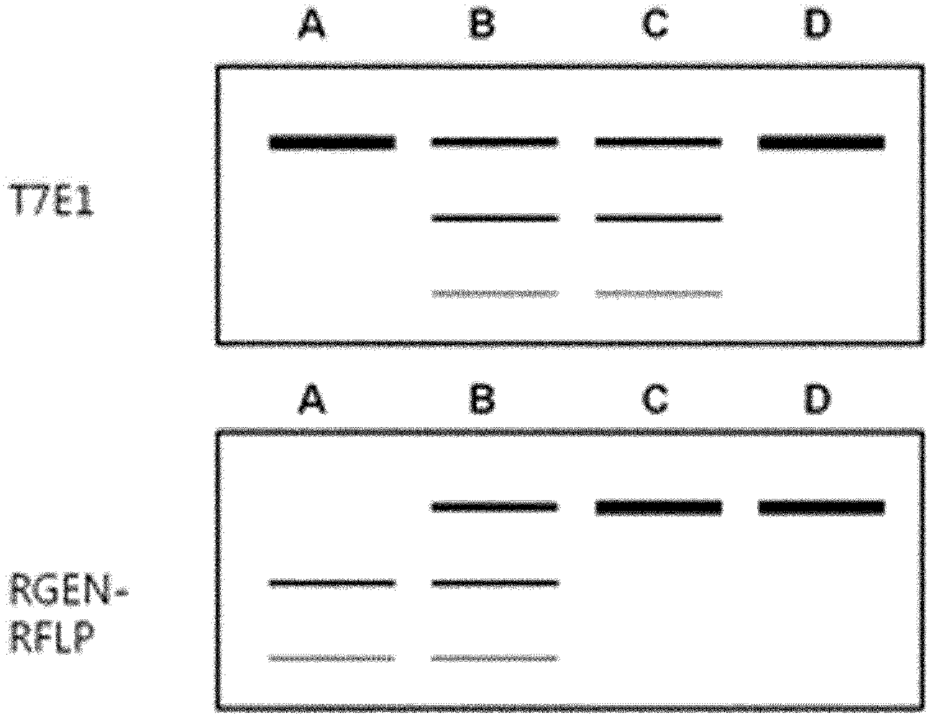
[Fig. 21c]


AS2 off-1 (AS2 + S3 sgRNA)
Cas9-WT
Cas9-D10A
Empty
FF
RR
FF
RR
FF
RR
Translocation
Internal control
S3 off-1 (AS2 + S3 sgRNA)
Cas9-WT
Cas9-D10A
Empty
FF
RR
FF
RR
FF
RR
Translocation
Internal control
S3 off-3 (AS2 + S3 sgRNA)
Cas9-WT
Cas9-D10A
Empty
FF
RR
FF
RR
FF
RR
Translocation
Internal control

[Fig. 22a]


RGEN-RFLP
T7E1
cas9
cas9
cas9
T7E1
T7E1
T7E1
T7E1
Wild type
A
Mono allele mutation
B
Bi allele mutation (hetro)
C
Bi allele mutation (homo)
D

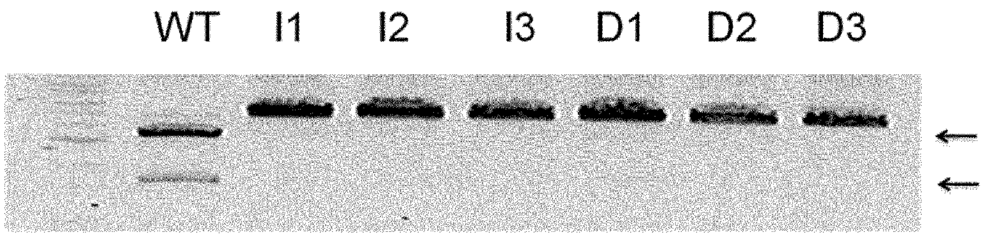
[Fig. 22b]



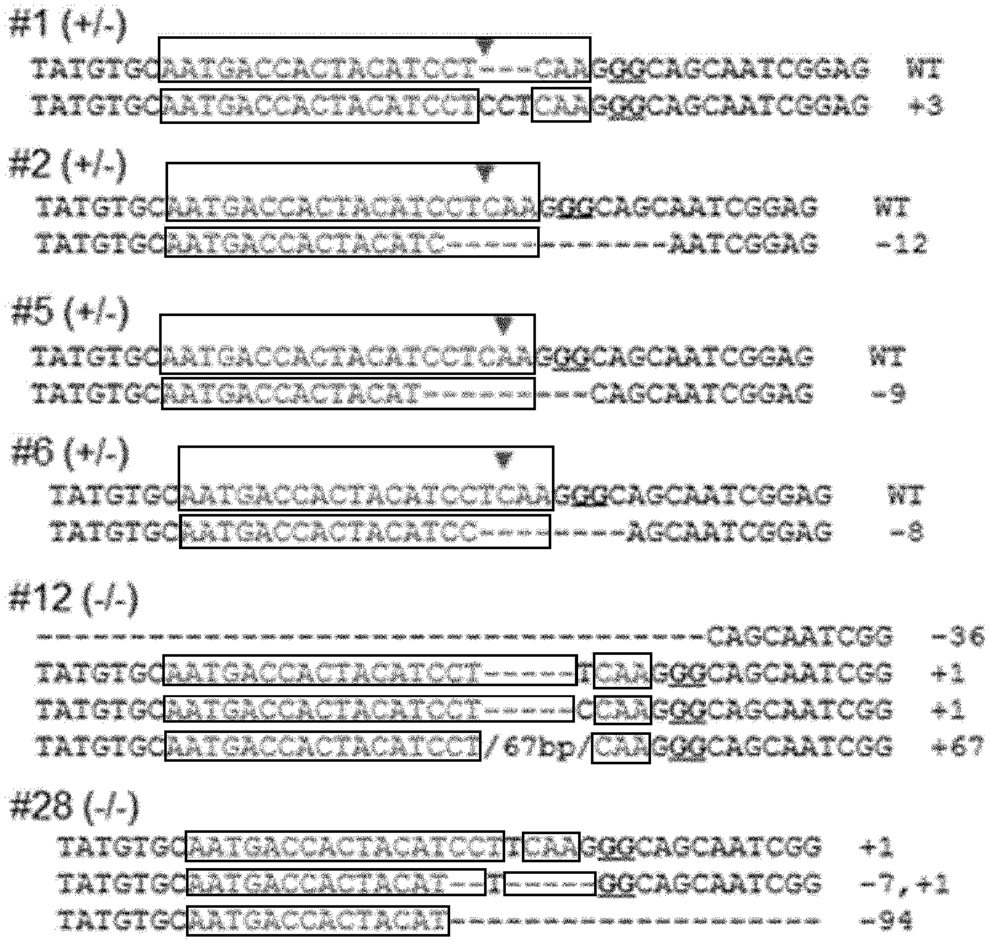
[Fig. 23]

Plasmid target sequence

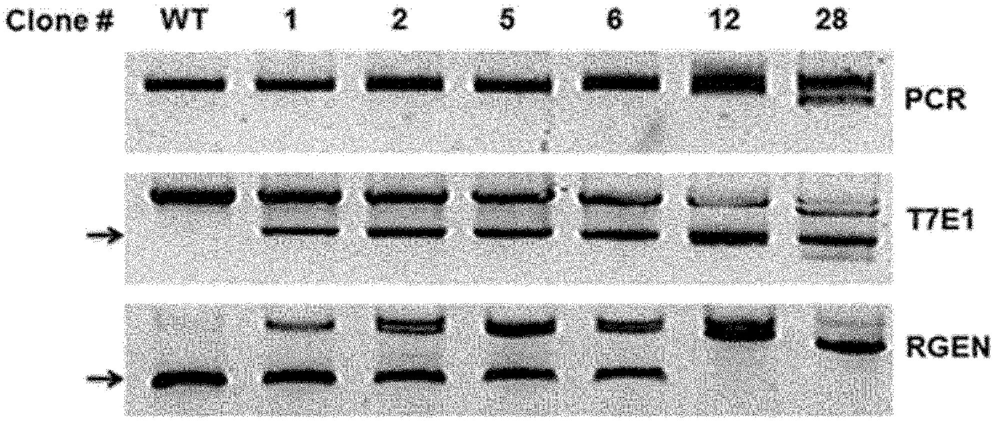
```
AATGACCACTACATCCT---CAAGGG   WT
AATGACCACTACATCCTT--CAAGGG   I1
AATGACCACTACATCCTTT-CAAGGG   I2
AATGACCACTACATCCTTTTCAAGGG   I3
AATGACCACTACATCCT----AAGGG   D1
AATGACCACTACATCCT-----AGGG   D2
AATGACCACTACATCCT------GGG   D3
```



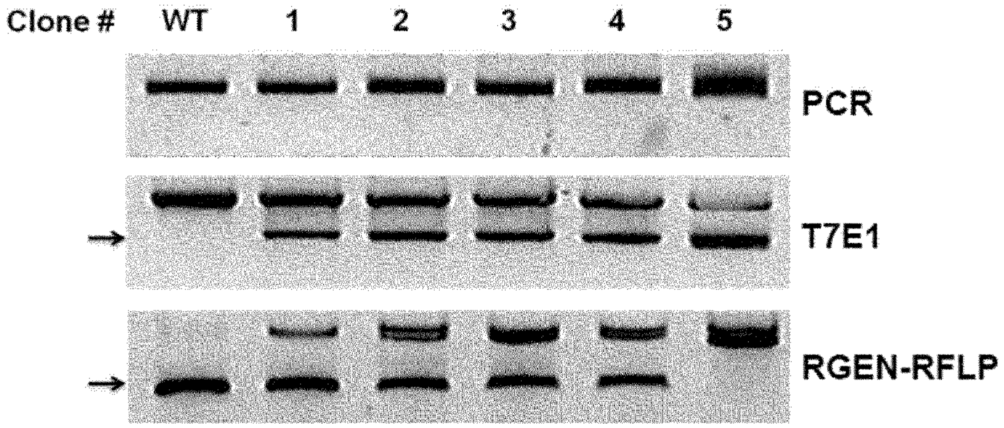
[Fig. 24a]



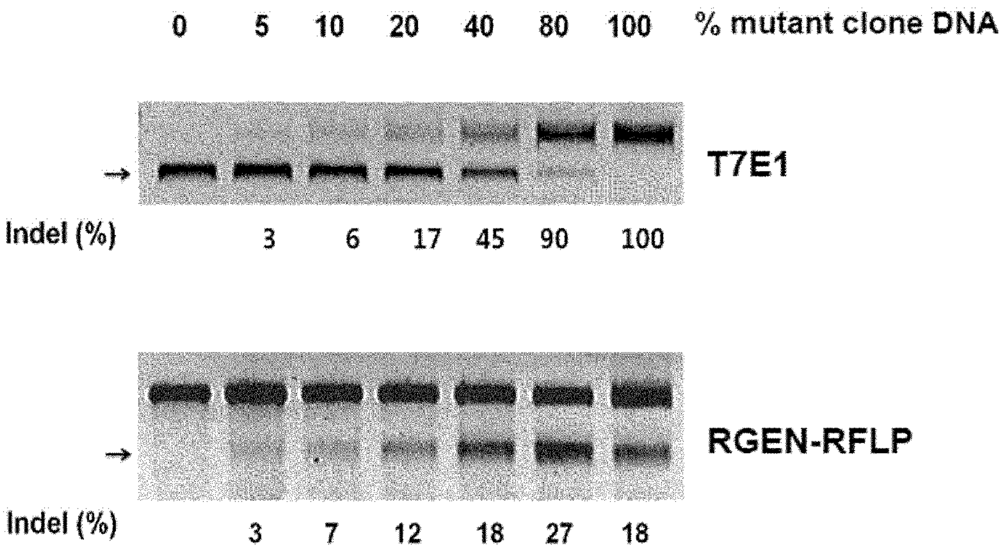
[Fig. 24b]



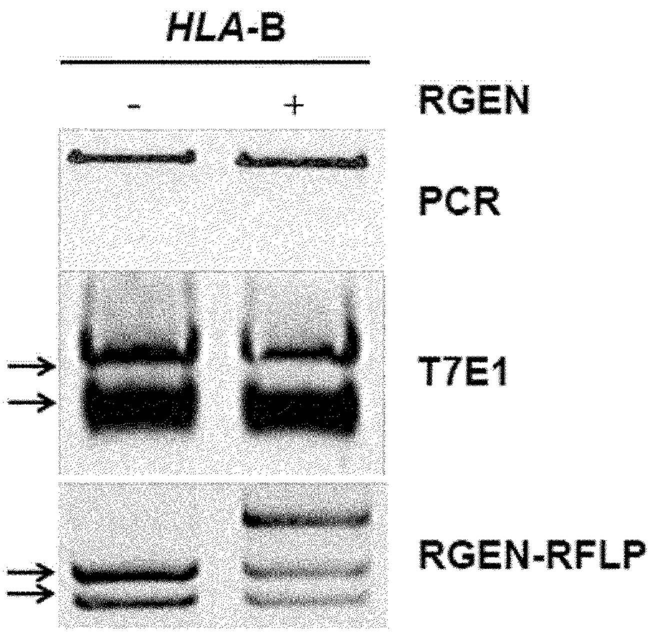
[Fig. 25a]

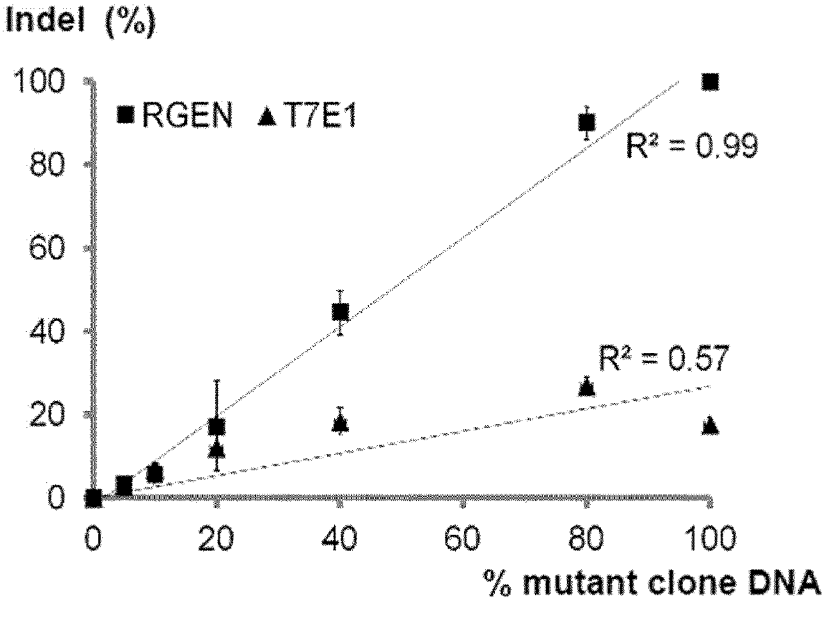


[Fig. 25b]

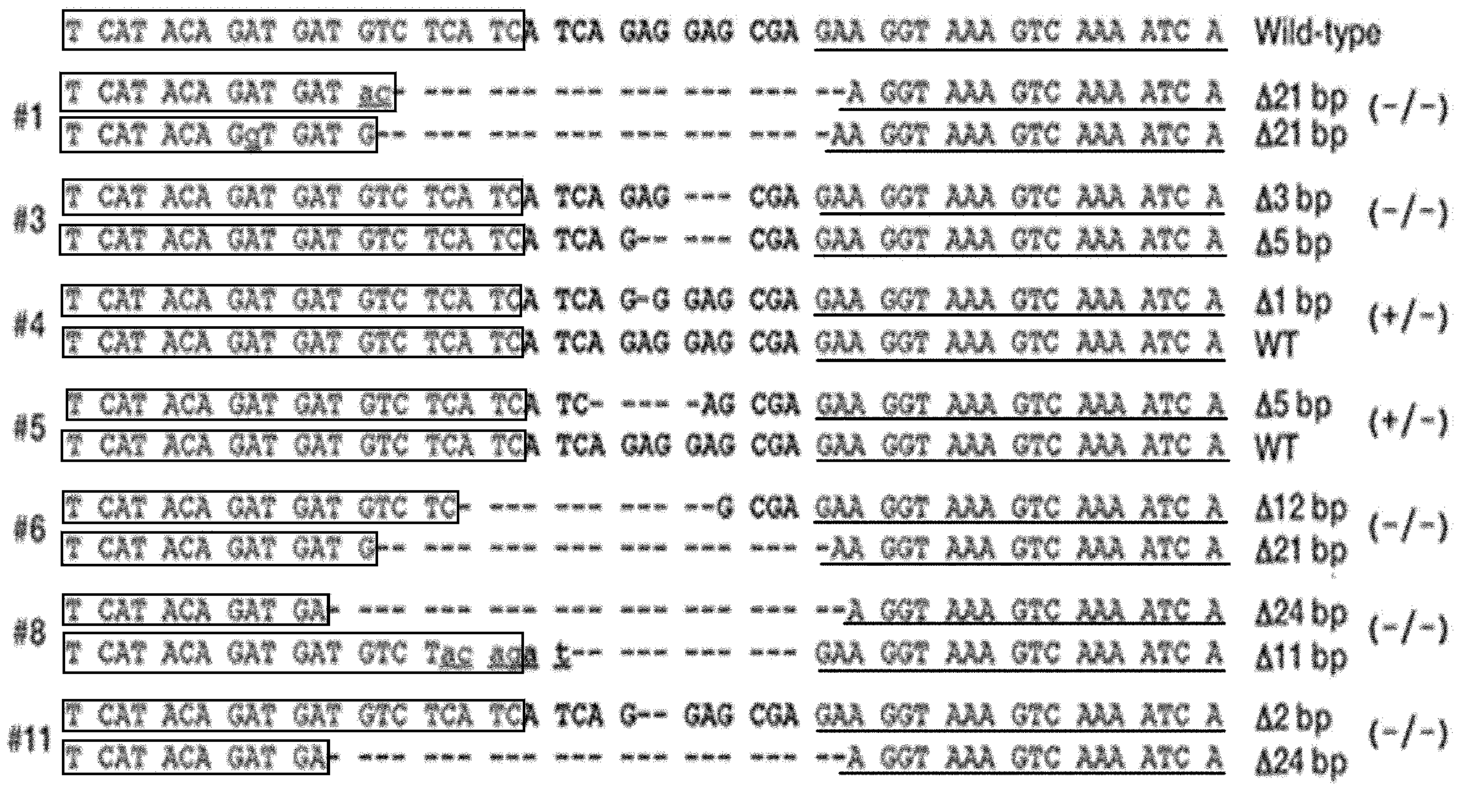


[Fig. 25c]

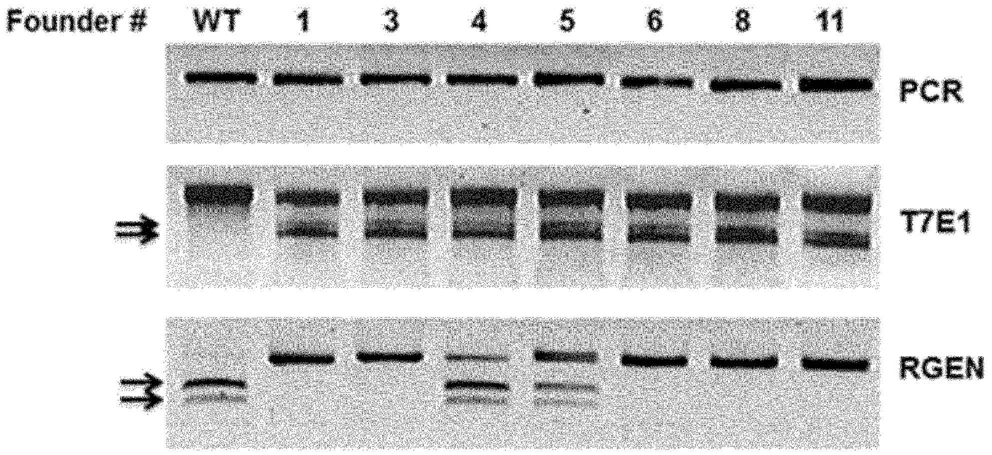




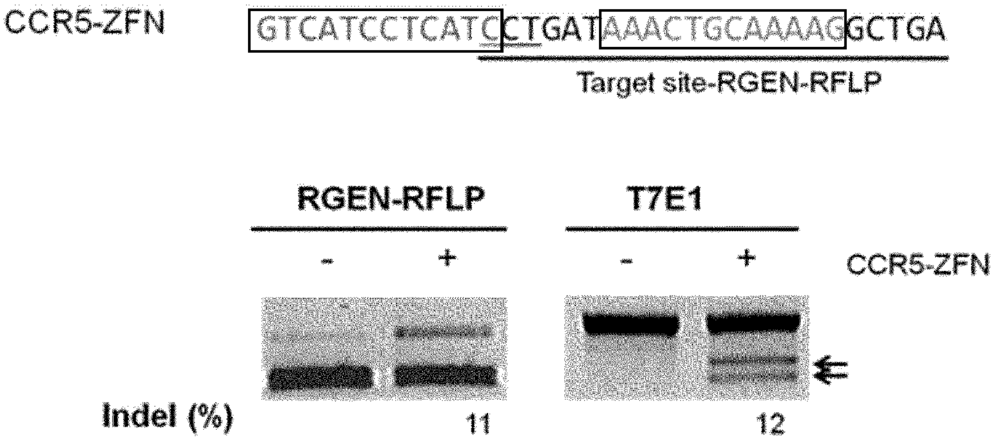
[Fig. 26a]


T CAT ACA GAT GAT GTC TCA TCA TCA GAG GAG CGA GAA GGT AAA GTC AAA ATC A Wild-type
#1
T CAT ACA GAT GAT ac- --- --- --- --- --- --- --A GGT AAA GTC AAA ATC A Δ21 bp (−/−)
T CAT ACA GgT GAT G-- --- --- --- --- --- --- -AA GGT AAA GTC AAA ATC A Δ21 bp
#3
T CAT ACA GAT GAT GTC TCA TCA TCA GAG --- CGA GAA GGT AAA GTC AAA ATC A Δ3 bp (−/−)
T CAT ACA GAT GAT GTC TCA TCA TCA G-- --- CGA GAA GGT AAA GTC AAA ATC A Δ5 bp
#4
T CAT ACA GAT GAT GTC TCA TCA TCA G-G GAG CGA GAA GGT AAA GTC AAA ATC A Δ1 bp (+/−)
T CAT ACA GAT GAT GTC TCA TCA TCA GAG GAG CGA GAA GGT AAA GTC AAA ATC A WT
#5
T CAT ACA GAT GAT GTC TCA TCA TC- --- -AG CGA GAA GGT AAA GTC AAA ATC A Δ5 bp (+/−)
T CAT ACA GAT GAT GTC TCA TCA TCA GAG GAG CGA GAA GGT AAA GTC AAA ATC A WT
#6
T CAT ACA GAT GAT GTC TC- --- --- --- --G CGA GAA GGT AAA GTC AAA ATC A Δ12 bp (−/−)
T CAT ACA GAT GAT G-- --- --- --- --- --- --- -AA GGT AAA GTC AAA ATC A Δ21 bp
#8
T CAT ACA GAT GA- --- --- --- --- --- --- --- --A GGT AAA GTC AAA ATC A Δ24 bp (−/−)
T CAT ACA GAT GAT GTC Tac aga t-- --- --- --- GAA GGT AAA GTC AAA ATC A Δ11 bp
#11
T CAT ACA GAT GAT GTC TCA TCA TCA G-- GAG CGA GAA GGT AAA GTC AAA ATC A Δ2 bp (−/−)
T CAT ACA GAT GA- --- --- --- --- --- --- --- --A GGT AAA GTC AAA ATC A Δ24 bp

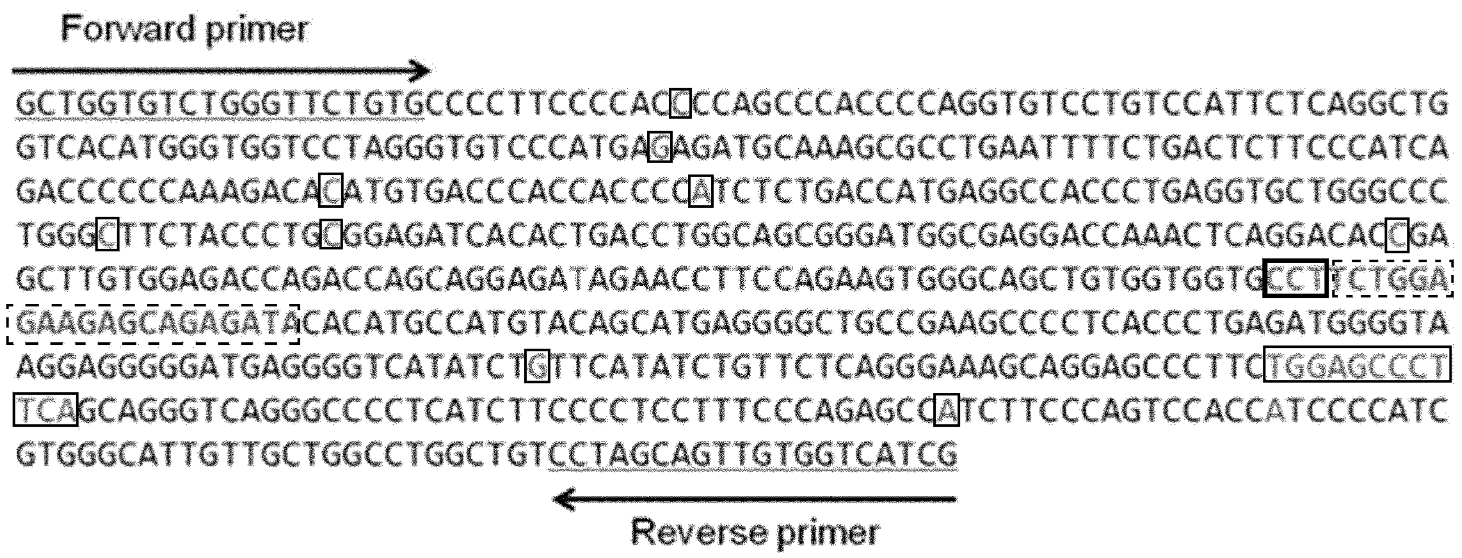
[Fig. 26b]



[Fig. 27]

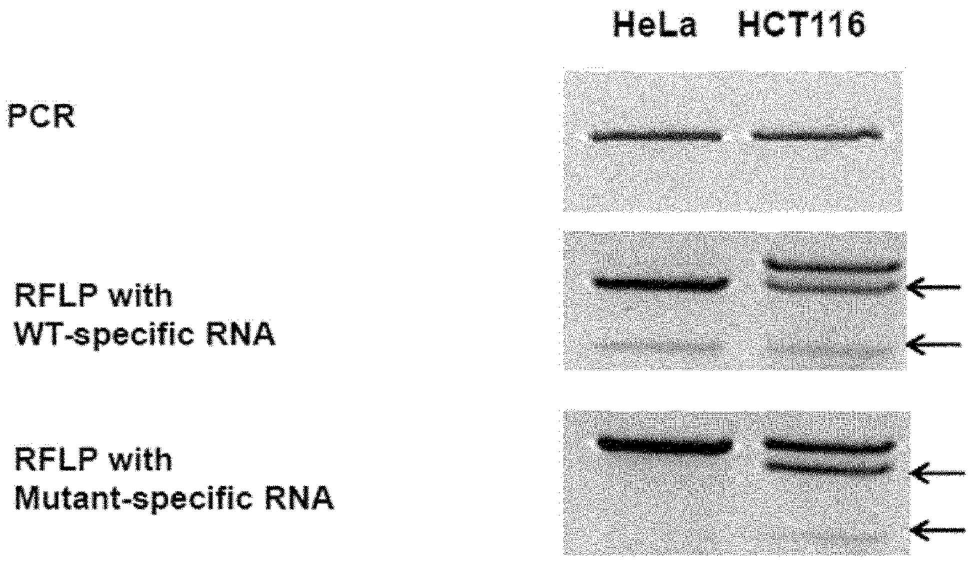


[Fig. 28]



GCTGGTGTCTGGGTTCTGTGCCCCTTCCCCACCCCAGCCCACCCCAGGTGTCCTGTCCATTCTCAGGCTG
GTCACATGGGTGGTCCTAGGGTGTCCCATGAGAGATGCAAAGCGCCTGAATTTTCTGACTCTTCCCATCA
GACCCCCCAAAGACACATGTGACCCACCACCCCATCTCTGACCATGAGGCCACCCTGAGGTGCTGGGCCC
TGGGCTTCTACCCTGCGGAGATCACACTGACCTGGCAGCGGGATGGCGAGGACCAAACTCAGGACACCGA
GCTTGTGGAGACCAGACCAGCAGGAGATAGAACCTTCCAGAAGTGGGCAGCTGTGGTGGTGCCTTCTGGA
GAAGAGCAGAGATACACATGCCATGTACAGCATGAGGGGCTGCCGAAGCCCCTCACCCTGAGATGGGGTA
AGGAGGGGGATGAGGGGTCATATCTGTTCATATCTGTTCTCAGGGAAAGCAGGAGCCCTTCTGGAGCCCT
TCAGCAGGGTCAGGGCCCCTCATCTTCCCCTCCTTTCCCAGAGCCATCTTCCCAGTCCACCATCCCCATC
GTGGGCATTGTTGCTGGCCTGGCTGTCCTAGCAGTTGTGGTCATCG

[Fig. 29a]



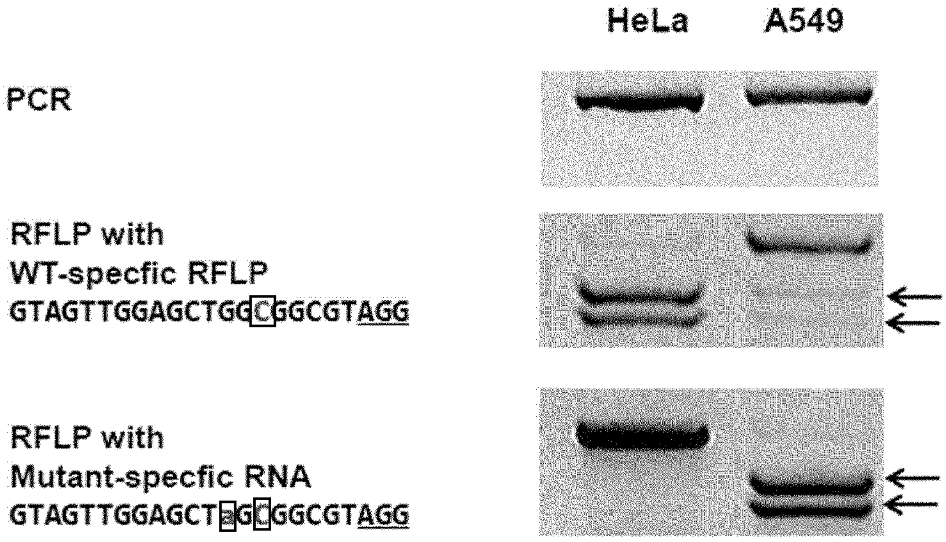
HeLa

ACTACCACAGCTCCTTCTCTGAGTGG wild-type

HCT116

ACTACCACAGCTCCTTCTCTGAGTGG wild-type
ACTACCACAGCTCCT---CTGAGTGG c.133-135 del TCT

[Fig. 29b]



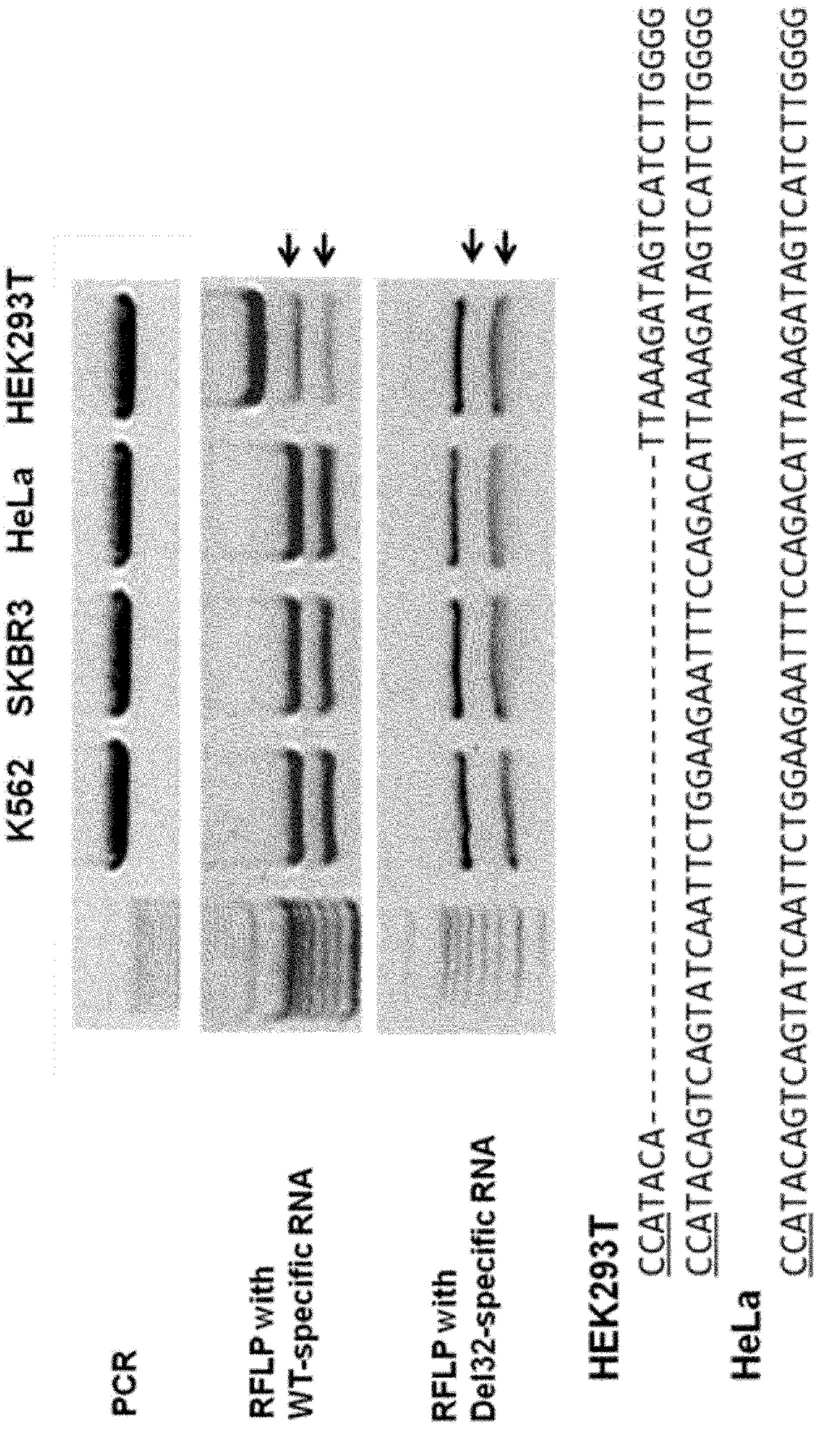
HeLa

GTAGTTGGAGCTGGTGGCGTAGG wild-type

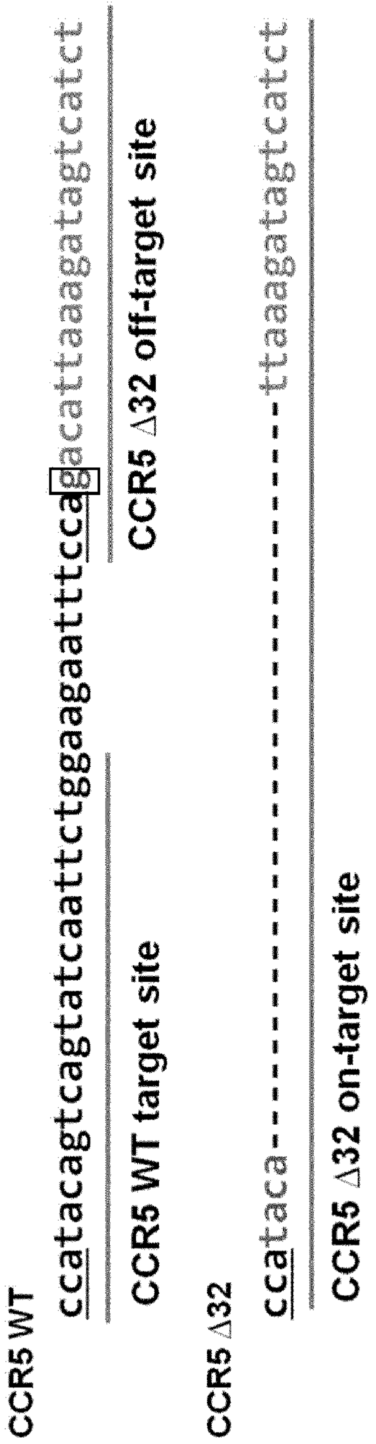
A549

GTAGTTGGAGCTaGTGGCGTAGG c.34G>A

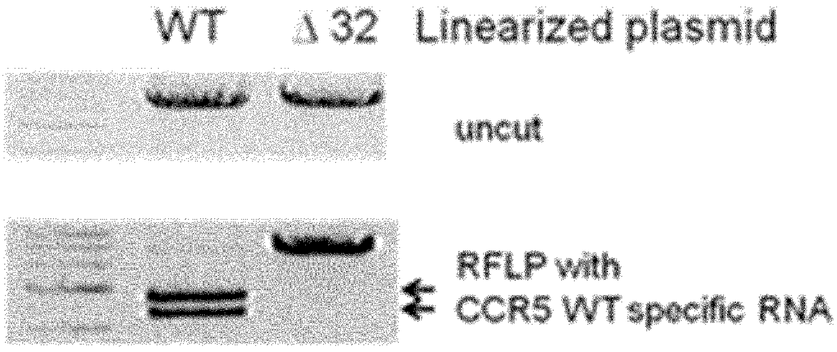
[Fig. 30a]


K562
SKBR3
HeLa
HEK293T
PCR
RFLP with WT-specific RNA
RFLP with Del32-specific RNA
HEK293T
CCATACA--------------------------------TTAAAGATAGTCATCTTGGGG
CCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGG
HeLa
CCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGG

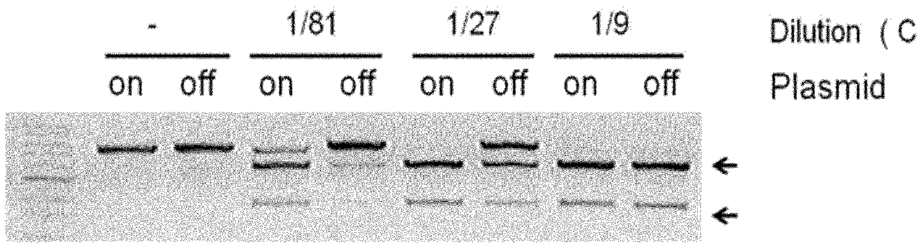
[Fig. 30b]



[Fig. 30c]

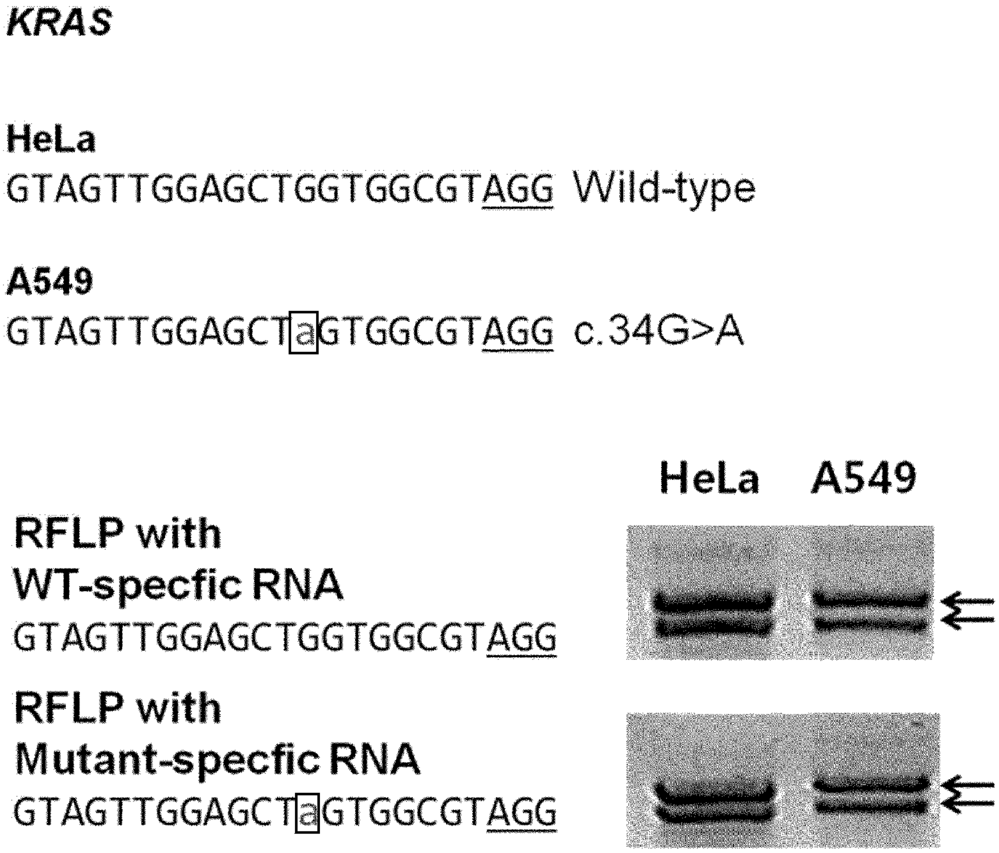


[Fig. 30d]

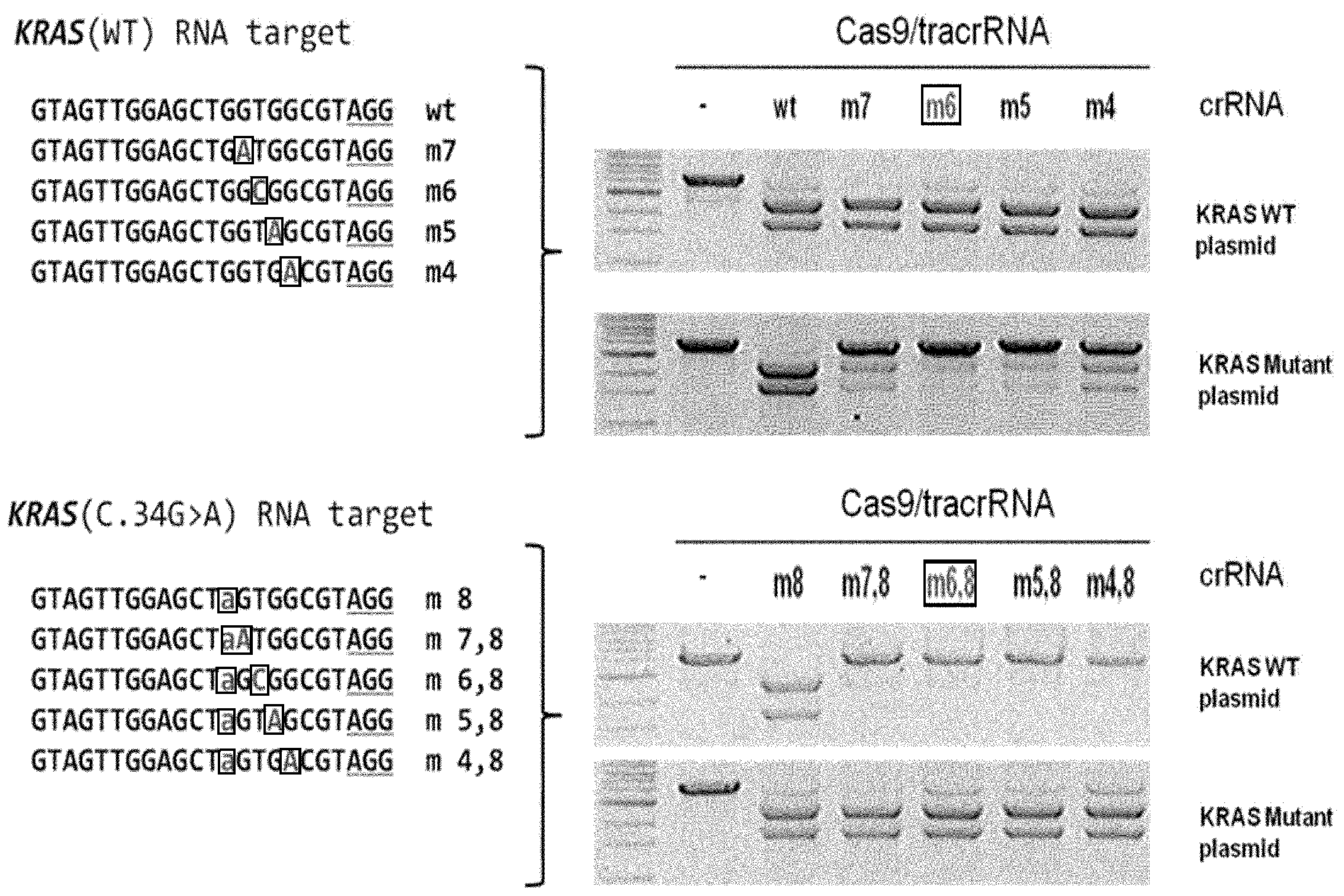


On agatgactatctttaatgtc<u>tgg</u>
Off agatgactatctttaatgt$\boxed{a}$<u>tgg</u>

[Fig. 31a]


KRAS
HeLa
GTAGTTGGAGCTGGTGGCGTAGG Wild-type
A549
GTAGTTGGAGCTaGTGGCGTAGG c.34G>A
HeLa
A549
RFLP with
WT-specfic RNA
GTAGTTGGAGCTGGTGGCGTAGG
RFLP with
Mutant-specfic RNA
GTAGTTGGAGCTaGTGGCGTAGG

[Fig. 31b]


KRAS(WT) RNA target
GTAGTTGGAGCTGGTGGCGTAGG wt
GTAGTTGGAGCTGATGGCGTAGG m7
GTAGTTGGAGCTGGCGGCGTAGG m6
GTAGTTGGAGCTGGTAGCGTAGG m5
GTAGTTGGAGCTGGTGACGTAGG m4
Cas9/tracrRNA
-
wt
m7
m6
m5
m4
crRNA
KRAS WT plasmid
KRAS Mutant plasmid
KRAS(C.34G>A) RNA target
GTAGTTGGAGCTaGTGGCGTAGG m 8
GTAGTTGGAGCTaATGGCGTAGG m 7,8
GTAGTTGGAGCTaGCGGCGTAGG m 6,8
GTAGTTGGAGCTaGTAGCGTAGG m 5,8
GTAGTTGGAGCTaGTGACGTAGG m 4,8
Cas9/tracrRNA
-
m8
m7,8
m6,8
m5,8
m4,8
crRNA
KRAS WT plasmid
KRAS Mutant plasmid

[Fig. 32a]

PIK3CA

HeLa

CAAATGAATGATGCACATCATGG Wild-type

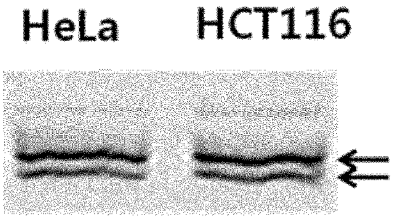
HCT116

CAAATGAATGATGCACATCATGG Wild-type

CAAATGAATGATGCACgTCATGG C.3140A>G

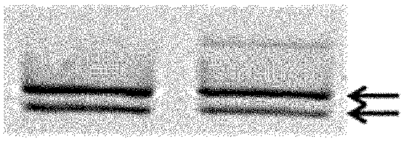
RFLP with WT-specfic RNA

CAAATGAATGATGCACATCATGG

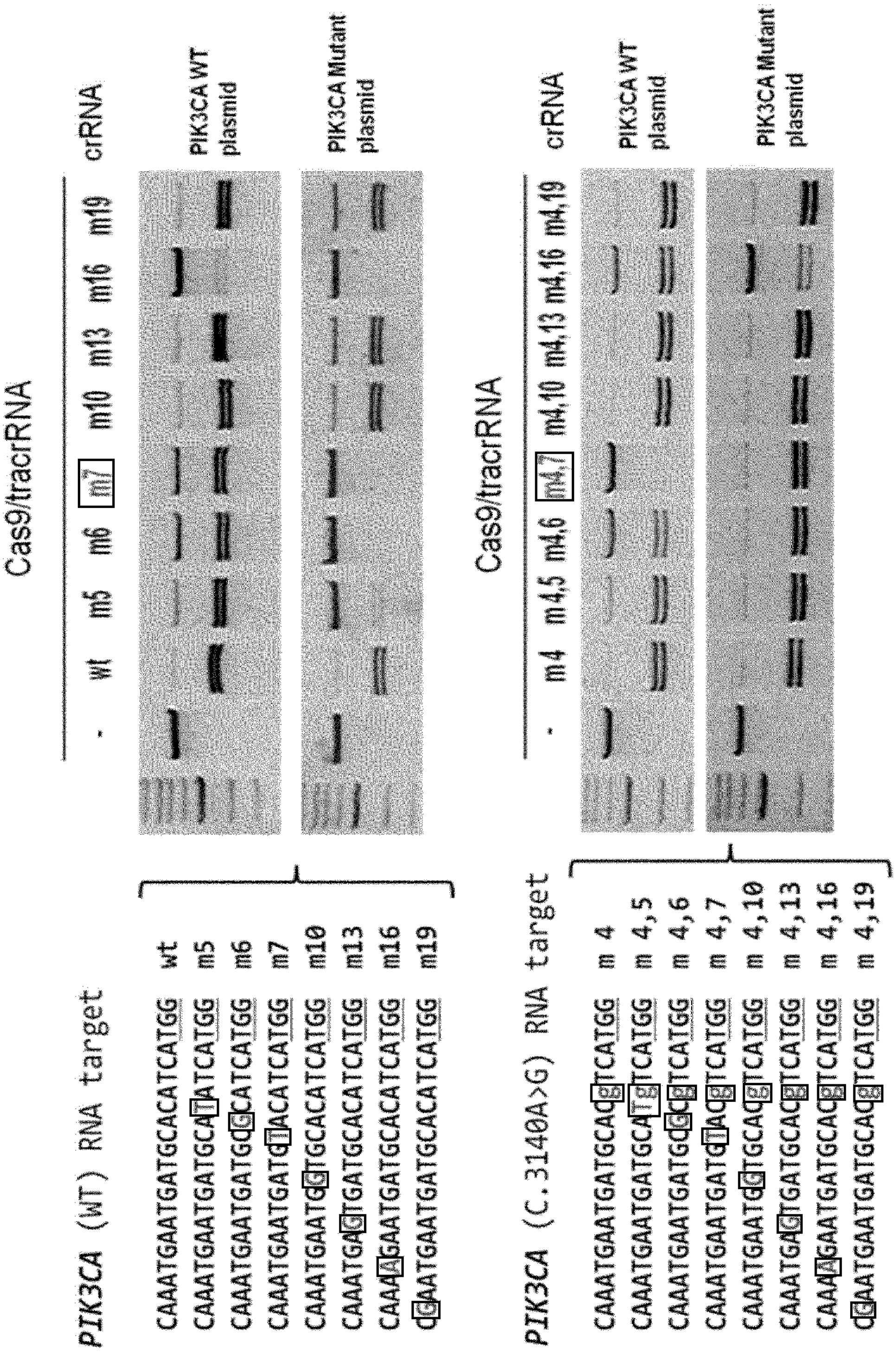


RFLP with Mutant-specfic RNA

CAAATGAATGATGCACgTCATGG



[Fig. 32b]


Cas9/tracrRNA
- wt m5 m6 m7 m10 m13 m16 m19 crRNA
PIK3CA WT plasmid
PIK3CA Mutant plasmid
Cas9/tracrRNA
- m4 m4,5 m4,6 m4,7 m4,10 m4,13 m4,16 m4,19 crRNA
PIK3CA WT plasmid
PIK3CA Mutant plasmid
PIK3CA (WT) RNA target
CAAATGAATGATGCACATCATGG wt
CAAATGAATGATGCATATCATGG m5
CAAATGAATGATGCGCATCATGG m6
CAAATGAATGATGTACATCATGG m7
CAAATGAATGGTGCACATCATGG m10
CAAATGAGTGATGCACATCATGG m13
CAAAAGAATGATGCACATCATGG m16
CGAATGAATGATGCACATCATGG m19
PIK3CA (C.3140A>G) RNA target
CAAATGAATGATGCACgTCATGG m 4
CAAATGAATGATGCATgTCATGG m 4,5
CAAATGAATGATGCGCgTCATGG m 4,6
CAAATGAATGATGTACgTCATGG m 4,7
CAAATGAATGGTGCACgTCATGG m 4,10
CAAATGAGTGATGCACgTCATGG m 4,13
CAAAAGAATGATGCACgTCATGG m 4,16
CGAATGAATGATGCACgTCATGG m 4,19

[Fig. 33a]

IDH1

HeLa

ATCATAGGTCGTCATGCTTATGG Wild-type

HT1080

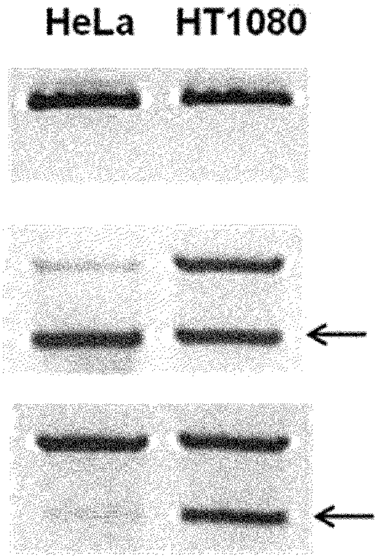
ATCATAGGTCGTCATGCTTATGG Wild-type

ATCATAGGTtGTCATGCTTATGG c.394C>T

PCR

RFLP with
WT-Specific RNA
ATCATAGGTCGTCCTGCTTATGG

RFLP with
Mutant-specific RNA
ATCATAGGTtGTCCTGCTTATGG



[Fig. 33b]

PIK3CA

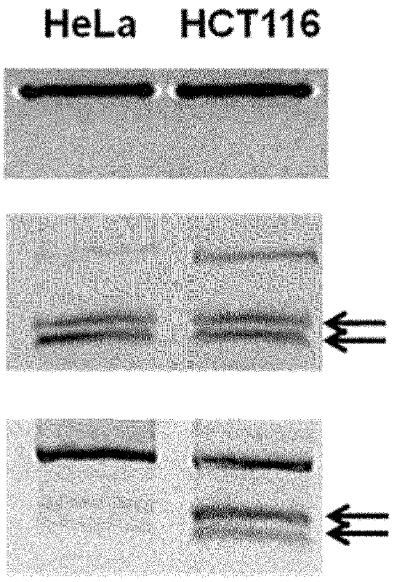
HeLa

CAAATGAATGATGCACATCATGG Wild-type

HCT116

CAAATGAATGATGCACATCATGG Wild-type

CAAATGAATGATGCACgTCATGG C.3140A>G



PCR

RFLP with
WT-Specific RNA

CAAATGAATGATGTACATCATGG

RFLP with
Mutant-specific RNA

CAAATGAATGATGTACgTCATGG

[Fig. 33c]

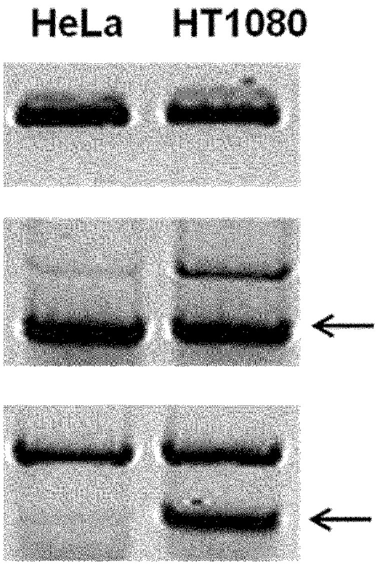
NRAS

HeLa

CTGGACAAGAAGAGTACAGTGCC Wild-type

HT1080

CTGGACAAGAAGAGTACAGTGCC Wild-type
CTGGAaAAGAAGAGTACAGTGCC c.181C>A

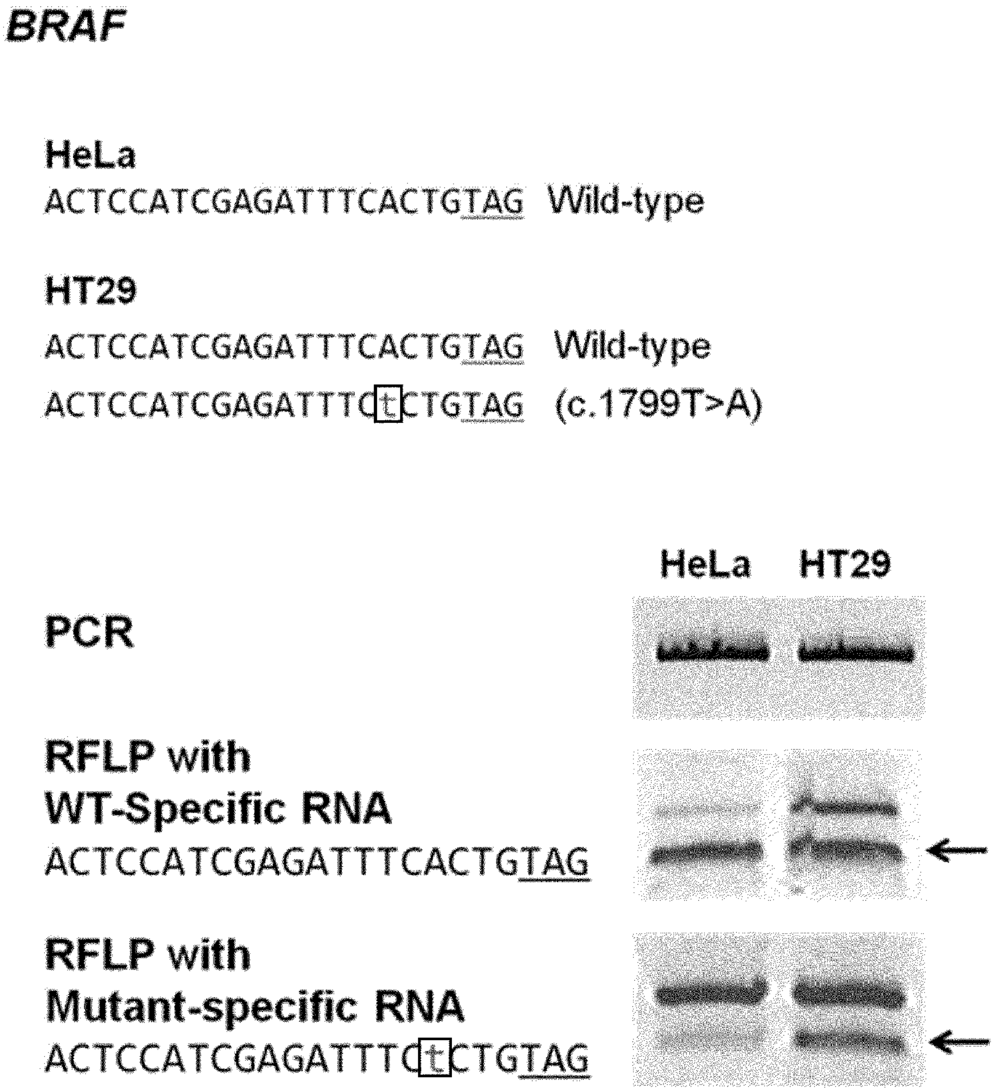


PCR

RFLP with
WT-Specific RNA
CTGGACAAGAAGAGTACAGTGCC

RFLP with
Mutant-specific RNA
CTGGAaAAGAAGAGTACAGTGCC

[Fig. 33d]


BRAF
HeLa
ACTCCATCGAGATTTCACTGTAG Wild-type
HT29
ACTCCATCGAGATTTCACTGTAG Wild-type
ACTCCATCGAGATTTCtCTGTAG (c.1799T>A)
HeLa
HT29
PCR
RFLP with
WT-Specific RNA
ACTCCATCGAGATTTCACTGTAG
RFLP with
Mutant-specific RNA
ACTCCATCGAGATTTCtCTGTAG

<110>   TOOLGEN INCORPORATED

<120>   Composition for cleaving a target DNA comprising a guide RNA
         specific for the target DNA and Cas protein-encoding nucleic acid
         or Cas protein, and use thereof

<130>   OPA13141PCT

<150>   US 61/717,324
<151>   2012-10-23

<150>   US 61/803,599
<151>   2013-03-20

<150>   US 61/837,481
<151>   2013-06-20

<160>   111

<170>   KopatentIn 2.0

<210>   1
<211>   4107
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Cas9-coding sequence

<400>   1
atggacaaga agtacagcat cggcctggac atcggtacca acagcgtggg ctgggccgtg   60

atcaccgacg agtacaaggt gcccagcaag aagttcaagg tgctgggcaa caccgaccgc   120

cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag   180

gccacccgcc tgaagcgcac cgcccgccgc cgctacaccc gccgcaagaa ccgcatctgc   240

tacctgcagg agatcttcag caacgagatg gccaaggtgg acgacagctt cttccaccgc   300

ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccaccc catcttcggc   360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgcgcaag   420

aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac   480

atgatcaagt tccgcggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac   540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc   600

atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc   660

cgcctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac   720

ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag   780

gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc   840

cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc   900

ctgctgagcg acatcctgcg cgtgaacacc gagatcacca aggccccccт gagcgccagc   960

atgatcaagc gctacgacga gcaccaccag gacctgaccc tgctgaaggc cctggtgcgc   1020

cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc   1080

ggctacatcg acggcggcgc cagccaggag gagttctaca agttcatcaa gcccatcctg   1140

```
gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc      1200
aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgcac      1260
gccatcctgc gccgccagga ggacttctac cccttcctga aggacaaccg cgagaagatc      1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc      1380
cgcttcgcct ggatgacccg caagagcgag gagaccatca cccccctggaa cttcgaggag      1440
gtggtggaca agggcgccag cgcccagagc ttcatcgagc gcatgaccaa cttcgacaag      1500
aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg      1560
tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttcctg      1620
agcggcgagc agaagaaggc catcgtggac ctgctgttca agaccaaccg caaggtgacc      1680
gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc      1740
agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc      1800
atcaaggaca aggacttcct ggacaacgag gagaacgagg acatcctgga ggacatcgtg      1860
ctgaccctga ccctgttcga ggaccgcgag atgatcgagg agcgcctgaa gacctacgcc      1920
cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggctggggc      1980
cgcctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg      2040
gacttcctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac      2100
agcctgacct tcaaggagga catccagaag gcccaggtga gcggccaggg cgacagcctg      2160
cacgagcaca tcgccaacct ggccggcagc cccgccatca agaagggcat cctgcagacc      2220
gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg      2280
atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc      2340
atgaagcgca tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc      2400
gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc      2460
gacatgtacg tggaccagga gctggacatc aaccgcctga gcgactacga cgtggaccac      2520
atcgtgcccc agagcttcct gaaggacgac agcatcgaca acaaggtgct gacccgcagc      2580
gacaagaacc gcggcaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaag      2640
aactactggc gccagctgct gaacgccaag ctgatcaccc agcgcaagtt cgacaacctg      2700
accaaggccg agcgcggcgg cctgagcgag ctggacaagg ccggcttcat caagcgccag      2760
ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac      2820
accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc      2880
aagctggtga gcgacttccg caaggacttc cagttctaca aggtgcgcga gatcaacaac      2940
taccaccacg cccacgacgc ctacctgaac gccgtggtgg gcaccgccct gatcaagaag      3000
taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag      3060
atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtactt cttctacagc      3120
aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc      3180
```

```
ccccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc       3240

gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg       3300

cagaccggcg gcttcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgatc       3360

gcccgcaaga aggactggga ccccaagaag tacggcggct tcgacagccc caccgtggcc       3420

tacagcgtgc tggtggtggc caaggtggag aagggcaaga gcaagaagct gaagagcgtg       3480

aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac       3540

ttcctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgcccaag       3600

tacagcctgt tcgagctgga gaacggccgc aagcgcatgc tggccagcgc cggcgagctg       3660

cagaagggca acgagctggc cctgcccagc aagtacgtga acttcctgta cctggccagc       3720

cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag       3780

cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg       3840

atcctggccg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag       3900

cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc       3960

cccgccgcct tcaagtactt cgacaccacc atcgaccgca agcgctacac cagcaccaag       4020

gaggtgctgg acgccaccct gatccaccag agcatcaccg gtctgtacga gacccgcatc       4080

gacctgagcc agctgggcgg cgactaa                                          4107


<210>   2
<211>   21
<212>   PRT
<213>   Artificial Sequence

<220>
<223>   peptide tag


<400>   2
Gly Gly Ser Gly Pro Pro Lys Lys Lys Arg Lys Val Tyr Pro Tyr Asp
  1               5                  10                  15

Val Pro Asp Tyr Ala
             20


<210>   3
<211>   34
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F primer for CCR5


<400>   3
aattcatgac atcaattatt atacatcgga ggag                                    34


<210>   4
<211>   34
<212>   DNA
<213>   Artificial Sequence
```

```
<220>
<223>    R primer for CCR5


<400>    4
gatcctcctc cgatgtataa taattgatgt catg                                   34


<210>    5
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F1 primer for CCR5


<400>    5
ctccatggtg ctatagagca                                                   20


<210>    6
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F2 primer for CCR5


<400>    6
gagccaagct ctccatctag t                                                 21


<210>    7
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for CCR5


<400>    7
gccctgtcaa gagttgacac                                                   20


<210>    8
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F1 primer for C4BPB


<400>    8
tatttggctg gttgaaaggg                                                   20


<210>    9
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R1 primer for C4BPB
```

```
<400>    9
aaagtcatga aataaacaca ccca                                              24


<210>    10
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F2 primer for C4BPB


<400>    10
ctgcattgat atggtagtac catg                                              24


<210>    11
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R2 primer for C4BPB


<400>    11
gctgttcatt gcaatggaat g                                                 21


<210>    12
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F1 primer for ADCY5


<400>    12
gctcccacct tagtgctctg                                                   20


<210>    13
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R1 primer for ADCY5


<400>    13
ggtggcagga acctgtatgt                                                   20


<210>    14
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F2 primer for ADCY5


<400>    14
```

```
gtcattggcc agagatgtgg a                                                21


<210>    15
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R2 primer for ADCY5


<400>    15
gtcccatgac aggcgtgtat                                                  20


<210>    16
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F primer for KCNJ6


<400>    16
gcctggccaa gtttcagtta                                                  20


<210>    17
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R1 primer for KCNJ6


<400>    17
tggagccatt ggtttgcatc                                                  20


<210>    18
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R2 primer for KCNJ6


<400>    18
ccagaactaa gccgtttctg ac                                               22


<210>    19
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F1 primer for CNTNAP2


<400>    19
atcaccgaca accagtttcc                                                  20
```

<210> 20
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> F2 primer for CNTNAP2

<400> 20
tgcagtgcag actctttcca 20

<210> 21
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> R primer for CNTNAP2

<400> 21
aaggacacag ggcaactgaa 20

<210> 22
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> F1 primer for N/A Chr. 5

<400> 22
tgtggaacga gtggtgacag 20

<210> 23
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> R1 primer for N/A Chr. 5

<400> 23
gctggattag gaggcaggat tc 22

<210> 24
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> F2 primer for N/A Chr. 5

<400> 24
gtgctgagaa cgcttcatag ag 22

<210> 25
<211> 23
<212> DNA

```
<213>    Artificial Sequence

<220>
<223>    R2 primer for N/A Chr. 5


<400>    25
ggaccaaacc acattcttct cac                                               23


<210>    26
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F primer for deletion


<400>    26
ccacatctcg ttctcggttt                                                   20


<210>    27
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for deletion


<400>    27
tcacaagccc acagatattt                                                   20


<210>    28
<211>    105
<212>    RNA
<213>    Artificial Sequence

<220>
<223>    sgRNA for CCR5


<400>    28
ggugacauca auuauuauac auguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag       60

uccguuauca acuugaaaaa guggcaccga gucggugcuu uuuuu                      105


<210>    29
<211>    44
<212>    RNA
<213>    Artificial Sequence

<220>
<223>    crRNA for CCR5


<400>    29
ggugacauca auuauuauac auguuuuaga gcuaugcugu uuug                        44


<210>    30
<211>    86
<212>    RNA
<213>    Artificial Sequence
```

<220>
<223>   tracrRNA for CCR5

<400>   30
ggaaccauuc aaaacagcau agcaaguuaa aauaaggcua guccguuauc aacuugaaaa   60

aguggcaccg agucggugcu uuuuuu   86

<210>   31
<211>   86
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Foxn1 #1 sgRNA

<400>   31
gaaattaata cgactcacta taggcagtct gacgtcacac ttccgtttta gagctagaaa   60

tagcaagtta aaataaggct agtccg   86

<210>   32
<211>   86
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Foxn1 #2 sgRNA

<400>   32
gaaattaata cgactcacta taggacttcc aggctccacc cgacgtttta gagctagaaa   60

tagcaagtta aaataaggct agtccg   86

<210>   33
<211>   86
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Foxn1 #3 sgRNA

<400>   33
gaaattaata cgactcacta taggccaggc tccacccgac tggagtttta gagctagaaa   60

tagcaagtta aaataaggct agtccg   86

<210>   34
<211>   86
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Foxn1 #4 sgRNA

<400>   34
gaaattaata cgactcacta taggactgga gggcgaaccc caaggtttta gagctagaaa   60

```
tagcaagtta aaataaggct agtccg                                            86


<210>  35
<211>  86
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Foxn1 #5 sgRNA


<400>  35
gaaattaata cgactcacta taggacccca aggggacctc atgcgtttta gagctagaaa      60

tagcaagtta aaataaggct agtccg                                            86


<210>  36
<211>  86
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Prkdc #1 sgRNA


<400>  36
gaaattaata cgactcacta taggttagtt ttttccagag acttgtttta gagctagaaa      60

tagcaagtta aaataaggct agtccg                                            86


<210>  37
<211>  86
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Prkdc #2 sgRNA


<400>  37
gaaattaata cgactcacta taggttggtt tgcttgtgtt tatcgtttta gagctagaaa      60

tagcaagtta aaataaggct agtccg                                            86


<210>  38
<211>  86
<212>  DNA
<213>  Artificial Sequence

<220>
<223>  Prkdc #3 sgRNA


<400>  38
gaaattaata cgactcacta taggcacaag caaaccaaag tctcgtttta gagctagaaa      60

tagcaagtta aaataaggct agtccg                                            86


<210>  39
<211>  86
<212>  DNA
<213>  Artificial Sequence
```

```
<220>
<223>   Prkdc #4 sgRNA


<400>   39
gaaattaata cgactcacta taggcctcaa tgctaagcga cttcgtttta gagctagaaa       60

tagcaagtta aaataaggct agtccg                                              86


<210>   40
<211>   29
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F1 primer for Foxn1


<400>   40
gtctgtctat catctcttcc cttctctcc                                           29


<210>   41
<211>   25
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F2 primer for Foxn1


<400>   41
tccctaatcc gatggctagc tccag                                               25


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

<220>
<223>   R1 primer for Foxn1


<400>   42
acgagcagct gaagttagca tgc                                                 23


<210>   43
<211>   32
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   R2 primer for Foxn1


<400>   43
ctactcaatg ctcttagagc taccaggctt gc                                       32


<210>   44
<211>   20
<212>   DNA
<213>   Artificial Sequence

<220>
```

```
<223>    F primer for Prkdc


<400>    44
gactgttgtg gggagggccg                                                   20


<210>    45
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F2 primer for Prkdc


<400>    45
gggagggccg aaagtcttat tttg                                              24


<210>    46
<211>    28
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R1 primer for Prkdc


<400>    46
cctgaagact gaagttggca gaagtgag                                          28


<210>    47
<211>    27
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R2 primer for Prkdc


<400>    47
ctttagggct tcttctctac aatcacg                                           27


<210>    48
<211>    38
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F primer for Foxn1


<400>    48
ctcggtgtgt agccctgacc tcggtgtgta gccctgac                               38


<210>    49
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for Foxn1
```

```
<400>   49
agactggcct ggaactcaca g                                                  21


<210>   50
<211>   23
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F primer for Foxn1


<400>   50
cactaaagcc tgtcaggaag ccg                                                23


<210>   51
<211>   21
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   R primer for Foxn1


<400>   51
ctgtggagag cacacagcag c                                                  21


<210>   52
<211>   19
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F primer for Foxn1


<400>   52
gctgcgacct gagaccatg                                                     19


<210>   53
<211>   26
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   R primer for Foxn1


<400>   53
cttcaatggc ttcctgctta ggctac                                             26


<210>   54
<211>   23
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   F primer for Foxn1


<400>   54
ggttcagatg aggccatcct ttc                                                23
```

```
<210>    55
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for Foxn1


<400>    55
cctgatctgc aggcttaacc cttg                                            24


<210>    56
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F primer for Prkdc


<400>    56
ctcacctgca catcacatgt gg                                              22


<210>    57
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for Prkdc


<400>    57
ggcatccacc ctatggggtc                                                 20


<210>    58
<211>    25
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F primer for Prkdc


<400>    58
gccttgacct agagcttaaa gagcc                                           25


<210>    59
<211>    25
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R primer for Prkdc


<400>    59
ggtcttgtta gcaggaagga cactg                                           25


<210>    60
<211>    27
```

<212> DNA
<213> Artificial Sequence

<220>
<223> F primer for Prkdc

<400> 60
aaaactctgc ttgatgggat atgtggg 27

<210> 61
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> R primer for Prkdc

<400> 61
ctctcactgg ttatctgtgc tccttc 26

<210> 62
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> F primer for Prkdc

<400> 62
ggatcaatag gtggtggggg atg 23

<210> 63
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> R primer for Prkdc

<400> 63
gtgaatgaca caatgtgaca gcttcag 27

<210> 64
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> F primer for Prkdc

<400> 64
cacaagacag acctctcaac attcagtc 28

<210> 65
<211> 32
<212> DNA
<213> Artificial Sequence

```
<220>
<223>    R primer for Prkdc


<400>    65
gtgcatgcat ataatccatt ctgattgctc tc                                       32


<210>    66
<211>    17
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F1 primer for Prkdc


<400>    66
gggaggcaga ggcaggt                                                        17


<210>    67
<211>    23
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    F2 primer for Prkdc


<400>    67
ggatctctgt gagtttgagg cca                                                 23


<210>    68
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    R1 primer for Prkdc


<400>    68
gctccagaac tcactcttag gctc                                                24


<210>    69
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer for Foxn1


<400>    69
ctactccctc cgcagtctga                                                     20


<210>    70
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer for Foxn1
```

```
<400>    70
ccaggcctag gttccaggta                                                   20


<210>    71
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer for Prkdc


<400>    71
ccccagcatt gcagatttcc                                                   20


<210>    72
<211>    23
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer for Prkdc


<400>    72
agggcttctt ctctacaatc acg                                               23


<210>    73
<211>    86
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    BRI1 target 1


<400>    73
gaaattaata cgactcacta taggtttgaa agatggaagc gcgggtttta gagctagaaa        60

tagcaagtta aaataaggct agtccg                                            86


<210>    74
<211>    86
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    BRI1 target 2


<400>    74
gaaattaata cgactcacta taggtgaaac taaactggtc cacagtttta gagctagaaa        60

tagcaagtta aaataaggct agtccg                                            86


<210>    75
<211>    64
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Universal
```

```
<400>   75
aaaaaagcac cgactcggtg ccactttttc aagttgataa cggactagcc ttattttaac        60

ttgc                                                                     64


<210>   76
<211>   65
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Templates for crRNA


<400>   76
gaaattaata cgactcacta taggnnnnnn nnnnnnnnnn nnnngtttta gagctatgct        60

gtttt                                                                    65


<210>   77
<211>   67
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   tracrRNA


<400>   77
gaaattaata cgactcacta taggaaccat tcaaaacagc atagcaagtt aaaataaggc        60

tagtccg                                                                  67


<210>   78
<211>   69
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   tracrRNA


<400>   78
aaaaaaagca ccgactcggt gccacttttt caagttgata acggactagc cttattttaa        60

cttgctatg                                                                69


<210>   79
<211>   20
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   79
ctccatggtg ctatagagca                                                    20


<210>   80
<211>   21
```

```
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    80
gagccaagct ctccatctag t                                                    21


<210>    81
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    81
gccctgtcaa gagttgacac                                                      20


<210>    82
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    82
gcacagggtg gaacaagatg ga                                                   22


<210>    83
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    83
gccaggtacc tatcgattgt cagg                                                 24


<210>    84
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    84
gagccaagct ctccatctag t                                                    21


<210>    85
<211>    20
<212>    DNA
<213>    Artificial Sequence
```

```
<220>
<223>    Primer


<400>    85
actctgactg ggtcaccagc                                                   20


<210>    86
<211>    20
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    86
tatttggctg gttgaaaggg                                                   20


<210>    87
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    87
aaagtcatga aataaacaca ccca                                              24


<210>    88
<211>    24
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    88
ctgcattgat atggtagtac catg                                              24


<210>    89
<211>    21
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    89
gctgttcatt gcaatggaat g                                                 21


<210>    90
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer
```

```
<400>   90
atggagttgg acatggccat gg                                              22


<210>   91
<211>   28
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   91
actcactatc cacagttcag catttacc                                        28


<210>   92
<211>   23
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   92
tggagatagc tgtcagcaac ttt                                             23


<210>   93
<211>   29
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   93
caacaaagca aaggtaaagt tggtaatag                                       29


<210>   94
<211>   25
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   94
ggtttcagga gatgtgttac aaggc                                           25


<210>   95
<211>   27
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   95
gattgtgcaa ttcctatgca atcggtc                                         27
```

```
<210>   96
<211>   25
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   96
cactgggtac ttaatctgta gcctc                                           25


<210>   97
<211>   23
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   97
ggttccaagt cattcccagt agc                                             23


<210>   98
<211>   30
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   98
catcactgca gttgtaggtt ataactatcc                                      30


<210>   99
<211>   26
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   99
ttgaaaacca cagatctggt tgaacc                                          26


<210>   100
<211>   22
<212>   DNA
<213>   Artificial Sequence

<220>
<223>   Primer


<400>   100
ggagtgccaa gagaatatct gg                                              22


<210>   101
```

```
<211>    32
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    101
ctgaaactgg tttcaaaata ttcgttttaa gg                                              32


<210>    102
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    102
gctctgtatg ccctgtagta gg                                                         22


<210>    103
<211>    22
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Primer


<400>    103
tttgcatctg accttacctt tg                                                         22


<210>    104
<211>    23
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Target sequence of RGEN


<400>    104
aatgaccact acatcctcaa ggg                                                        23


<210>    105
<211>    23
<212>    DNA
<213>    Artificial Sequence

<220>
<223>    Target sequence of RGEN


<400>    105
agatgatgtc tcatcatcag agg                                                        23


<210>    106
<211>    4170
<212>    DNA
<213>    Artificial Sequence
```

<220>
<223>   Cas9-coding sequence in p3s-Cas9HC (humanized, C-term tagging,
         human cell experiments)

<400>   106
atggacaaga agtacagcat cggcctggac atcggtacca acagcgtggg ctgggccgtg   60

atcaccgacg agtacaaggt gcccagcaag aagttcaagg tgctgggcaa caccgaccgc   120

cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag   180

gccacccgcc tgaagcgcac cgcccgccgc cgctacaccc gccgcaagaa ccgcatctgc   240

tacctgcagg agatcttcag caacgagatg gccaaggtgg acgacagctt cttccaccgc   300

ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccaccc catcttcggc   360

aacatcgtgg acgaggtggc ctaccacgag aagtacccca ccatctacca cctgcgcaag   420

aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac   480

atgatcaagt tccgcggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac   540

gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc   600

atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc   660

cgcctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac   720

ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag   780

gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc   840

cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc   900

ctgctgagcg acatcctgcg cgtgaacacc gagatcacca aggcccccct gagcgccagc   960

atgatcaagc gctacgacga gcaccaccag gacctgaccc tgctgaaggc cctggtgcgc   1020

cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc   1080

ggctacatcg acggcggcgc cagccaggag gagttctaca agttcatcaa gcccatcctg   1140

gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc   1200

aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgcac   1260

gccatcctgc gccgccagga ggacttctac cccttcctga aggacaaccg cgagaagatc   1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc   1380

cgcttcgcct ggatgacccg caagagcgag gagaccatca ccccctggaa cttcgaggag   1440

gtggtggaca agggcgccag cgcccagagc ttcatcgagc gcatgaccaa cttcgacaag   1500

aacctgccca acgagaaggt gctgcccaag cacagcctgc tgtacgagta cttcaccgtg   1560

tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttcctg   1620

agcggcgagc agaagaaggc catcgtggac ctgctgttca agaccaaccg caaggtgacc   1680

gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc   1740

agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc   1800

atcaaggaca aggacttcct ggacaacgag gagaacgagg acatcctgga ggacatcgtg   1860

```
ctgaccctga ccctgttcga ggaccgcgag atgatcgagg agcgcctgaa gacctacgcc       1920
cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggctggggc       1980
cgcctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg       2040
gacttcctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac       2100
agcctgacct tcaaggagga catccagaag gcccaggtga gcggccaggg cgacagcctg       2160
cacgagcaca tcgccaacct ggccggcagc cccgccatca agaagggcat cctgcagacc       2220
gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg       2280
atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc       2340
atgaagcgca tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc       2400
gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc       2460
gacatgtacg tggaccagga gctggacatc aaccgcctga gcgactacga cgtggaccac       2520
atcgtgcccc agagcttcct gaaggacgac agcatcgaca acaaggtgct gacccgcagc       2580
gacaagaacc gcggcaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaag       2640
aactactggc gccagctgct gaacgccaag ctgatcaccc agcgcaagtt cgacaacctg       2700
accaaggccg agcgcggcgg cctgagcgag ctggacaagg ccggcttcat caagcgccag       2760
ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac       2820
accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc       2880
aagctggtga gcgacttccg caaggacttc cagttctaca aggtgcgcga gatcaacaac       2940
taccaccacg cccacgacgc ctacctgaac gccgtggtgg gcaccgccct gatcaagaag       3000
taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag       3060
atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtactt cttctacagc       3120
aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc       3180
cccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc       3240
gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg       3300
cagaccggcg gcttcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgatc       3360
gcccgcaaga aggactggga ccccaagaag tacggcggct tcgacagccc caccgtggcc       3420
tacagcgtgc tggtggtggc caaggtggag aagggcaaga gcaagaagct gaagagcgtg       3480
aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac       3540
ttcctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgcccaag       3600
tacagcctgt tcgagctgga gaacggccgc aagcgcatgc tggccagcgc cggcgagctg       3660
cagaagggca acgagctggc cctgcccagc aagtacgtga acttcctgta cctggccagc       3720
cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag       3780
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg       3840
atcctggccg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag       3900
```

cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc 3960

cccgccgcct tcaagtactt cgacaccacc atcgaccgca agcgctacac cagcaccaag 4020

gaggtgctgg acgccaccct gatccaccag agcatcaccg gtctgtacga gacccgcatc 4080

gacctgagcc agctgggcgg cgacggcggc tccggacctc caaagaaaaa gagaaaagta 4140

tacccctacg acgtgcccga ctacgcctaa 4170

<210> 107
<211> 4194
<212> DNA
<213> Artificial Sequence

<220>
<223> Cas9 coding sequence in p3s-Cas9HN (humanized codon, N-term tagging (underlined), human cell experiments)

<400> 107
atggtgtacc cctacgacgt gcccgactac gccgaattgc ctccaaaaaa gaagagaaag 60

gtagggatcc gaattcccgg ggaaaaaccg gacaagaagt acagcatcgg cctggacatc 120

ggtaccaaca gcgtgggctg ggccgtgatc accgacgagt acaaggtgcc cagcaagaag 180

ttcaaggtgc tgggcaacac cgaccgccac agcatcaaga agaacctgat cggcgccctg 240

ctgttcgaca gcggcgagac cgccgaggcc acccgcctga agcgcaccgc ccgccgccgc 300

tacacccgcc gcaagaaccg catctgctac ctgcaggaga tcttcagcaa cgagatggcc 360

aaggtggacg acagcttctt ccaccgcctg gaggagagct tcctggtgga ggaggacaag 420

aagcacgagc gccaccccat cttcggcaac atcgtggacg aggtggccta ccacgagaag 480

taccccacca tctaccacct gcgcaagaag ctggtggaca gcaccgacaa ggccgacctg 540

cgcctgatct acctggccct ggcccacatg atcaagttcc gcggccactt cctgatcgag 600

ggcgacctga accccgacaa cagcgacgtg gacaagctgt tcatccagct ggtgcagacc 660

tacaaccagc tgttcgagga gaaccccatc aacgccagcg gcgtggacgc caaggccatc 720

ctgagcgccc gcctgagcaa gagccgccgc ctggagaacc tgatcgccca gctgcccggc 780

gagaagaaga acggcctgtt cggcaacctg atcgccctga gcctgggcct gacccccaac 840

ttcaagagca acttcgacct ggccgaggac gccaagctgc agctgagcaa ggacacctac 900

gacgacgacc tggacaacct gctggcccag atcggcgacc agtacgccga cctgttcctg 960

gccgccaaga acctgagcga cgccatcctg ctgagcgaca tcctgcgcgt gaacaccgag 1020

atcaccaagg ccccccctgag cgccagcatg atcaagcgct acgacgagca ccaccaggac 1080

ctgaccctgc tgaaggccct ggtgcgccag cagctgcccg agaagtacaa ggagatcttc 1140

ttcgaccaga gcaagaacgg ctacgccggc tacatcgacg gcggcgccag ccaggaggag 1200

ttctacaagt tcatcaagcc catcctggag aagatggacg gcaccgagga gctgctggtg 1260

aagctgaacc gcgaggacct gctgcgcaag cagcgcacct tcgacaacgg cagcatcccc 1320

caccagatcc acctgggcga gctgcacgcc atcctgcgcc gccaggagga cttctacccc 1380 ttcctgaagg acaaccgcga gaagatcgag aagatcctga ccttccgcat cccctactac 1440

gtgggccccc tggcccgcgg caacagccgc ttcgcctgga tgacccgcaa gagcgaggag 1500

accatcaccc cctggaactt cgaggaggtg gtggacaagg gcgccagcgc ccagagcttc 1560

atcgagcgca tgaccaactt cgacaagaac ctgcccaacg 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 cgccagcctg 1860

ggcacctacc acgacctgct gaagatcatc aaggacaagg acttcctgga caacgaggag 1920

aacgaggaca tcctggagga catcgtgctg accctgaccc tgttcgagga ccgcgagatg 1980

atcgaggagc gcctgaagac ctacgcccac ctgttcgacg acaaggtgat gaagcagctg 2040

aagcgccgcc gctacaccgg ctggggccgc ctgagccgca agcttatcaa cggcatccgc 2100

gacaagcaga gcggcaagac catcctggac ttcctgaaga gcgacggctt cgccaaccgc 2160

aacttcatgc agctgatcca cgacgacagc ctgaccttca aggaggacat ccagaaggcc 2220

caggtgagcg gccagggcga cagcctgcac gagcacatcg ccaacctggc cggcagcccc 2280

gccatcaaga agggcatcct gcagaccgtg aaggtggtgg acgagctggt gaaggtgatg 2340

ggccgccaca agcccgagaa catcgtgatc gagatggccc gcgagaacca gaccacccag 2400

aagggccaga agaacagccg cgagcgcatg aagcgcatcg aggagggcat caaggagctg 2460

ggcagccaga tcctgaagga gcaccccgtg gagaacaccc agctgcagaa cgagaagctg 2520

tacctgtact acctgcagaa cggccgcgac atgtacgtgg accaggagct ggacatcaac 2580

cgcctgagcg actacgacgt ggaccacatc gtgccccaga gcttcctgaa ggacgacagc 2640

atcgacaaca aggtgctgac ccgcagcgac aagaaccgcg gcaagagcga caacgtgccc 2700

agcgaggagg tggtgaagaa gatgaagaac tactggcgcc agctgctgaa cgccaagctg 2760

atcacccagc gcaagttcga caacctgacc aaggccgagc gcggcggcct gagcgagctg 2820

gacaaggccg gcttcatcaa gcgccagctg gtggagaccc gccagatcac caagcacgtg 2880

gcccagatcc tggacagccg catgaacacc aagtacgacg agaacgacaa gctgatccgc 2940

gaggtgaagg tgatcaccct gaagagcaag ctggtgagcg acttccgcaa ggacttccag 3000

ttctacaagg tgcgcgagat caacaactac caccacgccc acgacgccta cctgaacgcc 3060

gtggtgggca ccgccctgat caagaagtac cccaagctgg agagcgagtt cgtgtacggc 3120

gactacaagg tgtacgacgt gcgcaagatg atcgccaaga gcgagcagga gatcggcaag 3180

gccaccgcca agtacttctt ctacagcaac atcatgaact tcttcaagac cgagatcacc 3240

ctggccaacg gcgagatccg caagcgcccc ctgatcgaga ccaacggcga gaccggcgag 3300

atcgtgtggg acaagggccg cgacttcgcc accgtgcgca aggtgctgag catgccccag 3360

gtgaacatcg tgaagaagac cgaggtgcag accggcggct tcagcaagga gagcatcctg 3420 cccaagcgca acagcgacaa gctgatcgcc cgcaagaagg actgggaccc caagaagtac 3480

ggcggcttcg acagccccac cgtggcctac agcgtgctgg tggtggccaa ggtggagaag 3540

ggcaagagca agaagctgaa gagcgtgaag gagctgctgg gcatcaccat catggagcgc 3600

agcagcttcg agaagaaccc catcgacttc ctggaggcca agggctacaa ggaggtgaag 3660

aaggacctga tcatcaagct gcccaagtac agcctgttcg agctggagaa cggccgcaag 3720

cgcatgctgg ccagcgccgg cgagctgcag aagggcaacg agctggccct gcccagcaag 3780

tacgtgaact tcctgtacct ggccagccac tacgagaagc tgaagggcag ccccgaggac 3840

aacgagcaga agcagctgtt cgtggagcag cacaagcact acctggacga gatcatcgag 3900

cagatcagcg agttcagcaa gcgcgtgatc ctggccgacg ccaacctgga caaggtgctg 3960

agcgcctaca acaagcaccg cgacaagccc atccgcgagc aggccgagaa catcatccac 4020

ctgttcaccc tgaccaacct gggcgccccc gccgccttca agtacttcga caccaccatc 4080

gaccgcaagc gctacaccag caccaaggag gtgctggacg ccaccctgat ccaccagagc 4140

atcaccggtc tgtacgagac ccgcatcgac ctgagccagc tgggcggcga ctaa 4194

<210> 108
<211> 4107
<212> DNA
<213> Artificial Sequence

<220>
<223> Cas9-coding sequence in Streptococcus pyogenes

<400> 108
atggataaga aatactcaat aggcttagat atcggcacaa atagcgtcgg atgggcggtg 60

atcactgatg aatataaggt tccgtctaaa aagttcaagg ttctgggaaa tacagaccgc 120

cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcggaa 180

gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240

tatctacagg agatttttc aaatgagatg gcgaaagtag atgatagttt ctttcatcga 300

cttgaagagt cttttttggt ggaagaagac aagaagcatg aacgtcatcc tatttttgga 360

aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420

aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgcat 480

atgattaagt ttcgtggtca tttttgatt gagggagatt taaatcctga taatagtgat 540

gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct 600

attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660

cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720

ctcattgctt tgtcattggg tttgacccct aattttaaat caaattttga tttggcagaa 780

gatgctaaat tacagctttc aaaagatact tacgatgatg atttagataa tttattggcg 840

caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

```
ttactttcag atatcctaag agtaaatact gaaataacta aggctcccct atcagcttca        960

atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga       1020

caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca       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 aattccttat tatgttggtc cattggcgcg tggcaatagt       1380

cgttttgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa ttttgaagaa       1440

gttgtcgata aaggtgcttc agctcaatca tttattgaac gcatgacaaa ctttgataaa       1500

aatcttccaa atgaaaaagt actaccaaaa catagtttgc tttatgagta ttttacggtt       1560

tataacgaat tgacaaaggt caaatatgtt actgaaggaa tgcgaaaacc agcatttctt       1620

tcaggtgaac agaagaaagc cattgttgat ttactcttca aaacaaatcg aaaagtaacc       1680

gttaagcaat taaaagaaga ttatttcaaa aaaatagaat gttttgatag tgttgaaatt       1740

tcaggagttg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt       1800

attaaagata aagatttttt ggataatgaa gaaaatgaag atatcttaga ggatattgtt       1860

ttaacattga ccttatttga agatagggag atgattgagg aaagacttaa aacatatgct       1920

cacctctttg atgataaggt gatgaaacag cttaaacgtc gccgttatac tggttgggga       1980

cgtttgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta       2040

gatttttga aatcagatgg ttttgccaat cgcaatttta tgcagctgat ccatgatgat       2100

agtttgacat ttaaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagttta       2160

catgaacata ttgcaaattt agctggtagc cctgctatta aaaaaggtat tttacagact       2220

gtaaaagttg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcgtt       2280

attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt       2340

atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct       2400

gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga       2460

gacatgtatg tggaccaaga attagatatt aatcgtttaa gtgattatga tgtcgatcac       2520

attgttccac aaagtttcct taaagacgat tcaatagaca ataaggtctt aacgcgttct       2580

gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa       2640

aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta       2700

acgaaagctg aacgtggagg tttgagtgaa cttgataaag ctggttttat caaacgccaa       2760

ttggttgaaa ctcgccaaat cactaagcat gtggcacaaa ttttggatag tcgcatgaat       2820

actaaatacg atgaaaatga taaacttatt cgagaggtta aagtgattac cttaaaatct       2880

aaattagttt ctgacttccg aaaagatttc caattctata aagtacgtga gattaacaat       2940
```

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000

tatccaaaac ttgaatcgga gtttgtctat ggtgattata aagtttatga tgttcgtaaa 3060

atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120

aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180

cctctaatcg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240

gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300

cagacaggcg gattctccaa ggagtcaatt ttaccaaaaa gaaattcgga caagcttatt 3360

gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtggttgc taaggtggaa aaagggaaat cgaagaagtt aaaatccgtt 3480

aaagagttac tagggatcac aattatggaa agaagttcct ttgaaaaaaa tccgattgac 3540

tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600

tatagtcttt ttgagttaga aaacggtcgt aaacggatgc tggctagtgc cggagaatta 3660

caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttttata tttagctagt 3720

cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgtggag 3780

cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840

attttagcag atgccaattt agataaagtt cttagtgcat ataacaaaca tagagacaaa 3900

ccaatacgtg aacaagcaga aaatattatt catttattta cgttgacgaa tcttggagct 3960

cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020

gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080

gatttgagtc agctaggagg tgactaa 4107

<210> 109
<211> 1368
<212> PRT
<213> Artificial Sequence

<220>
<223> Amino acid sequence of Cas9 from S.pyogenes

<400> 109
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
```

```
    450                 455                 460

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

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

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

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

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

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

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

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

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605

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

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

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

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

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

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

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

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

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

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

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

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

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
                805                 810                 815
```

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

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

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

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

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

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

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

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

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

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

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

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

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

Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1025               1030                1035                1040

Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
               1045                1050                1055

Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
           1060                1065                1070

Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
       1075                1080                1085

Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
   1090                1095                1100

Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1105               1110                1115                1120

Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
               1125                1130                1135

Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
           1140                1145                1150

Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
       1155                1160                1165

Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
   1170                1175                1180
```

Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1185   1190   1195   1200

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

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

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

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

Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1265   1270   1275   1280

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
                1285   1290   1295

His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
            1300   1305   1310

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
        1315   1320   1325

Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
    1330   1335   1340

Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345   1350   1355   1360

Asp Leu Ser Gln Leu Gly Gly Asp
                1365

<210> 110
<211> 4221
<212> DNA
<213> Artificial Sequence

<220>
<223> 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)

<400> 110
atgggcagca gccatcatca tcatcatcat gtgtacccct acgacgtgcc cgactacgcc 60

gaattgcctc caaaaaagaa gagaaaggta gggatcgaga acctgtactt ccagggcgac 120

aagaagtaca gcatcggcct ggacatcggt accaacagcg tgggctgggc cgtgatcacc 180

gacgagtaca aggtgcccag caagaagttc aaggtgctgg gcaacaccga ccgccacagc 240

atcaagaaga acctgatcgg cgccctgctg ttcgacagcg gcgagaccgc cgaggccacc 300

cgcctgaagc gcaccgcccg ccgccgctac acccgccgca agaaccgcat ctgctacctg 360

caggagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca ccgcctggag 420

gagagcttcc tggtggagga ggacaagaag cacgagcgcc accccatctt cggcaacatc 480

gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgcg caagaagctg 540 gtggacagca ccgacaaggc cgacctgcgc ctgatctacc tggccctggc ccacatgatc 600

aagttccgcg gccacttcct gatcgagggc gacctgaacc ccgacaacag cgacgtggac 660

aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggagaa ccccatcaac 720

gccagcggcg tggacgccaa ggccatcctg agcgcccgcc tgagcaagag ccgccgcctg 780

gagaacctga tcgcccagct gcccggcgag aagaagaacg gcctgttcgg caacctgatc 840

gccctgagcc tgggcctgac ccccaacttc aagagcaact tcgacctggc cgaggacgcc 900

aagctgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc 960

ggcgaccagt acgccgacct gttcctggcc gccaagaacc tgagcgacgc catcctgctg 1020

agcgacatcc tgcgcgtgaa caccgagatc accaaggccc ccctgagcgc cagcatgatc 1080

aagcgctacg acgagcacca ccaggacctg accctgctga aggccctggt gcgccagcag 1140

ctgcccgaga agtacaagga gatcttcttc gaccagagca agaacggcta cgccggctac 1200

atcgacggcg gcgccagcca ggaggagttc tacaagttca tcaagcccat cctggagaag 1260

atggacggca ccgaggagct gctggtgaag ctgaaccgcg aggacctgct gcgcaagcag 1320

cgcaccttcg acaacggcag catcccccac cagatccacc tgggcgagct gcacgccatc 1380

ctgcgccgcc aggaggactt ctaccccttc ctgaaggaca accgcgagaa gatcgagaag 1440

atcctgacct tccgcatccc ctactacgtg ggcccccctgg cccgcggcaa cagccgcttc 1500

gcctggatga cccgcaagag cgaggagacc atcacccccct ggaacttcga ggaggtggtg 1560

gacaagggcg ccagcgccca gagcttcatc gagcgcatga ccaacttcga caagaacctg 1620

cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtacaac 1680

gagctgacca aggtgaagta cgtgaccgag ggcatgcgca agcccgcctt cctgagcggc 1740

gagcagaaga aggccatcgt ggacctgctg ttcaagacca accgcaaggt gaccgtgaag 1800

cagctgaagg aggactactt caagaagatc gagtgcttcg acagcgtgga gatcagcggc 1860

gtggaggacc gcttcaacgc cagcctgggc acctaccacg acctgctgaa gatcatcaag 1920

gacaaggact tcctggacaa cgaggagaac gaggacatcc tggaggacat cgtgctgacc 1980

ctgaccctgt tcgaggaccg cgagatgatc gaggagcgcc tgaagaccta cgcccacctg 2040

ttcgacgaca aggtgatgaa gcagctgaag cgccgccgct acaccggctg gggccgcctg 2100

agccgcaagc ttatcaacgg catccgcgac aagcagagcg gcaagaccat cctggacttc 2160

ctgaagagcg acggcttcgc caaccgcaac ttcatgcagc tgatccacga cgacagcctg 2220

accttcaagg aggacatcca gaaggcccag gtgagcggcc agggcgacag cctgcacgag 2280

cacatcgcca acctggccgg cagccccgcc atcaagaagg gcatcctgca gaccgtgaag 2340

gtggtggacg agctggtgaa ggtgatgggc cgccacaagc ccgagaacat cgtgatcgag 2400

atggcccgcg agaaccagac cacccagaag ggccagaaga acagccgcga gcgcatgaag 2460

cgcatcgagg agggcatcaa ggagctgggc agccagatcc tgaaggagca ccccgtggag 2520

aacacccagc tgcagaacga gaagctgtac ctgtactacc tgcagaacgg ccgcgacatg 2580 tacgtggacc aggagctgga catcaaccgc ctgagcgact acgacgtgga ccacatcgtg 2640

ccccagagct tcctgaagga cgacagcatc gacaacaagg tgctgacccg cagcgacaag 2700

aaccgcggca agagcgacaa cgtgcccagc gaggaggtgg tgaagaagat gaagaactac 2760

tggcgccagc tgctgaacgc caagctgatc acccagcgca agttcgacaa cctgaccaag 2820

gccgagcgcg gcggcctgag cgagctggac aaggccggct tcatcaagcg ccagctggtg 2880

gagacccgcc agatcaccaa gcacgtggcc cagatcctgg acagccgcat gaacaccaag 2940

tacgacgaga acgacaagct gatccgcgag gtgaaggtga tcaccctgaa gagcaagctg 3000

gtgagcgact tccgcaagga cttccagttc tacaaggtgc gcgagatcaa caactaccac 3060

cacgcccacg acgcctacct gaacgccgtg gtgggcaccg ccctgatcaa gaagtacccc 3120

aagctggaga gcgagttcgt gtacggcgac tacaaggtgt acgacgtgcg caagatgatc 3180

gccaagagcg agcaggagat cggcaaggcc accgccaagt acttcttcta cagcaacatc 3240

atgaacttct tcaagaccga gatcaccctg gccaacggcg agatccgcaa gcgccccctg 3300

atcgagacca acggcgagac cggcgagatc gtgtgggaca agggccgcga cttcgccacc 3360

gtgcgcaagg tgctgagcat gccccaggtg aacatcgtga agaagaccga ggtgcagacc 3420

ggcggcttca gcaaggagag catcctgccc aagcgcaaca gcgacaagct gatcgcccgc 3480

aagaaggact gggaccccaa gaagtacggc ggcttcgaca gccccaccgt ggcctacagc 3540

gtgctggtgg tggccaaggt ggagaagggc aagagcaaga agctgaagag cgtgaaggag 3600

ctgctgggca tcaccatcat ggagcgcagc agcttcgaga agaaccccat cgacttcctg 3660

gaggccaagg gctacaagga ggtgaagaag gacctgatca tcaagctgcc caagtacagc 3720

ctgttcgagc tggagaacgg ccgcaagcgc atgctggcca gcgccggcga gctgcagaag 3780

ggcaacgagc tggccctgcc cagcaagtac gtgaacttcc tgtacctggc cagccactac 3840

gagaagctga agggcagccc cgaggacaac gagcagaagc agctgttcgt ggagcagcac 3900

aagcactacc tggacgagat catcgagcag atcagcgagt tcagcaagcg cgtgatcctg 3960

gccgacgcca acctggacaa ggtgctgagc gcctacaaca agcaccgcga caagcccatc 4020

cgcgagcagg ccgagaacat catccacctg ttcaccctga ccaacctggg cgcccccgcc 4080

gccttcaagt acttcgacac caccatcgac cgcaagcgct acaccagcac caaggaggtg 4140

ctggacgcca ccctgatcca ccagagcatc accggtctgt acgagacccg catcgacctg 4200

agccagctgg gcggcgacta a 4221

<210> 111
<211> 1406
<212> PRT
<213> Artificial Sequence

<220>
<223> Amino acid sequence of Cas9 (pET-Cas9N3T)

<400> 111
Met Gly Ser Ser His His His His His His Val Tyr Pro Tyr Asp Val

```
  1               5                  10                  15

Pro Asp Tyr Ala Glu Leu Pro Pro Lys Lys Lys Arg Lys Val Gly Ile
             20                  25                  30

Glu Asn Leu Tyr Phe Gln Gly Asp Lys Lys Tyr Ser Ile Gly Leu Asp
         35                  40                  45

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys
     50                  55                  60

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser
 65                  70                  75                  80

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr
                 85                  90                  95

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg
            100                 105                 110

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met
        115                 120                 125

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu
    130                 135                 140

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile
145                 150                 155                 160

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu
                165                 170                 175

Arg Lys Lys Leu Val Asp Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile
            180                 185                 190

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile
        195                 200                 205

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile
    210                 215                 220

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn
225                 230                 235                 240

Ala Ser Gly Val Asp Ala Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys
                245                 250                 255

Ser Arg Arg Leu Glu Asn Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys
            260                 265                 270

Asn Gly Leu Phe Gly Asn Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro
        275                 280                 285

Asn Phe Lys Ser Asn Phe Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu
    290                 295                 300

Ser Lys Asp Thr Tyr Asp Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile
305                 310                 315                 320

Gly Asp Gln Tyr Ala Asp Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp
                325                 330                 335

Ala Ile Leu Leu Ser Asp Ile Leu Arg Val Asn Thr Glu Ile Thr Lys
            340                 345                 350

Ala Pro Leu Ser Ala Ser Met Ile Lys Arg Tyr Asp Glu His His Gln
        355                 360                 365
```

```
Asp Leu Thr Leu Leu Lys Ala Leu Val Arg Gln Gln Leu Pro Glu Lys
    370                 375                 380

Tyr Lys Glu Ile Phe Phe Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr
385                 390                 395                 400

Ile Asp Gly Gly Ala Ser Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro
                405                 410                 415

Ile Leu Glu Lys Met Asp Gly Thr Glu Glu Leu Leu Val Lys Leu Asn
            420                 425                 430

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

Pro His Gln Ile His Leu Gly Glu Leu His Ala Ile Leu Arg Arg Gln
    450                 455                 460

Glu Asp Phe Tyr Pro Phe Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys
465                 470                 475                 480

Ile Leu Thr Phe Arg Ile Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly
                485                 490                 495

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr
            500                 505                 510

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser
        515                 520                 525

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys
    530                 535                 540

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn
545                 550                 555                 560

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala
                565                 570                 575

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys
            580                 585                 590

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys
        595                 600                 605

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg
    610                 615                 620

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys
625                 630                 635                 640

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp
                645                 650                 655

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu
            660                 665                 670

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln
        675                 680                 685

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu
    690                 695                 700

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe
705                 710                 715                 720

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His
                725                 730                 735
```

```
Asp Asp Ser Leu Thr Phe Lys Glu Asp Ile Gln Lys Ala Gln Val Ser
            740                 745                 750

Gly Gln Gly Asp Ser Leu His Glu His Ile Ala Asn Leu Ala Gly Ser
        755                 760                 765

Pro Ala Ile Lys Lys Gly Ile Leu Gln Thr Val Lys Val Val Asp Glu
    770                 775                 780

Leu Val Lys Val Met Gly Arg His Lys Pro Glu Asn Ile Val Ile Glu
785                 790                 795                 800

Met Ala Arg Glu Asn Gln Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg
                805                 810                 815

Glu Arg Met Lys Arg Ile Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln
            820                 825                 830

Ile Leu Lys Glu His Pro Val Glu Asn Thr Gln Leu Gln Asn Glu Lys
        835                 840                 845

Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Arg Asp Met Tyr Val Asp Gln
    850                 855                 860

Glu Leu Asp Ile Asn Arg Leu Ser Asp Tyr Asp Val Asp His Ile Val
865                 870                 875                 880

Pro Gln Ser Phe Leu Lys Asp Asp Ser Ile Asp Asn Lys Val Leu Thr
                885                 890                 895

Arg Ser Asp Lys Asn Arg Gly Lys Ser Asp Asn Val Pro Ser Glu Glu
            900                 905                 910

Val Val Lys Lys Met Lys Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys
        915                 920                 925

Leu Ile Thr Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly
    930                 935                 940

Gly Leu Ser Glu Leu Asp Lys Ala Gly Phe Ile Lys Arg Gln Leu Val
945                 950                 955                 960

Glu Thr Arg Gln Ile Thr Lys His Val Ala Gln Ile Leu Asp Ser Arg
                965                 970                 975

Met Asn Thr Lys Tyr Asp Glu Asn Asp Lys Leu Ile Arg Glu Val Lys
            980                 985                 990

Val Ile Thr Leu Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe
        995                1000                1005

Gln Phe Tyr Lys Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp
   1010                1015                1020

Ala Tyr Leu Asn Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro
1025               1030                1035                1040

Lys Leu Glu Ser Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val
               1045                1050                1055

Arg Lys Met Ile Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala
           1060                1065                1070

Lys Tyr Phe Phe Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile
       1075                1080                1085

Thr Leu Ala Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn
```

```
   1090                1095                1100

Gly Glu Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr
1105               1110                1115                1120

Val Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
               1125                1130                1135

Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
           1140                1145                1150

Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys
       1155                1160                1165

Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val
   1170                1175                1180

Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu
1185               1190                1195                1200

Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro
               1205                1210                1215

Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu
           1220                1225                1230

Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg
       1235                1240                1245

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu
   1250                1255                1260

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1265               1270                1275                1280

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
               1285                1290                1295

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
           1300                1305                1310

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
       1315                1320                1325

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
   1330                1335                1340

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala
1345               1350                1355                1360

Ala Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
               1365                1370                1375

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
           1380                1385                1390

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
       1395                1400                1405
```