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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 January 2025 (02.01.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/006763 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/90* (2006.01)
C07K 14/22 (2006.01) *C12N 15/11* (2006.01)
C12N 15/113 (2010.01)

(21) International Application Number:
PCT/US2024/035845

(22) International Filing Date:
27 June 2024 (27.06.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/510,425 27 June 2023 (27.06.2023) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).

(72) Inventors: **ZHANG, Yan**; c/o The Regents of the University of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **CHEN, Jue**; c/o The Regents of the University of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US). **HOU, Zhonggang**; c/o The Regents of the University of Michigan, 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109-2590 (US).

(74) Agent: **BARTON, Kelly A.**; Casimir Jones, S.C., 2275 Deming Way, Suite 310, Middleton, Wisconsin 53562 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, 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, ME, 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: MODIFIED TYPE I CRISPR COMPONENTS WITH ENHANCED GENE EDITING ACTIVITY

(57) Abstract: The present invention relates to components and systems for modifying nucleic acids and gene expression. In particular, the present invention relates to engineered Cas proteins, fusion proteins and systems including the engineered Cas proteins, and methods for recruiting effector domains to target nucleic acids, modulating expression of a target gene, and altering a target nucleic acid sequence.

WO 2025/006763 A1

MODIFIED TYPE I CRISPR COMPONENTS WITH ENHANCED GENE EDITING ACTIVITY

FIELD

[001] The present invention relates to components and systems for modifying nucleic acids and gene expression. In particular, the present invention relates to engineered Cas proteins, fusion proteins and systems including the engineered Cas proteins, and methods for recruiting effector domains to target nucleic acids, modulating expression of a target gene, and altering a target nucleic acid sequence.

CROSS-REFERENCE TO RELATED APPLICATIONS

[002] This application claims the benefit of U.S. Provisional Application No. 63/510,425, filed June 27, 2023, the content of which is herein incorporated by reference in its entirety.

SEQUENCE LISTING STATEMENT

[003] The content of the electronic sequence listing titled UM_42152_601_SequenceListing.xml (Size: 42,561 bytes; and Date of Creation: June 24, 2024) is herein incorporated by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[004] This invention was made with Government support under GM137833 awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[005] CRISPR-Cas systems employ diverse RNA-guided nucleases to help microbes fend off bacteriophages and other mobile genetic elements. These systems have also been developed into tools that revolutionized genome engineering. Type I system is the most widespread and diversified type of CRISPR, which accounts for >50% of all identified CRISPR systems in nature, ten times more abundant than the currently widely used CRISPR/Cas9. Type I CRISPR is further classified into eight subtypes (I-A through I-F, I-Fv, and I-G) based on their *cas* gene composition. The Type I CRISPR DNA interference machinery consists of an RNA-guided Cascade complex for target site recognition and a helicase-nuclease enzyme Cas3 for processive target DNA degradation. Unlike widely used CRISPR/Cas9 systems that create a double-strand DNA break locally at the target site, Type I CRISPR creates targeted large chromosomal deletions in human cells. This is especially useful for the removal of disease-causing genomic loci such as toxic repeat expansions, large parasitic sequences, or integrated viral genomes. Furthermore, because Type I CRISPR can create a spectrum of large chromosomal deletions ranging from

a few hundred base pairs to up to 30-100 kb, it holds great potential as a screening tool to dissect key regulatory elements (e.g., enhancers, insulators, or repressors) along the mammalian genome. However, Type I based gene editing tools are associated with large target to target variations in editing efficiency, especially for base editing.

SUMMARY

[006] Provided herein are engineered Cas protein from a Type I CRISPR-Associated Complex for Anti-viral Defense (Cascade) complex. In some embodiments, the engineered Cas protein is from a Type I-C Cascade complex. In some embodiments, the engineered Cas protein has at least 70% identity to a wild-type protein and comprises one or more substitutions of a glutamate residue or aspartate residue within 10Å of a nucleic acid bound by the Cascade complex with a positively charged amino acid.

[007] In some embodiments, the positively charged amino acid is arginine, histidine, or lysine. In some embodiments, the positively charged amino acid is arginine.

[008] In some embodiments, the wild-type protein is a *Neisseria lactamica* (Nla) type I-C Cas protein.

[009] In some embodiments, the wild-type protein is a Cas8 protein comprising an amino acid sequence of SEQ ID NO: 1 and the one or more substitutions are selected from residues: E76, E230, E235, D268, D387, E391, D396, E426, E479, E484, D495, E524, E526, or combinations thereof, relative to SEQ ID NO: 1. In some embodiments, the one or more substitutions comprise: E76R, E230R, E235R, D268R, D387R, E391R, D396R, E426R, E479R, E484R, D495R, E524R, E526R, or combinations thereof. In some embodiments, the one or more substitutions are selected from residues: E235/D387, E235/E391, D387/E391, or E235/D387/E391. In some embodiments, the one or more substitutions comprise: E235R/D387R, E235R/E391R, D387R/E391R, or E235R/D387R/E391R.

[010] In some embodiments, the wild-type protein is a Cas5 protein comprising an amino acid sequence of SEQ ID NO: 2 and the one or more substitutions are selected from residues: E18, E22, E69, E76, E84, D85, or combinations thereof, relative to SEQ ID NO: 2. In some embodiments, the one or more substitutions comprise: E18R, E22R, E69R, E76R, E84R, D85R, or combinations thereof.

[011] In some embodiments, the wild-type protein is a Cas7 protein comprising an amino acid sequence of SEQ ID NO: 3 and the one or more substitutions are selected from residues: D23, D25, E69, D78, E90, D108, E144, E155, D157, E160, D163, or combinations thereof, relative to SEQ ID NO: 3. In some embodiments, the one or more substitutions comprise: D23R, D25R, E69R, D78R, E90R, D108R, E144R, E155R, D157R, E160R, D163R, or combinations thereof. In some embodiments, the one or more substitutions are selected from residues: E90/D78, D78/E160, E90/E160, D78/E90/E160. In some

embodiments, the one or more substitutions comprise: E90R/D78R, D78R/E160R, E90R/E160R, D78R/E90R/E160R.

[012] In some embodiments, the wild-type protein is a Cas11 protein comprising an amino acid sequence of SEQ ID NO: 4 and the one or more substitutions are selected from residues: E22, E27, D38, E67, E69, or combinations thereof, relative to SEQ ID NO: 4. In some embodiments, the one or more substitutions comprise: E22R, E27R, D38R, E67R, E69R, or combinations thereof.

[013] Also provided herein are fusion proteins comprising an engineered Cas protein as disclosed herein and at least one effector domain. In some embodiments, the at least one effector domain comprises a transcription activator, a transcription repressor, a base editor, an epigenetic modifier, or a combination thereof.

[014] Further provided are systems comprising an engineered Cas protein or fusion protein as disclosed herein. In some embodiments, the systems comprise Cas3, or a nucleic acid encoding thereof; one or more engineered Cas protein or fusion protein as disclosed herein, and/or one or more nucleic acids encoding thereof; and at least one guide RNA (gRNA), wherein each gRNA is configured to hybridize to a portion of a target nucleic acid sequence.

[015] In some embodiments, the one or more Type I Cas protein comprises a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1. In some embodiments, the one or more Type I Cas protein comprises: a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2. In some embodiments, the one or more type I Cas protein comprises a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3.

[016] In some embodiments, the one or more Type I Cas protein comprises: i) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2; ii) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3; iii) a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2; or iv) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations

thereof, relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2.

[017] In some embodiments, Cas3 is fully or partially catalytically inactive.

[018] Additionally provided are methods of using the disclosed engineered Cas protein, fusions proteins and system.

[019] In some embodiments, the methods comprise altering a target nucleic acid sequence comprising contacting a target nucleic acid sequence with a system disclosed herein. In some embodiments, the altering comprises a deletion. In some embodiments, the target nucleic acid sequence encodes a gene product. In some embodiments, the target nucleic acid sequence is a genomic DNA sequence. In some embodiments, contacting a target nucleic acid sequence comprises introducing the system into the cell.

[020] In some embodiments, the methods comprise recruiting one or more effector domains to a target nucleic acid in a cell and/or modulating expression of a target gene in a cell comprising: introducing into a cell a fusion protein or a system disclosed herein. In some embodiments, the Cas3 is fully or partially catalytically inactive. In some embodiments, the target nucleic acid comprises the promoter region or the upstream activator sequence of the target gene.

[021] In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell.

[022] In some embodiments, the introducing into the cell comprises administering the system to a subject. In some embodiments, the administering comprises in vivo administration.

[023] Other aspects and embodiments of the disclosure will be apparent in light of the following detailed description and accompanying figures.

BRIEF DESCRIPTION OF THE DRAWINGS

[024] FIGS. 1A-1D show cryo-EM maps of *Neisseria lactamica* type I-C CRISPR-Cas and its different functional states. FIG. 1A is a schematic of the *Neisseria lactamica* type I-C CRISPR-*cas* operon. FIG. 1B is a schematic of the R-loop formed by a guide RNA and the target dsDNA. FIGS. 1C and 1D are cryo-EM maps of *Neisseria lactamica* type I-C CRISPR-Cas at four different states.

[025] FIG. 2 is the structure of the *Neisseria lactamica* Type I-C cascade. All glutamate and aspartate residues of the Cas8 subunit that are within 10Å of the target strand DNA are labeled and highlighted by sphere representation.

[026] FIG. 3 is the structure of the *Neisseria lactamica* Type I-C cascade. All glutamate and aspartate residues of the Cas5 subunit that are within 10Å of the target strand DNA are labeled and highlighted by sphere representation.

[027] FIG. 4 is the structure of the *Neisseria lactamica* Type I-C cascade. All glutamate and aspartate residues of the Cas7 subunit that are within 10Å of the target strand DNA are labeled and highlighted by sphere representation.

[028] FIG. 5 is the structure of the *Neisseria lactamica* Type I-C cascade. All glutamate and aspartate residues of the Cas11 subunits that are within 10Å of the target strand DNA are labeled and highlighted by sphere representation.

[029] FIG. 6 is a bar graph showing the traditional gene deletion activity of Nla Cascade with either WT Cas8 or Cas8 mutants together with WT Cas3 and a GFP targeting crRNA in the HAP1 EGFP reporter cell line. A tdTomato expression cassette is inserted into Cas11 expression plasmid to track the transfected cell population (top right panel, grey: untransfected control, solid line: transfected cells). The activity in this graph is shown as the percentage of EGFP negative cells in the tdTomato positive population measured by flow cytometry.

[030] FIG. 7 is a bar graph showing the traditional gene deletion activity of Nla Cascade with either WT Cas8 or Cas8 mutants together with WT Cas3 and a GFP targeting crRNA in the HAP1 EGFP reporter cell line. A tdTomato expression cassette is inserted into Cas11 expression plasmid to track the transfected cell population (top right panel, grey: untransfected control, solid line: transfected cells). The activity in this graph is shown as the percentage of EGFP negative cells in the low tdTomato population measured by flow cytometry.

[031] FIG. 8 is a bar graph showing the traditional gene deletion activity of Nla Cascade with either WT Cas8 or Cas8 double mutants together with WT Cas3 and a GFP targeting crRNA in the HAP1 EGFP reporter cell line. A tdTomato expression cassette is inserted into Cas11 expression plasmid to track the transfected cell population (top right panel, grey: untransfected control, solid line: transfected cells). The activity in this graph is shown as the percentage of EGFP negative cells in the low tdTomato population (as shown in FIG. 7) measured by flow cytometry.

[032] FIG. 9 is a bar graph showing the traditional gene deletion activity of Nla Cascade with either WT Cas5 or Cas5 single mutants together with WT Cas3 and a GFP targeting crRNA in the HAP1 EGFP reporter cell line. A tdTomato expression cassette is inserted into Cas11 expression plasmid to track the transfected cell population (top right panel, grey: untransfected control, solid line: transfected cells). The

activity in this graph is shown as the percentage of EGFP negative cells in the low tdTomato population (as shown in FIG. 7) measured by flow cytometry.

[033] FIG. 10 is a bar graph showing the traditional gene deletion activity of Nla Cascade with either WT Cas7 or Cas7 mutants together with WT Cas3 and a GFP targeting crRNA in the HAP1 EGFP reporter cell line. A tdTomato expression cassette is inserted into Cas11 expression plasmid to track the transfected cell population (top right panel, grey: untransfected control, solid line: transfected cells). The activity in this graph is shown as the percentage of EGFP negative cells in the low tdTomato population (as shown in FIG. 7) measured by flow cytometry.

[034] FIGS. 11A and 11B show base editing of GFP using Nla-ABE with WT Cas8 or Cas8 mutants. FIG. 11A is target sequence of GFP (SEQ ID NOs: 28-29). PAM is underlined, guide sequence is indicated by a line above the sequence, and expected base editing window is boxed. The peak editing site is in bold and numbered. FIG. 11B is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas8 or Cas8 mutants at the EGFP locus in the 293T-AAVS1-EGFP cell line. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{46}) is shown.

[035] FIGS. 12A and 12B show base editing of HBB sickle using Nla-ABE with WT Cas8 or Cas8 mutants. FIG. 12A is the target sequence of HBB sickle (SEQ ID NOs: 30-31). PAM is underlined, guide sequence is indicated by a line above the sequence, and expected base editing window is boxed. The peak editing site is in bold and numbered. FIG. 12B is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas8 or Cas8 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (the sickle mutation, A_{43}) is shown.

[036] FIGS. 13A and 13B show base editing of HPRT1 locus using Nla-ABE with WT Cas8 or Cas8 mutants. FIG. 13A is the target sequence of HPRT1 (SEQ ID NOs: 32-33). PAM is underlined, guide sequence is indicated by a line above the sequence, and expected base editing window is boxed. The peak editing site is in bold and numbered. FIG. 13B is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas8 or Cas8 mutants at the HPRT1 locus in the 293T-AAVS1-EGFP cell line. Base conversion efficiencies were measured via amplicon sequencing and

calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[037] FIG. 14 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas5 or Cas5 mutants at the EGFP locus in the 293T-AAVS1-EGFP cell line. See FIG. 11A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{46}) is shown.

[038] FIG. 15 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas5 or Cas5 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown. The experiment was done in the 293T-HBB sickle cell line.

[039] FIG. 16 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas5 or Cas5 mutants at the HPRT1 locus in the 293T-AAVS1-EGFP cell line. See FIG. 13A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[040] FIG. 17 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas7 or Cas7 mutants at the EGFP locus in the 293T-AAVS1-EGFP cell line. See FIG. 11A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{46}) is shown.

[041] FIG. 18 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas7 or Cas7 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[042] FIG. 19 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas11 or Cas11 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and

calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[043] FIG. 20 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas7 or Cas7 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[044] FIG. 21 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT Cas8 or Cas8 mutants at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[045] FIG. 22 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[046] FIG. 23 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[047] FIG. 24 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[048] FIG. 25 is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the HBB sickle locus in the 293T-HBB sickle cell line. See FIG. 12A for sequence information. Base conversion efficiencies were measured via amplicon sequencing and

calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[049] FIGS. 26A and 26B show base editing of CFTR locus using Nla-ABE with WT Cas or Cas mutants. FIG. 26A is target sequence of CFTR (SEQ ID NOs: 34-35). PAM is underlined, guide sequence is indicated by a line above the sequence, and expected base editing window is boxed. The peak editing site is in bold and numbered. FIG. 26B is a representative bar graph showing the A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the CFTR locus in 293T-AAVS1-EGFP cells. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. The efficiency of the position with the highest A•T to G•C conversion (A_{43}) is shown.

[050] FIGS. 27A-27D show correction of CFTR G542X in iPS cells using Cas variants. FIG. 27A is a diagram showing the CFTR G542X mutation (A_{48} in the boxed region), location of the guide RNA (41nt, the line above the DNA sequence) and PAM (underlined) used, potential target adenosine residues (bold) and potential amino acid changes after edit. Expected editing window is boxed with dashed line. FIG. 27B is a representative bar graph showing the on target (A_{48}) A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the CFTR locus in iPS cells with CFTR G542X mutation. FIG. 27C is a representative bar graph showing the total indel (insertion and deletion) percentage at the CFTR locus in the cells edited by Nla-ABE with WT or mutant Cas subunits. FIG. 27D is a representative bar graph showing both the on target (A_{48}) and bystander A•T to G•C conversion efficiency of Nla-ABE with WT or mutant Cas subunits at the CFTR locus in iPS cells with CFTR G542X mutation. Base conversion efficiencies were measured via amplicon sequencing and calculated as the percentage of total reads with A•T to G•C edits at a specific location. Total indel rates were measured via amplicon sequencing and calculated as the percentage of total reads with insertions or deletions around the editing site. Abbreviations: NT: non-transfected. Cas8 DM1: Cas8 E235R/D387R; Cas8 DM2: Cas8 E235R/E391R; Cas8 DM3: Cas8 D387R/E391R; Cas8 TM: Cas8 E235R/D387R-E391R; Cas7DM3: E90R/E160R.

[051] FIGS. 28A and 28B show specificity of Cas8 variants in gene editing at the EGFP site. FIG. 28A is a diagram showing the WT and mismatched EGFP guides (SEQ ID NOs: 16-19). The Δ9nt *EGFP* guide has the deletion of nine nucleotides distal to PAM. The 1st nt mm *EGFP* guide has the 1st nucleotide proximal to PAM being mismatched. The 4 nt mm *EGFP* guide has the 13th to 16th nucleotides mismatched. FIG. 28B is a bar graph showing the gene deletion activity of Nla Cascade with either WT Cas8 or Cas8 activity enhancing mutants together with WT Cas3 and GFP targeting crRNA with various

mismatches in the HAP1 EGFP reporter cell line. The activity is shown as the percentage of EGFP negative cells in the total population measured by flow cytometry.

[052] FIG. 29 is a representative bar graph showing the traditional gene deletion activity of Nla Cascade with various Cas mutants together with WT Cas3 and a crRNA targeting the 3' UTR of GFP utilizing CTC PAM in the 293 EGFP reporter cell line. The activity in this graph is shown as the percentage of EGFP negative cells in the total population measured by flow cytometry. Abbreviations: NT: non-transfected. Cas8 DM1: Cas8 E235R/D387R; Cas8 DM2: Cas8 E235R/E391R; Cas8 DM3: Cas8 D387R/E391R; Cas8 TM: Cas8 E235R/D387R/E391R; Cas7DM3: E90R/E160R.

DETAILED DESCRIPTION OF THE INVENTION

[053] The present disclosure is directed to an engineered Type I CRISPR system and components thereof with enhanced affinity to target DNA. As described herein, single amino acid substitutions were installed in the Cascade subunits of Cas5, Cas7, Cas8, and Cas11 from *Neisseria lactamica* type I-C CRISPR-Cas system. These substitutions replace negatively charged residues (e.g., Glu and Asp) on Cascade subunits within about 10Å of the target strand DNA with positively charged residues (e.g., Arg) and enhance both the gene deletion and the base editing efficiency across multiple genomic sites. Further enhancement of activity was observed when individual substitutions are combined. Thus, the engineered systems and components described herein show decreased target to target variations resulting in more reliable and efficient genome engineering tools.

1. Definitions

[054] To facilitate an understanding of the present technology, a number of terms and phrases are defined below. Additional definitions are set forth throughout the detailed description.

[055] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. As used herein, comprising a certain sequence or a certain SEQ ID NO usually implies that at least one copy of said sequence is present in recited peptide or polynucleotide. However, two or more copies are also contemplated. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments "comprising," "consisting of" and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[056] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[057] Unless otherwise defined herein, scientific, and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. The meaning and scope of the terms should be clear; in the event, however of any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[058] As used herein, a "nucleic acid" or a "nucleic acid sequence" refers to a polymer or oligomer of pyrimidine and/or purine bases, preferably cytosine, thymine, and uracil, and adenine and guanine, respectively (See Albert L. Lehninger, *Principles of Biochemistry*, at 793-800 (Worth Pub. 1982)). The present technology contemplates any deoxyribonucleotide, ribonucleotide, or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated, or glycosylated forms of these bases, and the like. The polymers or oligomers may be heterogenous or homogenous in composition, and may be isolated from naturally occurring sources or may be artificially or synthetically produced. In addition, the nucleic acids may be DNA or RNA, or a mixture thereof, and may exist permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states. In some embodiments, a nucleic acid or nucleic acid sequence comprises other kinds of nucleic acid structures such as, for instance, a DNA/RNA helix, peptide nucleic acid (PNA), morpholino nucleic acid (see, e.g., Braasch and Corey, *Biochemistry*, *41*(14): 4503-4510 (2002)) and U.S. Patent 5,034,506), locked nucleic acid (LNA; see Wahlestedt et al., *Proc. Natl. Acad. Sci. U.S.A.*, *97*: 5633-5638 (2000)), cyclohexenyl nucleic acids (see Wang, *J. Am. Chem. Soc.*, *122*: 8595-8602 (2000)), and/or a ribozyme. Hence, the term "nucleic acid" or "nucleic acid sequence" may also encompass a chain comprising non-natural nucleotides, modified nucleotides, and/or non- nucleotide building blocks that can exhibit the same function as natural nucleotides (e.g., "nucleotide analogs"); further, the term "nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin, which may be single or double-stranded, and represent the sense or antisense strand. The terms "nucleic acid," "polynucleotide," "nucleotide sequence," and "oligonucleotide" are used interchangeably. They

refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof.

[059] The terms "complementary" and "complementarity" refer to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base-paring or other non-traditional types of pairing. The degree of complementarity between two nucleic acid sequences can be indicated by the percentage of nucleotides in a nucleic acid sequence which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 50%, 60%, 70%, 80%, 90%, and 100% complementary). Two nucleic acid sequences are "perfectly complementary" if all the contiguous nucleotides of a nucleic acid sequence will hydrogen bond with the same number of contiguous nucleotides in a second nucleic acid sequence. Two nucleic acid sequences are "substantially complementary" if the degree of complementarity between the two nucleic acid sequences is at least 60% (e.g., 65%, 70%, 75%, 80%, 85%, 90%, 95%. 97%, 98%, 99%, or 100%) over a region of at least 8 nucleotides (e.g., 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides), or if the two nucleic acid sequences hybridize under at least moderate, preferably high, stringency conditions. Exemplary moderate stringency conditions include overnight incubation at 37° C in a solution comprising 20% formamide, 5×SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1×SSC at about 37-50° C, or substantially similar conditions, e.g., the moderately stringent conditions described in Sambrook et al., *infra*. High stringency conditions are conditions that use, for example (1) low ionic strength and high temperature for washing, such as 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate (SDS) at 50° C, (2) employ a denaturing agent during hybridization, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin (BSA)/0.1% Ficoll/0.1% polyvinylpyrrolidone (PVP)/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride and 75 mM sodium citrate at 42° C, or (3) employ 50% formamide, 5×SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5×Denhardt's solution, sonicated salmon sperm DNA (50 μg/ml), 0.1% SDS, and 10% dextran sulfate at 42° C, with washes at (i) 42° C in 0.2×SSC, (ii) 55° C in 50% formamide, and (iii) 55° C in 0.1×SSC (preferably in combination with EDTA). Additional details and an explanation of stringency of hybridization reactions are provided in, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 3rd ed., Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

(2001); and Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates and John Wiley & Sons, New York (1994).

[060] As used herein, the term "percent sequence identity" refers to the percentage of nucleotides or nucleotide analogs in a nucleic acid sequence, or amino acids in an amino acid sequence, that is identical with the corresponding nucleotides or amino acids in a reference sequence after aligning the two sequences. In case a nucleic acid according to the technology is longer than a reference sequence, additional nucleotides in the nucleic acid, that do not align with the reference sequence, are not taken into account for determining sequence identity. Methods and computer programs for alignment are well known in the art, including BLAST, Align 2, and FASTA.

[061] As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (e.g., the strength of the association between the nucleic acids) is influenced by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, and the T_m of the formed hybrid. Hybridization methods involve the annealing of one nucleic acid to another, complementary nucleic acid, e.g., a nucleic acid having a complementary nucleotide sequence. The ability of two polymers of nucleic acid containing complementary sequences to find each other and "anneal" or "hybridize" through base pairing interaction is a well-recognized phenomenon. The initial observations of the "hybridization" process by Marmur and Lane, *Proc. Natl. Acad. Sci. USA*, *46*: 453 (1960) and Doty et al., *Proc. Natl. Acad. Sci. USA*, *46*: 461 (1960), have been followed by the refinement of this process into an essential tool of modern biology. For example, hybridization and washing conditions are now well known and exemplified in Sambrook et al., *supra*. The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

[062] As used herein, a "double-stranded nucleic acid" may be a portion of a nucleic acid, a region of a longer nucleic acid, or an entire nucleic acid. A "double-stranded nucleic acid" may be, e.g., without limitation, a double-stranded DNA, a double-stranded RNA, a double-stranded DNA/RNA hybrid, etc. A single-stranded nucleic acid having secondary structure (e.g., base-paired secondary structure) and/or higher order structure (e.g., a stem-loop structure) may also be considered a "double-stranded nucleic acid." For example, triplex structures are considered to be "double-stranded." In some embodiments, any base-paired nucleic acid is a "double-stranded nucleic acid."

[063] The term "gene" refers to a DNA sequence that comprises control and coding sequences necessary for the production of an RNA having a non-coding function (e.g., a ribosomal or transfer RNA), a polypeptide, or a precursor of any of the foregoing. The RNA or polypeptide can be encoded by a full

length coding sequence or by any portion of the coding sequence so long as the desired activity or function is retained. Thus, a "gene" refers to a DNA or RNA, or portion thereof, that encodes a polypeptide or an RNA chain that has functional role to play in an organism. For the purpose of this disclosure, it may be considered that genes include regions that regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites, and locus control regions.

[064] The term "wild-type" refers to a gene or a gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "engineered," "modified," "mutant," or "polymorphic" refers to a gene or gene product that displays modifications in sequence and or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

[065] As used herein, the term "variant" refers to the exhibition of qualities that have a pattern that deviates from what occurs in nature. In some embodiments, a variant may also be a mutant.

[066] The terms "non-naturally occurring," "engineered," and "synthetic" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature.

[067] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[068] "Binding" as used herein (e.g., with reference to an NA-binding domain of a polypeptide) refers to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e.g., when a molecule X is said to interact with a molecule Y, it is meant the molecule X binds

to molecule Y in a non-covalent manner). Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence specific. Binding interactions are generally characterized by a dissociation constant (K_d) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. "Affinity" refers to the strength of binding, increased binding affinity being correlated with a lower K_d.

[069] By "binding domain" it is meant a protein domain that is able to bind non-covalently to another molecule. A binding domain can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein binding protein). In the case of a protein domain-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins.

[070] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, e.g., an "insert," may be attached or incorporated so as to bring about the replication of the attached segment in a cell.

[071] A cell has been "genetically modified," "transformed," or "transfected" by exogenous DNA, e.g., a recombinant expression vector, when such DNA has been introduced inside the cell. The presence of exogenous DNA results in permanent or transient genetic change. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones that comprise a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

[072] A "subject" or "patient" may be human or non-human and may include, for example, animal strains or species used as "model systems" for research purposes, such a mouse model as described herein. Likewise, patient may include either adults, juveniles (e.g., children), or infants. Moreover, patient may mean any living organism, preferably a mammal (e.g., humans and non-humans) that may benefit from the administration of compositions contemplated herein. Examples of mammals include, but are not limited to, any member of the Mammalian class: humans, non-human primates such as chimpanzees, and

other apes and monkey species; farm animals such as cattle, horses, sheep, goats, swine; domestic animals such as rabbits, dogs, and cats; laboratory animals including rodents, such as rats, mice, and guinea pigs, and the like. Examples of non-mammals include, but are not limited to, birds, fish, and the like. In one embodiment, the mammal is a human.

[073] The term "contacting" as used herein refers to bring or put in contact, to be in or come into contact. The term "contact" as used herein refers to a state or condition of touching or of immediate or local proximity. Contacting to a target destination, such as, but not limited to, an organ, tissue, cell, or tumor, may occur by any means of administration known to the skilled artisan.

[074] As used herein, the terms "providing," "administering," and "introducing," are used interchangeably herein and refer to the placement of the proteins or systems of the disclosure into a subject by a method or route which results in at least partial localization to a desired site. Administration can use any appropriate route which results in delivery to a desired location in the subject.

[075] Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present disclosure. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

2. CRISPR/Cas

a) Engineered Cas Proteins

[076] In bacteria and archaea, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) systems provide immunity by incorporating fragments of invading phage, virus, and plasmid DNA into CRISPR loci and using corresponding CRISPR RNAs ("crRNAs") to guide the degradation of homologous sequences. Transcription of a CRISPR locus produces a "pre-crRNA," which is processed to yield crRNAs containing spacer-repeat fragments that guide effector nuclease complexes to cleave dsDNA sequences complementary to the spacer. Several different types of CRISPR systems are known, (e.g., type I, type II, or type III), and classified based on the Cas protein type and the use of a proto-spacer-adjacent motif (PAM) for selection of proto-spacers in invading DNA.

[077] Engineering CRISPR/Cas components for use in eukaryotic cells may require modification of Cas proteins, for example to increase efficiency and decrease off-target effects. Provided herein are engineered Cas proteins from a Type I CRISPR-Associated Complex for Anti-viral Defense (Cascade)

complex which increase DNA interactions of Cascade. The terms "Cascade (CRISPR-Associated Complex for Anti-viral Defense)" or "Cascade complex" as used herein, refer to a ribonucleoprotein complex comprised of multiple protein subunits (e.g., Cas proteins). The Cascade complex recognizes nucleic acid targets via direct base-pairing to guide RNA contained in the complex.

[078] The engineered Cas proteins have at least 70% identity (e.g., at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity) to a wild-type protein and include one or more substitutions of a glutamate residue or aspartate residue (e.g., within 10Å of a nucleic acid bound by the Cascade complex) with a positively charged amino acid (e.g., arginine, histidine, lysine). In certain embodiments, the glutamate or aspartate residue is substituted with an arginine. In some embodiments, one, two, three, four, five, six, seven, eight, nine, ten, or more, or all glutamate or aspartate residues are replaced with a positively charged amino acid. In some embodiments, one, two, three, four, five, six, seven, eight, nine, ten, or more, or all glutamate or aspartate residues within 10Å of a nucleic acid bound by the Cascade complex are replaced with a positively charged amino acid. In some embodiments, no glutamate or aspartate residue outside of 10Å of a nucleic acid bound by the Cascade complex are replaced with a positively charged amino acid.

[079] Identification of those residues within 10Å of a nucleic acid can be completed using detailed analysis of structures of the bound complex with DNA at various stages and conformations. For example, as described herein, high resolution cryo-EM structures Cascade at different stages of target recognition can be used to identify negatively charged residues in proximity to the bound DNA.

[080] The engineered Cas proteins may increase DNA binding or association, increase activity, or increase stability or kinetics of Cascade complex formation. Accordingly, the engineered Cas proteins or complexes and systems comprising them may show increased activity (e.g., gene deletion and the base editing efficiency) and/or decreased off-target or un-desired effects.

[081] In some embodiments, the wild-type Cas protein is a Cas protein from a *Neisseria* species (e.g., *Neisseria lactamica*). The genus *Neisseria* comprises many gram-negative β-proteobacteria that interact with eukaryotic hosts, but only two organisms, the *gonococcus* (Gc) and its close relative the *meningococcus* (Mc), are human pathogens, both of which colonize mucosal surfaces. Many non-pathogenic *Neisseria* species also colonize the human nasopharynx, and among them *N. lactamica* is the most widely studied commensal bacterium. In some embodiments, the wild-type Cas protein used in the

context of the present disclosure is derived from the Type I-C system of *Neisseria lactamica* (Nla), and as such the engineered Cas protein is derived from the Type I-C system of *Neisseria lactamica* (Nla).

[082] In some embodiments, the Cas protein is a Type I-C Cas protein. In some embodiments, the Cas protein is Cas8, Cas5, Cas7, or Cas11.

[083] In some embodiments, the wild-type protein is a Cas8 protein having an amino acid sequence of SEQ ID NO: 1. In some embodiments, the one or more substitutions are selected from residues: E76, E230, E235, D268, D387, E391, D396, E426, E479, E484, D495, E524, E526, or combinations thereof, relative to SEQ ID NO: 1. Thus, provided herein is a Cas8 protein comprising one or more substitutions of residues: E76, E230, E235, D268, D387, E391, D396, E426, E479, E484, D495, E524, E526, or combinations thereof, relative to SEQ ID NO: 1. The one or more substitutions may comprise any combination of the listed residues, for example, E230 and E235, E230 and D268, E230 and D387, E230 and E391, E230 and D396, E230 and E426, E230 and E479, E230 and E484, E230 and D495, E230 and E524, E230 and E526, E235 and D268, E235 and D387, E235 and E391, E235 and D396, E235 and E426, E235 and E479, E235 and E484, E235 and D495, E235 and E524, E235 and E526, D268 and D387, D268 and E391, D268 and D396, D268 and E426, D268 and E479, D268 and E484, D268 and D495, D268 and E524, D268 and E526, D387 and E391, D387 and D396, D387 and E426, D387 and E479, D387 and E484, D387 and D495, D387 and E524, D387 and E526, E391 and D396, E391 and E426, E391 and E479, E391 and E484, E391 and D495, E391 and E524, E391 and E526, D396 and E426, D396 and E479, D396 and E484, D396 and D495, D396 and E524, D396 and E526, E426 and E479, E426 and E484, E426 and D495, E426 and E524, E426 and E526, E479 and E484, E479 and D495, E479 and E524, E479 and E526, E484 and D495, E484 and E524, E484 and E526, D495 and E524, D495 and E526, E524 and E526, or combinations thereof. In select embodiments, the one or more substitutions are of residues: E235 and D387 (shown as E235/D387); E235 and E391; D387 and E391; or E235, D387 and E391, relative to SEQ ID NO: 1.

[084] In select embodiments, the one or more substitutions comprise: E76R, E230R, E235R, D268R, D387R, E391R, D396R, E426R, E479R, E484R, D495R, E524R, E526R, or combinations thereof, relative to SEQ ID NO: 1. In select embodiments, the one or more substitutions comprise: E235R and D387R (represented as E235R/D387R); E235R and E391R; D387R and E391R; or E235R, D387R, and E391R.

[085] In some embodiments, the wild-type protein is a Cas5 protein having an amino acid sequence of SEQ ID NO: 2. In some embodiments, the one or more substitutions are selected from residues: E18, E22,

E69, E76, E84, D85, or combinations thereof, relative to SEQ ID NO: 2. Thus, provided herein is a Cas5 protein comprising one or more substitutions of residues: E18, E22, E69, E76, E84, D85, or combinations thereof, relative to SEQ ID NO: 2. The one or more substitutions may comprise any combination of the listed residues, for example, E18 and E22, E18 and E69, E18 and E76, E18 and E84, E18 and D85, E22 and E69, E22 and E76, E22 and E84, E22 and D85, E69 and E76, E69 and E84, E69 and D85, E76 and E84, E76 and D85, E84 and D85, or combinations thereof. In select embodiments, the one or more substitutions comprise: E18R, E22R, E69R, E76R, E84R, D85R, or combinations thereof.

[086] In some embodiments, the wild-type protein is a Cas7 protein comprising an amino acid sequence of SEQ ID NO: 3. In some embodiments, the one or more substitutions are selected from residues: D23, D25, E69, D78, E90, D108, E144, E155, D157, E160, D163, or combinations thereof, relative to SEQ ID NO: 3. Thus, provided herein is a Cas7 protein comprising one or more substitutions of residues: D23, D25, E69, D78, E90, D108, E144, E155, D157, E160, D163, or combinations thereof, relative to SEQ ID NO: 3. The one or more substitutions may comprise any combination of the listed residues, for example, D23 and D25, D23 and E69, D23 and D78, D23 and E90, D23 and D108, D23 and E144, D23 and E155, D23 and D157, D23 and E160, D23 and D163, D25 and E69, D25 and D78, D25 and E90, D25 and D108, D25 and E144, D25 and E155, D25 and D157, D25 and E160, D25 and D163, E69 and D78, E69 and E90, E69 and D108, E69 and E144, E69 and E155, E69 and D157, E69 and E160, E69 and D163, D78 and E90, D78 and D108, D78 and E144, D78 and E155, D78 and D157, D78 and E160, D78 and D163, E90 and D108, E90 and E144, E90 and E155, E90 and D157, E90 and E160, E90 and D163, D108 and E144, D108 and E155, D108 and D157, D108 and E160, D108 and D163, E144 and E155, E144 and D157, E144 and E160, E144 and D163, E155 and D157, E155 and E160, E155 and D163, D157 and E160, D157 and D163, E160 and D163, or combinations thereof. In select embodiments, the one or more substitutions comprise: D23R, D25R, E69R, D78R, E90R, D108R, E144R, E155R, D157R, E160R, D163R, or combinations thereof.

[087] In some embodiments, the one or more substitutions are selected from residues: E90 and D78 (shown as E90/D78); D78 and E160; E90 and E160; D78, E90, and E160, relative to SEQ ID NO: 3. In select embodiments, the one or more substitutions comprise: E90R and D78R (represented as E90R/D78R); D78R and E160R; E90R and E160R; D78R, E90R, and E160R.

[088] In some embodiments, the wild-type protein is a Cas11 protein comprising an amino acid sequence of SEQ ID NO: 4 and the one or more substitutions are of residues: E22, E27, D38, E67, E69, or combinations thereof, relative to SEQ ID NO: 4. Thus, provided herein is a Cas11 protein comprising one

or more substitutions of residues: E22, E27, D38, E67, E69, or combinations thereof, relative to SEQ ID NO: 4. The one or more substitutions may comprise any combination of the listed residues, for example, E22 and E27, E22 and D38, E22 and E67, E22 and E69, E27 and D38, E27 and E67, E27 and E69, D38 and E67, D38 and E69, E67 and E69, or combinations thereof. In select embodiments, the one or more substitutions comprise: E22R, E27R, D38R, E67R, E69R, or combinations thereof.

[089] In some embodiments, the Type I-C Cas protein is derived from a *Bacillus* species (e.g., *Bacillus halodurans* (Bha)) system, or variants thereof. The genus *Bacillus* is a diverse group of spore-forming bacteria ubiquitous in the environment. *Bacillus anthracis*, the agent of anthrax, is the only obligate *Bacillus* pathogen in vertebrates. *Bacillus larvae, B. lentimorbus, B. popilliae, B. sphaericus,* and *B. thuringiensis* are pathogens of specific groups of insects. A number of other species, in particular *B. cereus*, are occasional pathogens of humans and livestock, but the large majority of *Bacillus* species are harmless saprophytes. Thus, the vast majority of *Bacillus* are nonpathogenic, environmental organisms found in soil, air, dust, and debris. In some embodiments, the Type I-C Cas protein is derived from the Type I-C system of *Bacillus halodurans* (Bha), or variants thereof.

[090] In some embodiments, the wild-type protein is a Cas8 (Csd1) protein having an amino acid sequence of SEQ ID NO: 8. In some embodiments, the one or more substitutions are of residue selected from: D305, E440, D445, D528, or combinations thereof, relative to SEQ ID NO: 8.

[091] In some embodiments, the wild-type protein is a Cas5 protein having an amino acid sequence of SEQ ID NO: 7. In some embodiments, the one or more substitutions include residue E26 to SEQ ID NO: 7.

[092] In some embodiments, the wild-type protein is a Cas7 (Csd2) protein having an amino acid sequence of SEQ ID NO: 6. In some embodiments, the one or more substitutions include residues: D23, D109, D160, or combinations thereof, relative to SEQ ID NO: 6.

[093] In some embodiments, the wild-type protein is a Cas11 protein having an amino acid sequence of SEQ ID NO: 9. In some embodiments, the one or more substitutions include residue D21, relative to SEQ ID NO: 9.

[094] In some embodiments, the Type I-C Cas protein is derived from a *Desulfovibrio* species (e.g., *Desulfovibrio vulgaris* (Dvu)) system, or variants thereof. Desulfovibrio is a genus of Gram-negative sulfate-reducing bacteria commonly found in aquatic environments. In some embodiments, the CRISPR-Cas system used in the context of the present disclosure is derived from the Type I-C system of *Desulfovibrio vulgaris* (Dvu), or variants thereof.

[095] In some embodiments, the wild-type protein is a Cas8 protein having an amino acid sequence of SEQ ID NO: 12. In some embodiments, the one or more substitutions include residues: E225, D230, D265, E416, E451, E509, D514, D525, E558, or combinations thereof, relative to SEQ ID NO: 12.

[096] In some embodiments, the wild-type protein is a Cas5 protein having an amino acid sequence of SEQ ID NO: 11. In some embodiments, the one or more substitutions include residues: E30, E77, D93, combinations thereof, relative to SEQ ID NO: 11.

[097] In some embodiments, the wild-type protein is a Cas7 protein having an amino acid sequence of SEQ ID NO: 13. In some embodiments, the one or more substitutions include residues: D24, D26, E71, E80, E151, D115, E164, D171, or combinations thereof, relative to SEQ ID NO: 13.

[098] In some embodiments, the wild-type protein is a Cas11 protein having an amino acid sequence of SEQ ID NO: 14. In some embodiments, the one or more substitutions include residues: E21, D26, D37, E70, or combinations thereof, relative to SEQ ID NO: 14.

b) Fusion proteins

[099] The disclosure also provides fusion proteins comprising, consisting of, or consisting essentially of any of the engineered Cas proteins disclosed herein and at least one effector domain. The fusion protein may comprise more than one (e.g., 2, 3, 4, 5, or more) effector domains.

[0100] Effector domains encompass any protein or fragments thereof that can modify, regulate, or tag a target nucleic acid. The effector domain may comprise a number of functionalities, including but not limited to, nuclease function, recombinase function, epigenetic modifying function, transposase function, integrase function, resolvase function, invertase function, protease function, DNA methyltransferase function, DNA demethylase function, histone acetylase function, histone deacetylase function, transcriptional repressor function, transcriptional activator function, DNA binding protein function, transcription factor recruiting protein function, nuclear-localization signal function, DNA editing function (e.g., deaminase) or any combination thereof. For example, some effector domains function in transcriptional regulation via their ability to interact with the basal transcriptional machinery and general co-activators, interact with other transcription factors to allow cooperative binding, and/or directly or indirectly recruit histone and chromatin modifying enzymes. In select embodiments, the at least one effector domain includes a transcription activator, a transcription repressor, a base editor, an epigenetic modifier, or a combination thereof. In some embodiments, any additional domains or proteins necessary for the functionality of the effector domain may be provided separately or as a fusion to another associated protein (e.g., another Cas protein).

[0101] In some embodiments, the effector domains are fragments of proteins that have been separated from their natural DNA binding domains and engineered to be part of a fusion protein with the Cas proteins described herein. In some embodiments, the effector domains are proteins which normally bind to other proteins or factors which result in their recruitment to a specific or non-specific nucleic acid.

[0102] The at least one effector domain may be appended to the N-terminus or the C-terminus of the Cas protein. When two or more effector domains are appended to the Cas protein, each may be individually N-terminal or C-terminal of the Cas protein, such that the Cas protein may be flanked by effector domains on its N- and C-terminus, or all the effector domains may be linking consecutively to either the N- or C-terminus. The at least one effector domain may be fused in any orientation in relationship to the Cas protein, for example, N-terminus to C-terminus, N-terminus to N-terminus, and C-terminus to C-terminus.

[0103] The at least one effector domain may be appended to the Cas protein by a linker. The linker may have any of a variety of amino acid sequences. Suitable linkers include polypeptides of between 4 amino acids and 40 amino acids in length, or between 4 amino acids and 25 amino acids in length. These linkers can be produced by using synthetic, linker-encoding oligonucleotides to couple the proteins, or can be encoded by a nucleic acid sequence encoding the fusion protein. Peptide linkers with a degree of flexibility can be used. The linking peptides may have virtually any amino acid sequence, bearing in mind that the preferred linkers will have a sequence that results in a generally flexible peptide. Small amino acids, such as glycine and alanine, are generally used in creating a flexible peptide. A variety of different linkers are commercially available and are considered suitable for use, including but not limited to, glycine-serine polymers, glycine-alanine polymers, and alanine-serine polymers.

[0104] Any of the engineered Cas proteins, or fusion proteins thereof, described herein may comprise one or more amino acid substitutions as compared to the corresponding wild-type protein in addition to those expressly recited above. Any of the engineered Cas proteins, or fusion proteins thereof, described herein may also comprise one or more amino acid deletions or additions as compared to the corresponding wild-type protein.

[0105] An amino acid "replacement" or "substitution" refers to the replacement of any one amino acid at a given position or residue by another amino acid at the same position or residue within a polypeptide sequence. Amino acids are broadly grouped as "aromatic" or "aliphatic." An aromatic amino acid includes an aromatic ring. Examples of "aromatic" amino acids include histidine (H or His), phenylalanine (F or Phe), tyrosine (Y or Tyr), and tryptophan (W or Trp). Non- aromatic amino acids are broadly grouped as "aliphatic." Examples of "aliphatic" amino acids include glycine (G or Gly), alanine

(A or Ala), valine (V or Val), leucine (L or Leu), isoleucine (I or He), methionine (M or Met), serine (S or Ser), threonine (T or Thr), cysteine (C or Cys), proline (P or Pro), glutamic acid (E or Glu), aspartic acid (A or Asp), asparagine (N or Asn), glutamine (Q or Gin), lysine (K or Lys), and arginine (R or Arg).

[0106] The amino acid replacement or substitution can be conservative, semi-conservative, or non-conservative. The phrase "conservative amino acid substitution" or "conservative mutation" refers to the replacement of one amino acid by another amino acid with a common property. A functional way to define common properties between individual amino acids is to analyze the normalized frequencies of amino acid changes between corresponding proteins of homologous organisms (Schulz and Schirmer, Principles of Protein Structure, Springer-Verlag, New York (1979)). According to such analyses, groups of amino acids may be defined where amino acids within a group exchange preferentially with each other, and therefore resemble each other most in their impact on the overall protein structure (Schulz and Schirmer, supra). Examples of conservative amino acid substitutions include substitutions of amino acids within the sub-groups described above, for example, lysine for arginine and vice versa such that a positive charge may be maintained, glutamic acid for aspartic acid and vice versa such that a negative charge may be maintained, serine for threonine such that a free -OH can be maintained, and glutamine for asparagine such that a free -NH2 can be maintained. "Semi-conservative mutations" include amino acid substitutions of amino acids within the same groups listed above, but not within the same sub-group. For example, the substitution of aspartic acid for asparagine, or asparagine for lysine, involves amino acids within the same group, but different sub-groups. "Non-conservative mutations" involve amino acid substitutions between different groups, for example, lysine for tryptophan, or phenylalanine for serine, etc.

c) Systems

[0107] Engineering CRISPR/Cas systems for use in eukaryotic cells may also involve reconstitution of the CRISPR/Cas complex. Typically, the RNA sequences necessary for CRISPR/Cas systems are referred to collectively as "guide RNA" (gRNA) or single guide RNA (sgRNA). Thus, the terms "guide RNA," "single guide RNA," and "synthetic guide RNA," are used interchangeably herein and may refer to a nucleic acid sequence comprising a tracrRNA and a pre-crRNA array containing a guide sequence. The terms "guide sequence," "guide," and "spacer" are used interchangeably herein and refer to the nucleotide sequence within a guide RNA that specifies the target site.

[0108] The system disclosed herein comprises an engineered Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated (Cas) (CRISPR-Cas) system, and/or one or more nucleic acids encoding the engineered CRISPR-Cas system, wherein the engineered CRISPR-Cas system

comprises: (a) Cas3; (b) one or more engineered Cas protein or fusion protein as disclosed herein; and (c) at least one guide RNA (gRNA), wherein each gRNA is configured to hybridize to a portion of a target nucleic acid sequence.

[0109] Target recognition by Cascade results in a conformational change which facilitates recruitment of a protein component referred to as Cas3. Cas3 may comprise a single protein unit which contains helicase and nuclease domains. After target validation by Cascade, Cas3 nicks the strand of DNA that is looped out by the R-loop formed by Cascade approximately 9-12 nucleotides inward from the PAM site. Cas3 then uses its helicase/nuclease activity to processively degrade substrate nucleic acids, moving in a 3' to 5' direction.

[0110] Variants of Cas3 can be obtained by disabling the functional activity of one or both domains of Cas3. Disabling the ATPase dependent helicase activity by deletion, knockout of the Cas3-helicase domain, through mutagenesis of critical residues (e.g., active site residues), or by assembling the reaction in the absence of ATP can modify the Cas3 endonuclease into a nickase as the protein retains a functional nuclease domain but no longer has the ability to unwind DNA to form a ssDNA substrate. Disabling the nuclease activity can be accomplished by any method known in the art, such as but not limited to, mutagenesis of critical residues of the catalytic acid site of the nuclease domain. Disabling both the helicase and nuclease activities, to create a catalytically inactive or dead form of Cas3, converts the Cas3 endonuclease into nucleic acid binding protein via Cascade. In some embodiments, Cas3 is catalytically active, having full catalytic activity from both the nuclease and helicase domains. In some embodiments, Cas3 is fully or partially catalytically inactive, for example, lacking a functional helicase domain or both the helicase and nuclease domains.

[0111] In some embodiments, the Type I CRISPR-Cas system is a Type I-C system. Elements or sequences from any suitable Type I-C CRISPR-Cas system may be used in the context of the disclosed system. See for example, those Cas proteins and system described in International Application No. PCT/US2022/031091.

[0112] In some embodiments, the system may be derived from CRISPR-Cas elements (e.g., Cascade-Cas3 proteins or variants thereof) from a *Neisseria* species (e.g., *Neisseria lactamica*). In some embodiments, the Cas3 protein comprises a sequence of SEQ ID NO: 5, or variant thereof. For example, in certain embodiments, the Cas3 protein comprises an amino acid sequence having at least 70% similarity to that of SEQ ID NO: 5.

[0113] In some embodiments, the system may be derived from CRISPR-Cas elements (e.g., Cascade-Cas3 proteins or variants thereof) from a *Bacillus* species (e.g., *Bacillus halodurans* (Bha)). In some embodiments, the Cas3 protein comprises a sequence of SEQ ID NO: 10, or variant thereof. For example, in certain embodiments, the Cas3 protein comprises an amino acid sequence having at least 70% similarity to that of SEQ ID NO: 10.

[0114] In some embodiments, the system may be derived from CRISPR-Cas elements (e.g., Cascade-Cas3 proteins or variants thereof) from a *Desulfovibrio* species (e.g., *Desulfovibrio vulgaris* (Dvu)). In some embodiments, the Cas3 protein comprises a sequence of SEQ ID NO: 15, or variant thereof. For example, in certain embodiments, the Cas3 protein comprises an amino acid sequence having at least 70% similarity to that of SEQ ID NO: 15.

[0115] In some embodiments, the system comprises Cas3, and one or more of: Cas5, Cas7, Cas8, and Cas11, wherein at least one of the Cas5, Cas7, Cas8, and Cas11 is an engineered protein or fusion protein as disclosed herein. In some embodiments, the system comprises one of Cas5, Cas7, Cas8, and Cas11, which is an engineered protein or fusion protein as disclosed herein. In some embodiments, the system comprises two, three or all of Cas5, Cas7, Cas8, and Cas11, only one of which is an engineered protein or fusion protein as disclosed herein and the other Cas proteins are wild-type proteins or other variants thereof, for example those disclosed in International Application No. PCT/US2022/031091, incorporated herein by reference. In some embodiments, the system comprises two, three or all of Cas5, Cas7, Cas8, and Cas11, all of which are an engineered protein or fusion protein as disclosed herein

[0116] In some embodiments, the system comprises a Cas8 protein having an amino acid sequence with one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1. In some embodiments, the system comprises a Cas5 protein having an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2. In some embodiments, the system comprises a Cas7 protein having an amino acid sequence comprises an E90R substitution relative to SEQ ID NO: 3.

[0117] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having an E235R substitution relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having an D387R

substitution relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having an E391R substitution relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having E235R and D387R substitutions relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having E235R and E391R substitutions relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having D387R and E391R substitutions relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having E235R, D387R, and E391R substitutions relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0118] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3. In certain embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having E235R, D387R, and E391R substitutions relative to SEQ ID NO: 1 and a Cas7 protein having an E90R substitution relative to SEQ ID NO: 3.

[0119] In select embodiments, the system comprises a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0120] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2

[0121] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having an E235R substitution relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid

sequence having an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0122] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having a D387R substitution relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0123] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having an E391R substitution relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0124] In select embodiments, the system comprises a Cas8 protein comprising an amino acid sequence having E235R, D387R, and E391R substitutions relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2.

[0125] The one or more nucleic acids encoding the engineered CRISPR-Cas system may be any nucleic acid including DNA, RNA, or combinations thereof. In some embodiments, the one or more nucleic acids comprise one or more messenger RNAs, one or more vectors, or any combination thereof.

[0126] In some embodiments, the Cas3 and the other Cas proteins, e.g., the engineered Cas protein(s) or fusion protein(s) are encoded by a single nucleic acid (e.g., a single vector). In some embodiments, the Cas3 and the other Cas proteins, e.g., the engineered Cas protein(s) or fusion protein(s) are encoded by different nucleic acids (e.g., multiple mRNAs or two or more vectors).

[0127] In certain embodiments, engineering the system for use in eukaryotic cells may involve codon-optimization or other modification (e.g., to include an appropriate nuclear localization signal (NLS) or purification tag). It will be appreciated that changing native codons to those most frequently used in mammals allows for maximum expression of the system proteins in mammalian cells (e.g., human cells). Such modified nucleic acid sequences are commonly described in the art as "codon-optimized," or as utilizing "mammalian-preferred" or "human-preferred" codons. In some embodiments, the nucleic acid sequence is considered codon-optimized if at least about 60% (e.g., 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 98%) of the codons encoded therein are mammalian preferred codons. Furthermore, in some embodiments, engineering the system involves incorporating elements of a native CRISPR array into the disclosed system.

[0128] The system and the nucleic acid disclosed herein may comprise at least one guide RNA (gRNA), wherein each gRNA is configured to hybridize to a target nucleic acid sequence. The gRNA may be a crRNA or a crRNA/tracrRNA (e.g., single guide RNA, sgRNA) fusion. The terms "gRNA" and "guide RNA" refer to any nucleic acid comprising a sequence that determines the binding specificity of the CRISPR-Cas complex. In instances in which the system comprises two or more guide RNAs, each guide RNA may hybridize to a different target nucleic acid sequence.

[0129] The at least one gRNA may be encoded on the same or different nucleic acid as any of the Cas3 and the other Cas proteins, e.g., the engineered Cas protein(s) or fusion protein(s). For example, a single vector may encode any or all of the at least one gRNA, the Cas3, and the other Cas proteins, e.g., the engineered Cas protein(s) or fusion protein(s).

[0130] The terms "target DNA sequence," "target nucleic acid," "target sequence," and "target site" are used interchangeably herein to refer to a polynucleotide (nucleic acid, gene, chromosome, genome, etc.) to which a guide sequence (e.g., a guide RNA) is designed to have complementarity, wherein hybridization between the target sequence and a guide sequence promotes the formation of a CRISPR/Cas complex, provided sufficient conditions for binding exist. The target sequence and guide sequence need not exhibit complete complementarity, provided that there is sufficient complementarity to cause hybridization and promote formation of a CRISPR complex. In some embodiments the system further comprises at least one target nucleic acid.

[0131] A target sequence may comprise any polynucleotide, such as DNA or RNA. Suitable DNA/RNA binding conditions include physiological conditions normally present in a cell. Other suitable DNA/RNA binding conditions (e.g., conditions in a cell-free system) are known in the art; see, e.g., Sambrook, referenced herein and incorporated by reference. The strand of the target DNA that is complementary to and hybridizes with the DNA-targeting RNA is referred to as the "complementary strand" and the strand of the target DNA that is complementary to the "complementary strand" (and is therefore not complementary to the DNA-targeting RNA) is referred to as the "noncomplementary strand" or "non-complementary strand."

[0132] The target nucleic acid sequence may include a protospacer adjacent motif (PAM). A PAM can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more nucleotides in length. In certain embodiments, a PAM is between 2-6 nucleotides in length. In some embodiments, the PAM is 3 nucleotides in length. The PAM may be "adjacent to" the target nucleic acid sequence in that it typically immediately precedes the target sequence. In some embodiments, the PAM is 5' of the target site.

[0133] PAM sequences are often specific to the particular Cas endonuclease being used in the CRISPR/Cas complex and the species from which it was derived. For example, Type I-C CRISPR-Cas3 elements typically are active in a host cell genome which comprises a protospacer adjacent motif (PAM) comprising the nucleic acid sequence 5'-TTC-3' or 5'-TTT-3' located adjacent to the target genomic DNA sequence. PAM sequences and methods of determining PAM sequences for specific Cas proteins are known in the art. The gRNA or portion thereof that hybridizes to a target nucleic acid sequence (e.g., the guide sequence) may be between any length.

[0134] The guide sequence of the gRNA does not need to be completely complementary to the target site. In some embodiments, the guide sequence of the gRNA is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or at least 100% complementary to the target site. In some embodiments, the gRNA sequence is at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, or at least 100% complementary to the 3' end of the target site (e.g., the last 5, 6, 7, 8, 9, or 10 nucleotides of the 3' end of the target site). "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule, which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence.

[0135] To facilitate gRNA design, many computational tools have been developed (See Prykhozhij et al. (PLoS ONE, 10(3): (2015)); Zhu et al. (PLoS ONE, 9(9) (2014)); Xiao et al. (Bioinformatics. Jan 21 (2014)); Heigwer et al. (Nat Methods, 11(2): 122–123 (2014)). Methods and tools for guide RNA design are discussed by Zhu (Frontiers in Biology, 10 (4) pp 289-296 (2015)), which is incorporated by reference herein. Additionally, there are many publicly available software tools that can be used to facilitate the design of sgRNA(s); including but not limited to, Genscript Interactive CRISPR gRNA Design Tool, WU-CRISPR, and Broad Institute GPP sgRNA Designer.

[0136] In addition to the guide sequence, in some embodiments, a gRNA may also comprise a scaffold sequence (e.g., tracrRNA). Exemplary scaffold sequences will be evident to one of skill in the art and can be found, for example, in Jinek, et al. Science (2012) 337(6096):816-821, and Ran, et al. Nature Protocols (2013) 8:2281-2308, incorporated herein by reference in their entireties.

[0137] In some embodiments, at least one gRNA is within a crRNA array. A crRNA array comprises multiple guide RNAs (sgRNA) derived from the fusion of CRISPR RNA (crRNA) and *trans*-activating crRNA (tracrRNA) expressed a single transcript, which after processing by a nuclease are cleaved into

separate gRNAs. The crRNA array may contain multiple repeats separated by unique spacers. For example, an engineered crRNA array may comprise two repeats and one spacer, or three repeats and two identical spacers.

[0138] One or all of the at least one gRNAs may be a non-naturally occurring gRNA.

[0139] In some embodiments, the system is a cell-free system.

[0140] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids encoding components of the present system into cells, tissues, or a subject. Such methods can be used to administer nucleic acids encoding components of the present system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, cosmids, RNA (e.g., a transcript of a vector described herein), and a nucleic acid complexed with a delivery vehicle.

[0141] Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. A variety of viral constructs may be used to deliver the present system and/or components to the cells, tissues, and/or a subject. Viral vectors include, for example, retroviral, lentiviral, adenoviral, adeno-associated and herpes simplex viral vectors. Nonlimiting examples of such recombinant viruses include recombinant adeno-associated virus (AAV), recombinant adenoviruses, recombinant lentiviruses, recombinant retroviruses, recombinant herpes simplex viruses, recombinant poxviruses, phages, etc. The present disclosure provides vectors capable of integration in the host genome, such as retrovirus or lentivirus. See, e.g., Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989; Kay, M. A., et al., 2001 Nat. Medic. 7(1):33-40; and Walther W. and Stein U., 2000 Drugs, 60(2): 249-71.

[0142] Drug selection strategies may be adopted for positively selecting for cells comprising the nucleic acid sequences encoding the present system or components thereof.

[0143] The present disclosure also provides for DNA segments encoding the proteins and nucleic acids disclosed herein, vectors containing these segments and cells containing the vectors. The vectors may be used to propagate the segment in an appropriate cell and/or to allow expression from the segment (e.g., an expression vector). The person of ordinary skill in the art would be aware of the various vectors available for propagation and expression of a nucleic acid sequence.

[0144] To construct cells that express the present system, expression vectors for stable or transient expression of the present system may be constructed via conventional methods and introduced into cells. For example, nucleic acids encoding the components of the present system may be cloned into a suitable expression vector, such as a plasmid or a viral vector in operable linkage to a suitable promoter. The

selection of expression vectors/plasmids/viral vectors should be suitable for integration and replication in eukaryotic cells.

[0145] In certain embodiments, vectors of the present disclosure can drive the expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, Nature (1987) 329:840, incorporated herein by reference) and pMT2PC (Kaufman, et al., EMBO J. (1987) 6:187, incorporated herein by reference). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., incorporated herein by reference.

[0146] Vectors of the present disclosure can comprise any of a number of promoters known to the art, wherein the promoter is constitutive, regulatable or inducible, cell type specific, tissue-specific, or species specific. In addition to the sequence sufficient to direct transcription, a promoter sequence of the invention can also include sequences of other regulatory elements that are involved in modulating transcription (e.g., enhancers, Kozak sequences and introns). Many promoter/regulatory sequences useful for driving constitutive expression of a gene are available in the art and include, but are not limited to, for example, CMV (cytomegalovirus promoter), EF1a (human elongation factor 1 alpha promoter), SV40 (simian vacuolating virus 40 promoter), PGK (mammalian phosphoglycerate kinase promoter), Ubc (human ubiquitin C promoter), human beta-actin promoter, rodent beta-actin promoter, CBh (chicken beta-actin promoter), CAG (hybrid promoter contains CMV enhancer, chicken beta actin promoter, and rabbit beta-globin splice acceptor), TRE (Tetracycline response element promoter), H1 (human polymerase III RNA promoter), U6 (human U6 small nuclear promoter), and the like. Additional promoters that can be used for expression of the components of the present system, include, without limitation, cytomegalovirus (CMV) intermediate early promoter, a viral LTR such as the Rous sarcoma virus LTR, HIV-LTR, HTLV-1 LTR, Maloney murine leukemia virus (MMLV) LTR, myeloproliferative sarcoma virus (MPSV) LTR, spleen focus-forming virus (SFFV) LTR, the simian virus 40 (SV40) early promoter, herpes simplex tk virus promoter, elongation factor 1-alpha (EF1-α) promoter with or without the EF1-α intron. Additional promoters include any constitutively active promoter. Alternatively, any regulatable promoter may be used, such that its expression can be modulated within a cell.

[0147] Moreover, inducible expression can be accomplished by placing the nucleic acid encoding such a molecule under the control of an inducible promoter/regulatory sequence. Promoters well known in the art can be induced in response to inducing agents such as metals, glucocorticoids, tetracycline, hormones, and the like, are also contemplated for use with the invention. Thus, it will be appreciated that the present disclosure includes the use of any promoter/regulatory sequence known in the art that is capable of driving expression of the desired protein operably linked thereto.

[0148] The vectors of the present disclosure may direct the expression of the nucleic acid in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Such regulatory elements include promoters that may be tissue specific or cell specific. The term "tissue specific" as it applies to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest to a specific type of tissue (e.g., seeds) in the relative absence of expression of the same nucleotide sequence of interest in a different type of tissue. The term "cell type specific" as applied to a promoter refers to a promoter that is capable of directing selective expression of a nucleotide sequence of interest in a specific type of cell in the relative absence of expression of the same nucleotide sequence of interest in a different type of cell within the same tissue. The term "cell type specific" when applied to a promoter also means a promoter capable of promoting selective expression of a nucleotide sequence of interest in a region within a single tissue. Cell type specificity of a promoter may be assessed using methods well known in the art, e.g., immunohistochemical staining.

[0149] Additionally, the vector may contain, for example, some or all of the following: a selectable marker gene, such as the neomycin gene for selection of stable or transient transfectants in host cells; enhancer/promoter sequences from the immediate early gene of human CMV for high levels of transcription; transcription termination and RNA processing signals from SV40 for mRNA stability; 5'- and 3'-untranslated regions for mRNA stability and translation efficiency from highly-expressed genes like α-globin or β-globin; SV40 polyoma origins of replication and ColE1 for proper episomal replication; internal ribosome binding sites (IRESes), versatile multiple cloning sites; T7 and SP6 RNA promoters for in vitro transcription of sense and antisense RNA; a "suicide switch" or "suicide gene" which when triggered causes cells carrying the vector to die (e.g., HSV thymidine kinase, an inducible caspase such as iCasp9), and reporter gene for assessing expression.

[0150] When introduced into a cell, the vectors may be maintained as an autonomously replicating sequence or extrachromosomal element or may be integrated into host DNA.

[0151] The present system or components thereof (e.g., Cas proteins or fusion proteins as described herein) may be delivered to a cell by any suitable means. In certain embodiments, the system is delivered in vivo. In other embodiments, the system is delivered to isolated/cultured cells in vitro or ex vivo to provide modified cells useful for in vivo delivery to patients afflicted with a disease or condition.

[0152] Vectors according to the present disclosure can be transformed, transfected, or otherwise introduced into a wide variety of host cells. Transfection refers to the taking up of a vector by a cell whether or not any coding sequences are in fact expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, lipofectamine, calcium phosphate co-precipitation, electroporation, DEAE-dextran treatment, microinjection, viral infection, and other methods known in the art. Transduction refers to entry of a virus into the cell and expression (e.g., transcription and/or translation) of sequences delivered by the viral vector genome. In the case of a recombinant vector, "transduction" generally refers to entry of the recombinant viral vector into the cell and expression of a nucleic acid of interest delivered by the vector genome.

[0153] Any of the vectors comprising a nucleic acid sequence that encodes the components of the present system is also within the scope of the present disclosure. Such a vector may be delivered into cells by a suitable method. Methods of delivering vectors to cells are well known in the art and may include DNA or RNA electroporation, transfection reagents such as liposomes or nanoparticles to delivery DNA or RNA; delivery of DNA, RNA, or protein by mechanical deformation (see, e.g., Sharei et al. Proc. Natl. Acad. Sci. USA (2013) 110(6): 2082-2087, incorporated herein by reference); or viral transduction. In some embodiments, the vectors are delivered to host cells by viral transduction. Nucleic acids can be delivered as part of a larger construct, such as a plasmid or viral vector, or directly, e.g., by electroporation, lipid vesicles, viral transporters, microinjection, and biolistics (high-speed particle bombardment). In some embodiments, the construct or the nucleic acid encoding the components of the present system is a DNA molecule. In some embodiments, the nucleic acid encoding the components of the present system is a DNA vector and may be electroporated to cells. In some embodiments, the nucleic acid encoding the components of the present system is an RNA molecule, which may be electroporated to cells.

[0154] Additionally, delivery vehicles such as nanoparticle- and lipid-based mRNA or protein delivery systems can be used. Further examples of delivery vehicles include lentiviral vectors, ribonucleoprotein (RNP) complexes, lipid-based delivery system, gene gun, hydrodynamic, electroporation or nucleofection microinjection, and biolistics. Various gene delivery methods are discussed in detail by Nayerossadat et

al. (Adv Biomed Res. 2012; 1: 27) and Ibraheem et al. (Int J Pharm. 2014 Jan 1;459(1-2):70-83), incorporated herein by reference.

[0155] In other embodiments, various components of the system may be introduced into a host cell as a ribonucleoprotein (RNP) complex. The term "ribonucleoprotein complex," as used herein, refers to a complex of ribonucleic acid and RNA-binding protein(s). In the context of CRISPR-Cas systems, an RNP complex typically comprises Cas protein(s) (e.g., Cas5, Cas7, and Cas8) in complex with a gRNA. RNPs may be assembled *in vitro* and can be delivered directly to cells using standard electroporation, cationic lipids, gold nanoparticles, or other transfection techniques (see, e.g., Kim et al., *Genome Res.*, *24*: 1012-1019 (2014); Zuris et al., *Nat. Biotechnol.*, *33*: 73-80 (2015); and Mout et al., *ACS Nano.*, *11*: 2452-2458 (2017)).

[0156] As such, the disclosure provides an isolated cell comprising the systems, the vector(s), nucleic acid(s), or proteins disclosed herein. The disclosure also provides populations of cells comprising the systems, the vector(s), nucleic acid(s), or proteins disclosed herein.

[0157] Preferred cells are those that can be easily and reliably grown, have reasonably fast growth rates, have well characterized expression systems, and can be transformed or transfected easily and efficiently, including both eukaryotic and prokaryotic cells. Examples of suitable prokaryotic cells include, but are not limited to, cells from the genera *Bacillus* (such as *Bacillus subtilis* and *Bacillus brevis*), *Escherichia* (such as *E. coli*), *Pseudomonas*, *Streptomyces*, *Salmonella*, and *Erwinia*. Suitable eukaryotic cells are known in the art and include, for example, yeast cells, insect cells, and mammalian cells. Examples of suitable yeast cells include those from the genera *Kluyveromyces*, *Pichia*, *Rhino-sporidium*, *Saccharomyces*, and *Schizosaccharomyces*. Exemplary insect cells include Sf-9 and HIS (Invitrogen, Carlsbad, Calif.) and are described in, for example, Kitts et al., *Biotechniques*, *14*: 810-817 (1993); Lucklow, *Curr. Opin. Biotechnol.*, *4*: 564-572 (1993); and Lucklow et al., *J. Virol.*, *67*: 4566-4579 (1993), incorporated herein by reference.

[0158] Desirably, the cell is a mammalian cell, and in some embodiments, the cell is a human cell. A number of suitable mammalian and human host cells are known in the art, and many are available from the American Type Culture Collection (ATCC, Manassas, Va.). Examples of suitable mammalian cells include, but are not limited to, Chinese hamster ovary cells (CHO) (ATCC No. CCL61), CHO DHFR- cells (Urlaub et al., Proc. Natl. Acad. Sci. USA, 97: 4216-4220 (1980)), human embryonic kidney (HEK) 293 or 293T cells (ATCC No. CRL1573), and 3T3 cells (ATCC No. CCL92). Other suitable mammalian cell lines are the monkey COS-1 (ATCC No. CRL1650) and COS-7 cell lines (ATCC No. CRL1651), as

well as the CV-1 cell line (ATCC No. CCL70). Further exemplary mammalian host cells include primate, rodent, and human cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as well as primary explants, are also suitable. Other suitable mammalian cell lines include, but are not limited to, mouse neuroblastoma N2A cells, HeLa, HEK, A549, HepG2, mouse L-929 cells, and BHK or HaK hamster cell lines.

[0159] Methods for selecting suitable mammalian cells and methods for transformation, culture, amplification, screening, and purification of cells are known in the art.

[0160] The systems may further comprise components in addition to those listed, including, but not limited to: sequence tags, protein markers or marker proteins, spacers, capture sequences, and the like.

3. Methods

[0161] Disclosed herein are methods for utilizing the disclosed proteins and systems. The descriptions and embodiments provided above for the proteins and systems are applicable to the methods described herein.

a) Recruiting Effectors and Modulating Expression

[0162] The disclosure provides methods for recruiting one or more effector domains to a target nucleic acid and for modulating expression of a target gene. The methods comprise contacting a target nucleic acid sequence or target gene with a fusion protein or a system as disclosed herein, or a composition comprising the fusion protein or system. In some embodiments, the target nucleic acid or target gene are in a cell and the methods comprise introducing the fusion protein, the system, or the composition comprising the fusion protein or system into the cell.

[0163] When utilizing a system as disclosed herein in the methods for recruiting one or more effector domains to a target nucleic acid and for modulating expression of a target gene, the system may be engineered such that Cas3 function is fully or partially inactivated, as described above. Thus, in some embodiments, the methods utilize a system having a fully or partially inactivated Cas3.

[0164] Polynucleotides containing the target nucleic acid sequence may include, but is not limited to, purified chromosomal DNA, total cDNA, cDNA fractionated according to tissue or expression state (e.g., after heat shock or after cytokine treatment other treatment) or expression time (after any such treatment) or developmental stage, plasmid, cosmid, BAC, YAC, phage library, etc.

[0165] In some embodiments, the target nucleic acid is a nucleic acid endogenous to a target cell. In some embodiments, the target nucleic acid encodes a gene or gene product. The term "gene product," as used herein, refers to any biochemical product resulting from expression of a gene. Gene products may be

RNA or protein. RNA gene products include non-coding RNA, such as tRNA, rRNA, micro RNA (miRNA), and small interfering RNA (siRNA), and coding RNA, such as messenger RNA (mRNA). In some embodiments, the target nucleic acid sequence encodes a protein or polypeptide.

[0166] In some embodiments, the target nucleic acid comprises a promoter region of a gene of interest. In some embodiments, the target nucleic acid comprises an upstream activator sequence. In some embodiments, the gene of interest is located on a chromosome in a cell.

[0167] As described above the fusion protein or system may be introduced into eukaryotic or prokaryotic cells by a variety of methods. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell.

[0168] In some embodiments, introducing the fusion protein or a system into a cell comprises administering the system to a subject. In some embodiments, the subject is human. The administering may comprise in vivo administration. In alternative embodiments, a vector is contacted with a cell in vitro or ex vivo and the treated cell, containing the system, is transplanted into a subject.

b) Altering a Target Sequence

[0169] The disclosure also provides a method of altering a target nucleic acid sequence. The phrase "altering a DNA sequence," as used herein, refers to modifying at least one physical feature of a DNA sequence of interest. DNA alterations include, for example, single or double strand DNA breaks, deletion, or insertion of one or more nucleotides, and other modifications that affect the structural integrity or nucleotide sequence of the DNA sequence.

[0170] The methods comprise contacting a target nucleic acid sequence with a system disclosed herein or a composition comprising the system.

[0171] In one embodiment, the method introduces a single strand or double strand break in the target DNA sequence. In this respect, the disclosed systems may direct cleavage of one or both strands of a target DNA sequence, such as within the target genomic DNA sequence and/or within the complement of the target sequence.

[0172] In some embodiments, altering a DNA sequence comprises a deletion. The deletion may be upstream or downstream of the PAM binding side, so called unidirectional deletions. The deletion may encompass sequences on either side of the PAM binding site, a bidirectional deletion. In some embodiments, the system introduces unidirectional DNA deletions. In some embodiments, the system introduces bidirectional DNA deletions. In some embodiments, the system introduces a deletion without prominent off-target activity.

[0173] The deletion of the DNA sequence may be of any size. For example, in some embodiments the deletion of the DNA sequence comprises from about 500 nucleotides to about 100,000 nucleotides (e.g., about 1,000, 5,000, 10,000, or 50,000 nucleotides, or a range defined by any two of the foregoing values). In other embodiments, the deletion of the DNA sequence comprises from about 5,000 nucleotides to about 20,000 nucleotides (e.g., about 6,000, 6,500, 7,000, 7,500, 8,000, 8,500, 9,000, 9,500, 10,000, 10,500, 11,000, 11,500, 12,000, 12,500, 13,000, 13,500, 14,000, 14,500, 15,000, 15,500, 16,000, 16,500, 17,000, 17,500, 18,000, 18,500, 19,000, or 19,500 nucleotides, or a range defined by any two of the foregoing values).

[0174] In some embodiments, the contacting comprises introducing the system into the cell. As described above the system may be introduced into eukaryotic or prokaryotic cells by methods known in the art. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a human cell.

[0175] In some embodiments, introducing the system into a cell comprises administering the system to a subject. In some embodiments, the subject is human. The administering may comprise in vivo administration. In alternative embodiments, a vector is contacted with a cell in vitro or ex vivo and the treated cell, containing the system, is transplanted into a subject.

[0176] In some embodiments, the target nucleic acid is a nucleic acid endogenous to a target cell. In some embodiments, the target nucleic acid is a genomic DNA sequence. The term "genomic," as used herein, refers to a nucleic acid sequence (e.g., a gene or locus) that is located on a chromosome in a cell.

[0177] In some embodiments, the target nucleic acid encodes a gene or gene product. The term "gene product," as used herein, refers to any biochemical product resulting from expression of a gene. Gene products may be RNA or protein. RNA gene products include non-coding RNA, such as tRNA, rRNA, micro RNA (miRNA), and small interfering RNA (siRNA), and coding RNA, such as messenger RNA (mRNA). In some embodiments, the target nucleic acid sequence encodes a protein or polypeptide.

[0178] The disclosed method may alter a target DNA sequence in a host cell so as to modulate expression of the target DNA sequence, e.g., expression of the target DNA sequence is increased, decreased, or completely eliminated (e.g., via deletion of a gene). In one embodiment, the disclosed system cleaves a target DNA sequence of the host cell to produce double strand DNA breaks. The double strand breaks can be repaired by the host cell by either non-homologous end joining (NHEJ) or homologous recombination. In NHEJ, the double-strand breaks are repaired by direct ligation of the break ends to one another. In homologous recombination repair, a donor nucleic acid molecule comprising a second DNA sequence with homology to the cleaved target DNA sequence is used as a template for repair of the cleaved target

DNA sequence, resulting in the transfer of genetic information from the donor nucleic acid molecule to the target DNA. As a result, new nucleic acid material is inserted/copied into the DNA break site. The modifications of the target sequence due to NHEJ and/or homologous recombination repair may lead to, for example, gene correction, gene replacement, gene tagging, transgene insertion, nucleotide deletion, gene disruption, gene mutation, gene knock-down, etc.

[0179] In some embodiments, the systems and methods described herein may be used to correct one or more defects or mutations in a gene (referred to as "gene correction"). In such cases, the target sequence encodes a defective version of a gene, and the disclosed system further comprises a donor nucleic acid molecule which encodes a wild-type or corrected version of the gene. Thus, in other words, the target sequence is a "disease-associated" gene. The term "disease-associated gene," refers to any gene or polynucleotide whose gene products are expressed at an abnormal level or in an abnormal form in cells obtained from a disease-affected individual as compared with tissues or cells obtained from an individual not affected by the disease. A disease-associated gene may be expressed at an abnormally high level or at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene, the mutation or genetic variation of which is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. Examples of genes responsible for such "single gene" or "monogenic" diseases include, but are not limited to, adenosine deaminase, α-1 antitrypsin, cystic fibrosis transmembrane conductance regulator (CFTR), β-hemoglobin (HBB), oculocutaneous albinism II (OCA2), Huntingtin (HTT), dystrophia myotonica-protein kinase (DMPK), low-density lipoprotein receptor (LDLR), apolipoprotein B (APOB), neurofibromin 1 (NF1), polycystic kidney disease 1 (PKD1), polycystic kidney disease 2 (PKD2), coagulation factor VIII (F8), dystrophin (DMD), phosphate-regulating endopeptidase homologue, X-linked (PHEX), methyl-CpG-binding protein 2 (MECP2), and ubiquitin-specific peptidase 9Y, Y-linked (USP9Y). Other single gene or monogenic diseases are known in the art and described in, e.g., Chial, H. *Rare Genetic Disorders: Learning About Genetic Disease Through Gene Mapping, SNPs, and Microarray Data*, *Nature Education* 1(1):192 (2008); Online Mendelian Inheritance in Man (OMIM); and the Human Gene Mutation Database (HGMD). In another embodiment, the target genomic DNA sequence can comprise a gene, the mutation of which contributes to a particular disease in combination with mutations in other genes. Diseases caused by the contribution of multiple genes which lack simple (i.e., Mendelian) inheritance patterns are referred to in the art as a "multifactorial" or "polygenic" disease. Examples of multifactorial or polygenic diseases include, but are not limited to, asthma, diabetes,

epilepsy, hypertension, bipolar disorder, and schizophrenia. Certain developmental abnormalities also can be inherited in a multifactorial or polygenic pattern and include, for example, cleft lip/palate, congenital heart defects, and neural tube defects.

[0180] In another embodiment, the method of altering a target sequence can be used to delete nucleic acids from a target sequence in a host cell by cleaving the target sequence and allowing the host cell to repair the cleaved sequence in the absence of an exogenously provided donor nucleic acid molecule. Deletion of a nucleic acid sequence in this manner can be used in a variety of applications, such as, for example, to remove disease-causing trinucleotide repeat sequences in neurons, to create gene knock-outs or knock-downs, and to generate mutations for disease models in research.

[0181] The components of the present systems or fusion proteins cells may be administered with a pharmaceutically acceptable carrier or excipient as a pharmaceutical composition. In some embodiments, the components of the present system may be mixed, individually or in any combination, with a pharmaceutically acceptable carrier to form pharmaceutical compositions, which are also within the scope of the present disclosure.

[0182] In some embodiments, an effective amount of the components of the present system or compositions as described herein can be administered. Within the context of the present disclosure, the term "effective amount" refers to that quantity of the components of the system such that recruitment of one or more effector domains and, if desired, modulation of expression of a target gene is achieve.

[0183] When utilized as a method of treatment, the effective amount may depend on the particular condition being treated, the severity of the condition, the individual patient parameters including age, physical condition, size, gender and weight, the duration of the treatment, the nature of concurrent therapy (if any), the specific route of administration and like factors within the knowledge and expertise of the health practitioner. In some embodiments, the effective amount alleviates, relieves, ameliorates, improves, reduces the symptoms, or delays the progression of any disease or disorder in the subject. In some embodiments, the subject is a human.

[0184] In the context of the present disclosure insofar as it relates to any of the disease conditions recited herein, the terms "treat," "treatment," and the like mean to relieve or alleviate at least one symptom associated with such condition, or to slow or reverse the progression of such condition. Within the meaning of the present disclosure, the term "treat" also denotes to arrest, delay the onset (e.g., the period prior to clinical manifestation of a disease) and/or reduce the risk of developing or worsening a disease.

For example, in connection with cancer the term "treat" may mean eliminate or reduce a patient's tumor burden, or prevent, delay, or inhibit metastasis, etc.

[0185] The phrase "pharmaceutically acceptable," as used in connection with compositions and/or cells of the present disclosure, refers to molecular entities and other ingredients of such compositions that are physiologically tolerable and do not typically produce untoward reactions when administered to a subject (e.g., a mammal, a human). Preferably, as used herein, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, and more particularly in humans. "Acceptable" means that the carrier is compatible with the active ingredient of the composition (e.g., the nucleic acids, vectors, cells, or therapeutic antibodies) and does not negatively affect the subject to which the composition(s) are administered. Any of the pharmaceutical compositions and/or cells to be used in the present methods can comprise pharmaceutically acceptable carriers, excipients, or stabilizers in the form of lyophilized formations or aqueous solutions.

[0186] Pharmaceutically acceptable carriers, including buffers, are well known in the art, and may comprise phosphate, citrate, and other organic acids; antioxidants including ascorbic acid and methionine; preservatives; low molecular weight polypeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; amino acids; hydrophobic polymers; monosaccharides; disaccharides; and other carbohydrates; metal complexes; and/or non-ionic surfactants. See, e.g., Remington: The Science and Practice of Pharmacy 20th Ed. (2000) Lippincott Williams and Wilkins, Ed. K. E. Hoover.

4. Kits

[0187] The disclosure further provides kits containing one or more reagents or other components useful, necessary, or sufficient for practicing any of the methods described herein. For example, kits may include the disclosed engineered Cas proteins, fusion proteins, other CRISPR/Cas components (e.g., Cas proteins, guide RNAs) vectors, compositions, transfection or administration reagents, negative and positive control samples (e.g., cells, template DNA), cells, containers housing one or more components (e.g., microcentrifuge tubes, boxes), detectable labels, detection and analysis instruments, software, instructions, and the like.

[0188] Any element of any suitable CRISPR/Cas gene editing system known in the art can be employed in the systems and methods described herein, as appropriate. CRISPR/Cas gene editing technology is described in detail in, for example, U.S. Patent Nos. 8,546,553, 8,697,359; 8,771,945; 8,795,965; 8,865,406; 8,871,445; 8,889,356; 8,889,418; 8,895,308; 8,9066,616; 8,932,814; 8,945,839; 8,993,233;

8,999,641; 9,115,348; 9,149,049; 9,493,844; 9,567,603; 9,637,739; 9,663,782; 9,404,098; 9,885,026; 9,951,342; 10,087,431; 10,227,610; 10,266,850; 10,601,748; 10,604,771; and 10,760,064; and U.S. Patent Application Publication Nos. US2010/0076057; US2014/0113376; US2015/0050699; US2015/0031134; US2014/0357530; US2014/0349400; US2014/0315985; US2014/0310830; US2014/0310828; US2014/0309487; US2014/0294773; US2014/0287938; US2014/0273230; US2014/0242699; US2014/0242664; US2014/0212869; US2014/0201857; US2014/0199767; US2014/0189896; US2014/0186919; US2014/0186843; and US2014/0179770, each incorporated herein by reference.

[0189] The following examples further illustrate the invention but should not be construed as in any way limiting its scope.

EXAMPLES

Materials and Methods

[0190] *Neisseria lactamica type I-C CRISPR-Cas variants creation* All glutamate (Glu) and aspartate (Asp) residues that are within 10Å of the target strand DNA were selected based on Neisseria lactamica type I-C CRISPR-Cas cryo-EM structure by using the PyMOL selection algebra command line. Plasmids encoding cascade subunits with arginine (Arg) substitutions of the selected Glu/Asp residues were generated by KLD site-directed mutagenesis. All clones were sequence confirmed by Sanger sequencing.

[0191] *Cell culture* HAP1-EGFP reporter cells were cultured in IMDM (Gibco) supplemented with 10% FBS. HEK293T-EGFP reporter cells and HEK293T-HBB sickle cells were cultured in 10% FBS supplemented DMEM/F12 (Gibco) supplemented with 10% FBS. All cells were cultured in a tissue culture incubator at 37°C with 5% CO_2.

[0192] *Plasmid transfection* CRISPR-Cas3 plasmid transfection was conducted using Lipofectamine 3000 Transfection Reagent (ThermoFisher) per manufacturer's instructions. HAP1-EGFP reporter cells or 293T cells were seeded one day before transfection at 0.65x105 or 1.5x105 cells per well of a 24-well plate, respectively. For each transfection, 1 μL P3000 Enhancer Reagent, 1.5 μL Lipofectamine 3000 reagent, and a total of 500 ng crispr-cas plasmids was used. To monitor genome targeting efficiency, cells were analyzed by flow cytometry 4-5 days post transfection. For GFP disruption experiments, 45, 22.5, 67.5, 270, 45 and 50 ng of Cas3, Cas5, Cas7, Cas8, Cas11 and CRISPR plasmids were used, respectively. For base editing experiments, WT Cas11 subunit plasmid was substituted with the same amount of Cas11-TadA* fusion derivative plasmid and WT Cas3 is substituted with Cas3 nickase constructs.

[0193] *In vitro transcription* 5' capped and 3' polyadenylated cas mRNAs were synthesized by in vitro transcription using AmpliScribe T7-Flash Transcription Kit (Lucigen) per manufacturer's instructions with the following modifications. UTP is substituted with N1-Methylpseudouridine-5'-Triphosphate (TriLink). CleanCap Reagent AG (TriLink) (7.2mM final concentration) was added to the reaction mix for co-transcriptional capping. In vitro transcribed RNAs were purified with LiCl precipitation and resuspended in nuclease free water. DNA templates used were purified PCR amplifications with a modified T7 promoter for the co-transcriptional capping, optimized UTR sequences and polyA sequences incorporated.

[0194] *mRNA electroporation into iPS cells* CFTR G542X iPS cells were washed once with 1xPBS on plate and then individualized with Accutase. Individualized cells were pelleted and then washed once with 1x PBS and resuspended in Neon buffer R to a concentration of $4x10^7$ cells/mL. Approximately $2x10^5$ cells were mixed with 40, 165, 220, 70, 55 ng of cas3, cas5, cas7, cas8, cas11 mRNAs, along with 100 pmol synthetic CRISPR RNA, in buffer R in a total volume of 11 µL. Each mixture was electroporated with a 10 µL Neon tip (1100 V 20 ms 2 pulses for iPS cells,) and plated in 24-well Matrigel coated tissue culture plates containing 500 µL E8 medium (iPS cells).

[0195] *High-throughput sequencing of genomic DNA samples* For base editing experiments, cells were lysed 4 days after plasmid transfection by adding 250 µL of lysis buffer per well (10mM Tris-HCl pH 8.0, 0.05% SDS, 0.04mg/mL proteinase K). The lysates were incubated at 37°C for 60 min, 80°C for 30 min and then used directly for NGS library construction. A 200-300 bp region surrounding the target genomic site was PCR amplified. Following Illumina barcoding, PCR amplicons were pooled and purified through 2% agarose gel electrophoresis and gel extraction using a Zymo DNA Gel Extraction Kit. Final elution is done with 20 µL 10mM Tris pH8.0. DNA concentration was quantified with a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) and sequenced on an Illumina MiSeq instrument (300 bp single end read) according to the manufacturer's protocols.

[0196] *Bioinformatic analysis of NGS sequencing datasets* Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina), and the FASTQ files were analyzed using CRISPResso2. Base editing efficiency values were reported as the percentage of reads with A•T to G•C conversion at a specific adenine location in the total aligned reads. The indel frequency was calculated as the ratio between the number of reads for each indel category and the total number of aligned reads. Heatmaps were generated in GraphPad Prism version 9.

Example 1

[0197] The type I-C CRISPR from *Neisseria meningitidis* (Nla) was the focus of initial engineering since it is currently the most streamlined and compact type I system that has been shown to work in human cells. High resolution Cryo-EM structures of this system at different stages of target recognition (FIG. 1) show that the Cascade complex makes extensive contact with the target strand DNA both before and after full R-loop formation.

[0198] All Glu and Asp residues on the Cascade subunits that are within 10Å of the target strand were identified DNA (FIGS. 2-5). Those identified residues were mutated to arginine one by one and tested to determine the effect on activity. The gene disruption activity of Cascade with subunits having those mutations was measured in GPF reporter cells along with WT Cas3 and a GFP targeting crRNA. As shown in FIGS. 6, 7, 9 and 10, a subset of those single arginine substitutions in Cas7 and Cas8 enhanced the GFP disruption efficiency of Cascade, especially when looking at the cell population with weak Cascade expression (FIGS. 7 and 10). Double or triple substitutions were made by combining promising single substitutions. All Cas8 double or triple substitutions showed a more robust GFP disruption in the reporter cell line (FIG. 8), while Cas7 double or triple substitutions did not (FIG. 10). Taken together, these data indicate creation of a series of mutants of Nla type I-C CRISPR that have enhanced activity at gene disruption.

[0199] The effect and potential enhancement of base editing efficiency for the mutants was tested using a Nla I-C CRISPR based adenine base editor (Nla-ABE). Nla ABE is created by fusing an evolved tadA domain to the C terminus of Cas11 subunit and replacing WT Cas3 with Cas3 nickase. As shown in FIGS. 11-26, a subset of the mutants significantly enhanced the editing efficiency of Nla ABE. The base editing activity enhancing mutations are in general the same mutations that also enhanced Cascade gene disruption activity with a few exceptions. Interestingly, the improvements vary from target to target. Targets that have low base editing efficiency (HBB, CFTR) with the WT Nla ABE showed the highest fold enhancement by the mutants. The mutants only have minor improvements on targets with moderate to high editing efficiency with the WT Nla ABE (HPRT1 and EGFP). By combining subunits with activity enhancing mutations, the base editing efficiency can be further increased at hard to edit sites.

[0200] To show the potential of the activity enhancing Cas mutants at a therapeutically relevant site, we tried to correct CFTR G542X mutation in iPS cells using Nla-ABE (FIG. 27A). CFTR G542X is the most common nonsense mutation that causes cystic fibrosis (CF). CF patients with this mutation do not benefit from currently FDA approved CFTR potentiators and correctors. This mutation has remained

inaccessible by Cas9 editors, due to the lack of suitable NGG or NGN PAMs nearby but is a perfect target for Nla-ABE with a TTT PAM (FIG. 27A). All Cas genes were delivered through electroporation in the form of mRNA and crRNA in the form of chemically synthesized small RNA with exonuclease resistant end modifications. As shown in FIG. 27B, all Cas mutants tested enhanced the on-target editing efficiency of Nla-ABE, with the best variant combination showing a >200-fold enhancement in base editing efficiency. All mutants showed modest indel generation and bystander edits (FIGS. 27C and 27D).

[0201] The guide-target mismatch tolerance of Cascade was tested with a subset of activity enhancing substitutions in Cas8 using a GFP disruption assay (FIG. 28). While enhanced Cascade showed higher tolerance for a single nucleotide mismatch, they showed similar profile with WT Cascade for guides with more significant mismatches (9 nt deletion and 4 nt mismatch).

[0202] A subset of Cascade mutants was tested using a crRNA targeting GFP with a sub-optimal CTC PAM in a GFP disruption assay. Wild type Cascade showed low GFP disruption activity using this crRNA. All Cas mutants tested enhanced the GFP disruption efficiency compared to that of the WT Cascade (FIG. 29). In general, Cas8 double (DM1 to 3) and triple (TM) mutants and the combination of Cas8 DM and TM with either Cas5 E76R or Cas7 E90R led to the most efficiency boost. Interestingly, combining Cas5, Cas7 and Cas8 mutants led to a decrease in GFP disruption efficiency.

[0203] It is contemplated that these activity enhancing mutations find use to enhance Cascade's activity in other applications, such as CRISPR activation and CRISPR inhibition. Furthermore, this approach is generally applicable to all type I systems to enhance their gene editing activity.

Sequences

Nla-Cas8 protein sequence (SEQ ID NO: 1)

MILHALTQYYQRKAESDGGIAQEGFENKEIPFIIVIDKQGNFIQLEDTRELKVKKKVGRTFLVPKG
LGRSGSKSYEVSNLLWDHYGYVLAYAGEKGQEQADKQHASFTAKVNELKQALPDDAGVTAVA
AFLSSAEEKSKVMQAANWAECAKVKGCNLSFRLVDEAVDLVCQSKAVREYVSQANQTQSDNV
QKGICLVTGKAAPIARLHNAVKGVNAKPAPFASVNLSAFESYGKEQGFIFPVGEQAMFEYTTAL
NTLLASENRFRIGDVTAVCWGAKRTPLEESLASMINGGGKDKPDEHIDAVKTLYKSLYNGQYQK
PDGKEKFYLLGLSPNSARIVVRFWHETTVAALSESIAAWYDDLQMVRGENSPYPEYMPLPRLLG
NLVLDGKMENLPSDLIAQITDAALNNRVLPVSLLQAALRRNKAEQKITYGRASLLKAYINRAIRA
GRLKNMKELTMGLDRNRQDIGYVLGRLFAVLEKIQAEANPGLNATIADRYFGSASSTPIAVFGTL

MRLLPHHLNKLEFEGRAVQLQWEIRQILEHCQRFPNHLNLEQQGLFAIGYYHETQFLFTKDALK
NLFNEAKTA

Nla-Cas5 (SEQ ID NO: 2)
MRFILEISGDLACFTRSELKVERVSYPVITPSAARNILMAILWKPAIRWKVLKIEILKPIQWTNIRRN
EVGTKMSERSGSLYIEDNRQQRASMLLKDVAYRIHADFDMTSEAGESDNYVKFAEMFKRRAKK
GQYFHQPYLGCREFPCDFRLLEKAEDGLPLEDITQDFGFMLYDMDFSKSDPRDSNNAEPMFYQC
KAVNGVITVPPADSEEVKR

Nla-Cas7 protein sequence (SEQ ID NO: 3)
MTIEKRYDFVFLFDVQDGNPNGDPDAGNLPRIDPQTGEGLVTDVCLKRKVRNFIQMTQNDEHHD
IFIREKGILNNLIDEAHEQENVKGKEKGEKTEAARQYMCSRYYDIRTFGAVMTTGKNAGQVRGP
VQLTFSRSIDPIMTLEHSITRMAVTNEKDASETGDNRTMGRKFTVPYGLYRCHGFISTHFAKQTG
FSENDLELFWQALVNMFDHDHSAARGQMNARGLYVFEHSNNLGDAPADSLFKRIQVVKKDGV
EVVRSFDDYLVSVDDKNLEETKLLRKLG

Nla-Cas11 protein sequence (SEQ ID NO: 4)
MGLDRNRQDIGYVLGRLFAVLEKIQAEANPGLNATIADRYFGSASSTPIAVFGTLMRLLPHHLNK
LEFEGRAVQLQWEIRQILEHCQRFPNHLNLEQQGLFAIGYYHETQFLFTKDALKNLFNEAKTA

Nla-Cas3 protein sequence (SEQ ID NO: 5)
MNFDYIAHARQDSSKNWHSHPLQKHLQKVAQLAKRFAGRYGSLFAEYAGLLHDLGKFQESFQK
YIRNASGFEKENAHLEDVESTKLRKIPHSTAGAKYAVERLNPFFGHLLAYLIAGHHAGLADWYD
KGSLKRRLQQADDELAASLSGFVESSLPEDFFPLSDDDLMRDFFAFWEDGAKLEELHIWMRFLFS
CLVDADFLDTEAFMNGYADADTAQAAGLRPKFPGLDELHRRYEQYMAQLSEKADKNSSLNQE
RHAILQQCFSAAETDRTLFSLTVPTGGGKTLASLGFALKHALKFGKKRIIYAIPFTSIIEQNANVFR
NALGDDVVLEHHSNLEVKEDKETAKTRLATENWDAPLIVTTNVQLFESLFAAKTSRCRKIHNIA
DSVVILDEAQQLPRDFQKPITDMMRVLARDYGVTFVLCTATQPELGKNIDAFGRTILEGLPDVRE
IVADKIALSEKLRRVRIKMPPPNGETQSWQKIADEIAARPCVLAVVNTRKHAQKLFAALPSNGIK
LHLSANMCATHCSEVIALVRRYLALYRAGSLHKPLWLVSTQLIEAGVDLDFPCVYRAMAGLDSI
AQAAGRCNREGKLPQLGEVVVFRAEEGAPSGSLKQGQDITEEMLKAGLLDDPLSPLAFAEYFRR

FNGKGDVDKHGITTLLTAEASNENPLAIKFRTAAERFHLIDNQGVALIVPFIPLAHWEKDGSPQIV
EANELDDFFRRHLDGVEVSEWQDILDKQRFPQPPDNSFGQTDQPLLPEPFESWFGLLESDPLKHK
WVYRKLQRYTITVYEHELKKLPEHAVFSRAGLLVLDKGYYKAVLGADFDDAAWLPENSVL

Bha-IC Cas 7 (Csd2) protein sequence (SEQ ID NO: 6)
MTILDHKIDFAVILSVTKANPNGDPLNGNRPRQNYDGHGEISDVAIKRKIRNRLLDMEEPIFVQSD
DRKADSFKSLRDRADSNPELAKMLKAKNASVDEFAKIACQEWMDVRSFGQVFAFKGSNLSVGV
RGPVSIHTATSIDPIDIVSTQITKSVNSVTGDKRSSDTMGMKHRVDFGVYVFKGSINTQLAEKTGF
TNEDAEKIKRALITLFENDSSSARPDGSMEVHKVYWWEHSSKLGQYSSAKVHRSLKIESKTDTPK
SFDDYAVELYELDGLGVEVIDGQ

Bha-IC Cas5 protein sequence (SEQ ID NO: 7)
MRNEVQFELFGDYALFTDPLTKIGGEKLSYSVPTYQALKGIAESIYWKPTIVFVIDELRVMKPIQM
ESKGVRPIEYGGGNTLAHYTYLKDVHYQVKAHFEFNLHRPDLAFDRNEGKHYSILQRSLKAGGR
RDIFLGARECQGYVAPCEFGSGDGFYDGQGKYHLGTMVHGFNYPDETGQHQLDVRLWSAVME
NGYIQFPRPEDCPIVRPVKEMEPKIFNPDNVQSAEQLLHDLGGE

Bha-IC Cas 8 (Csd1) protein sequence (SEQ ID NO: 8)
MSWLLHLYETYEANLDQVGKTVKKGEDREYTLLPISHTTQNAHIEVTLDEDGDFLRAKALTKES
TLIPCTEEAASRSGSKVAPYPLHDKLSYVAGDFVKYGGKIKNQDDAPFDTYIKNLGEWANSPYA
TEKVKCIYTYLKKGRLIEDLVDAGVLKLDENQQLIEKWEKRYEELLGEKPAIFSSGATDQASAFV
RFNVFHPESIDDVWKDKEMFDSFISFYNDKLGEEDICFVTGNRLPSTERHANKIRHAADKAKLISA
NDNSGFTFRGRFKTSREAVGISYEVSQKAHNALKWLIHRQSKSIDDRVFLVWSNDNSLVPNPDE
DAVDIMKHANRELERDPDTGQIFAGEVKKAIGGYRSDLNYQPEVHILVLDSATTGRMAVLYYRS
LNKELYLNRLEAWHDSCAWEHRYRRDEKEFISFYGAPATKDIAFAAYGPRASEKVIKDLMERML
PCIVDGRRVPKDIVRSAFQRASNPVSMERWEWEKTLSITCALIRKMHIEQKEEWGVPLDKSSTDR
SYLFGRLLAVADVLERGALGKDETRATNAIRYMNSYSKNPGRTWKTIQESLQPYQAKLGTKAT
YLSKLVDEIGDQFEPGDFNNNPLTEQYLLGFYSQRRELYKKKEEETNQ

Bha-IC Cas11 protein sequence (SEQ ID NO: 9)

MPLDKSSTDRSYLFGRLLAVADVLERGALGKDETRATNAIRYMNSYSKNPGRTWKTIQESLQPY
QAKLGTKATYLSKLVDEIGDQFEPGDFNNNPLTEQYLLGFYSQRRELYKKKEEETNQ

Bha-IC Cas3 protein sequence (SEQ ID NO: 10)
MYIAHIREVDKVIQTLKEHLCGVQCLAETFGAKLRLQHVAGLAGLLHDLGKYTNEFKDYIYKAV
FEPELAEKKRGQVDHSTAGGRLLYQMLHDRENSFHEKLLAEVVGNAIISHHSNLQDYISPTIESNF
LTRVLEKELPEYESAVERFFQEVMTEAELARYVAKAVDEIKQFTDNSPTQSFFLTKYIFSCLIDAD
RTNTRMFDEQAREEEPTQPQQLFEHYHQQLLNHLASLKESDSAQKPINVLRSAMSEQCESFAMR
PSGIYTLSIPTGGGKTLASLRYALKHAQEYNKQRIIYIVPFTTIIEQNAQEVRNILGDDENILEHHSN
VVEDSENGDEQEDGVITKKERLRLARDNWDRPIIFTTLVQFLNVFYAKGNRNTRRLHNLSHSVLI
FDEVQKVPTKCVSLFNEALNFLKEFAHCSILLCTATQPTLENVKHSLLKDRDGEIVQNLTEVSEAF
KRVEILDKTDQPMTNERLAEWVRDEAPSWGSTLIILNTKKVVKDLYEKLEGGPLPVFHLSTSMC
AAHRKDQLDEIRALLKEGTPFICVTTQLIEAGVDVSFKCVIRSLAGLDSIAQAAGRCNRHGEEQL
QYVYVIDHAEETLSKLKEIEVGQEIAGNVLARFKKKAEKYEGNLLSQAAMREYFRYYYSKMDA
NLNYFVKEVDKDMTKLLMSHAVENSYVTYYQKNTGTHFPLLLNGSYKTAADHFRVIDQNTTSA
IVPYGEGQDIIAQLNSGEWVDDLSKVLKKAQQYTVNLYSQEIDQLKKEGAIVMHLDGMVYELK
ESWYSHQYGVDFKGEGGMDFMSF

Dvu-IC Cas5 protein sequence (SEQ ID NO: 11)
MTHGAVKTYGIRLRVWGDYACFTRPEMKVERVSYDVMPPSAARGILEAIHWKPAIRWIVDRIHV
LRPIVFDNVRRNEVSSKIPKPNPATAMRDRKPLYFLVDDGSNRQQRAATLLRNVDYVIEAHFELT
DKAGAEDNAGKHLDIFRRRARAGQSFQQPCLGCREFPASFELLEGDVPLSCYAGEKRDLGYMLL
DIDFERDMTPLFFKAVMEDGVITPPSRTSPEVRA

Dvu-IC Cas8 protein sequence (SEQ ID NO: 12)
MILQALHGYYQRMSADPDAGMPPYGTSMENISFALVLDAKGTLRGIEDLREQEGKKLRPRKML
VPIAEKKGNGIKPNFLWENTSYILGVDAKGKQERTDKCHAAFIAHIKAYCDTADQDLAAVLQFL
EHGEKDLSAFPVSEEVIGSNIVFRIEGEPGFVHERPAARQAWANCLNRREQGLCGQCLITGERQK
PIAQLHPSIKGGRDGVRGAQAVASIVSFNNTAFESYGKEQSINAPVSQEAAFSYVTALNYLLNPSN
RQKVTIADATVVFWAERSSPAEDIFAGMFDPPSTTAKPESSNGTPPEDSEEGSQPDTARDDPHAA
ARMHDLLVAIRSGKRATDIMPDMDESVRFHVLGLSPNAARLSVRFWEVDTVGHMLDKVGRHY

RELEIIPQFNNEQEFPSLSTLLRQTAVLNKTENISPVLAGGLFRAMLTGGPYPQSLLPAVLGRIRAE
HARPEDKSRYRLEVVTYYRAALIKAYLIRNRKLEVPVSLDPARTDRPYLLGRLFAVLEKAQEDA
VPGANATIKDRYLASASANPGQVFHMLLKNASNHTAKLRKDPERKGSAIHYEIMMQEIIDNISDF
PVTMSSDEQGLFMIGYYHQRKALFTKKNKEN

Dvu-IC Cas7 protein sequence (SEQ ID NO: 13)
MTAIANRYEFVLLFDVENGNPNGDPDAGNMPRIDPETGHGLVTDVCLKRKIRNHVALTKEGAER
FNIYIQEKAILNETHERAYTACDLKPEPKKLPKKVEDAKRVTDWMCTNFYDIRTFGAVMTTEVN
CGQVRGPVQMAFARSVEPVVPQEVSITRMAVTTKAEAEKQQGDNRTMGRKHIVPYGLYVAHGF
ISAPLAEKTGFSDEDLTLFWDALVNMFEHDRSAARGLMSSRKLIVFKHQNRLGNAPAHKLFDLV
KVSRAEGSSGPARSFADYAVTVGQAPEGVEVKEML

Dvu-IC Cas11 protein sequence (SEQ ID NO: 14)
MSLDPARTDRPYLLGRLFAVLEKAQEDAVPGANATIKDRYLASASANPGQVFHMLLKNASNHT
AKLRKDPERKGSAIHYEIMMQEIIDNISDFPVTMSSDEQGLFMIGYYHQRKALFTKKNKEN

Dvu-IC Cas3 protein sequence (SEQ ID NO: 15)
MADGGDEHASGHDNTLKNARYYAHSTPNPDKSDWQGLDAHLENVANLAATFAEAFGAREWG
KAAGLLHDAGKATAQFTQRLEGRPVRVNHSICGARLAQEQGSTCGLLLSYAIAGHHGGLPDGGL
QDGQLHHRLKHERLPADVSPPSVDIRPDVLKPPFTCRPEHPGFSLSFFTRMLFSCLTDADFLDTEA
FCTPEKASARNGRSALGLVALRDALNTHLDTVERKALPSRVNDIRKTVLHDCRARASETPGLFSL
TVPTGGGKTLSSMAFALDHAVTHGLRRVIYAIPFTSIIEQNAKVFSDVFGQDNVLEHHCNYRSKD
EPEEQGYDKWRGLAAENWDAPVVVTTNVQFFESLFSNRPSRCRKLHNIARSVIVLDEAQAIPTEY
LEPCLYALKELVGQYGCTVVLCTATQPAVDDASLPERVRLHHVREIIADPQRLYTDLKRTEVTLA
GRLTDAALAARLDGHGQVLCIVGTKPQAQAVFSLLQEREGAFHLSTNMYPEHRRRVLGTIRQRL
ADRLPCRVVSTSLIEAGVDVDFPVVYRAMAGLDSIAQAAGRCNREGRLPEPGQVVVYEPEKPAR
MPWMQRCASRAQETLRTLPEADPLGLEAIRRYFGLIYDVQELDRKDIFKRLRGQVDRDMVFKFR
EIANDFRFIDDEGTALVIPTGPEVEDLVRRLRGCEFPRPVLRKLQQYSVTVRHRELEKLRSAGAVE
MIGDAYPVLRNLAAYSEDMGLCVDSVEVWQPEGLVS

Nla type I-C CRISPR repeat (show as DNA equivalent):

TCAGCCGCCTCTAGGCGGCTGTGTGTTGAAAC (SEQ ID NO: 20)

crRNA spacer sequences used (show as DNA equivalents):

For GFP disruption

GFP guide: GTGACCGCCGCCGGGATCACTCTCGGCATGGACGA (SEQ ID NO: 21)

GFP 3'UTR CTC PAM: CCACTGTCCTTTCCTAATAAAATGAGGAAATTGCA (SEQ ID NO: 22)

For base editing

GFP: GAGGGCGACACCCTGGTGAACCGCATCGAGCTGAA (SEQ ID NO: 23)

HPRT1: CTCATCTGTAAAATGGTAATAATCATACCATTGCT (SEQ ID NO: 24)

HBB: ACTAGCAACCTCAAACAGACACCATGGTGCATCTG (SEQ ID NO: 25)

CFTR-35nt: GGTAATAGGACATCTCCAAGTTTGCAGAGAAAGAC (SEQ ID NO: 26)

CFTR-41nt: GGTAATAGGACATCTCCAAGTTTGCAGAGAAAGACAATATA (SEQ ID NO: 27)

[0204] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[0205] Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

CLAIMS

What is claimed is:

1. An engineered Cas protein from a Type I-C CRISPR-Associated Complex for Anti-viral Defense (Cascade) complex having at least 70% identity to a wild-type protein and comprising one or more substitutions of a glutamate residue or aspartate residue within 10Å of a nucleic acid bound by the Cascade complex with a positively charged amino acid.

2. The engineered Cas protein of claim 1, wherein the positively charged amino acid is arginine, histidine, or lysine.

3. The engineered Cas protein of claim 1 or 2, wherein the positively charged amino acid is arginine.

4. The engineered Cas protein of any of claims 1-3, wherein the wild-type protein is a *Neisseria lactamica* type I-C Cas protein.

5. The engineered Cas protein of any of claims 1-4, wherein the wild-type protein is a Cas8 protein comprising an amino acid sequence of SEQ ID NO: 1 and the one or more substitutions are selected from residues: E76, E230, E235, D268, D387, E391, D396, E426, E479, E484, D495, E524, E526, or combinations thereof, relative to SEQ ID NO: 1.

6. The engineered Cas protein of claim 5, wherein the one or more substitutions comprise: E76R, E230R, E235R, D268R, D387R, E391R, D396R, E426R, E479R, E484R, D495R, E524R, E526R, or combinations thereof, relative to SEQ ID NO: 1.

7. The engineered Cas protein of claim 5 or 6, wherein the one or more substitutions are selected from residues: E235/D387, E235/E391, D387/E391, or E235/D387/E391, relative to SEQ ID NO: 1.

8. The engineered Cas protein of any of claims 5-7, wherein the one or more substitutions comprise: E235R/D387R, E235R/E391R, D387R/E391R, or E235R/D387R/E391R, relative to SEQ ID NO: 1.

9. The engineered Cas protein of any of claims 1-4, wherein the wild-type protein is a Cas5 protein comprising an amino acid sequence of SEQ ID NO: 2 and the one or more substitutions are selected from residues: E18, E22, E69, E76, E84, D85, or combinations thereof, relative to SEQ ID NO: 2.

10. The engineered Cas protein of claim 9, wherein the one or more substitutions comprise: E18R, E22R, E69R, E76R, E84R, D85R, or combinations thereof, relative to SEQ ID NO: 2.

11. The engineered Cas protein of any of claims 1-4, wherein the wild-type protein is a Cas7 protein comprising an amino acid sequence of SEQ ID NO: 3 and the one or more substitutions are selected from residues: D23, D25, E69, D78, E90, D108, E144, E155, D157, E160, D163, or combinations thereof, relative to SEQ ID NO: 3.

12. The engineered Cas protein of claim 11, wherein the one or more substitutions comprise: D23R, D25R, E69R, D78R, E90R, D108R, E144R, E155R, D157R, E160R, D163R, or combinations thereof, relative to SEQ ID NO: 3.

13. The engineered Cas protein of claim 11 or 12, wherein the one or more substitutions are selected from residues: E90/D78, D78/E160, E90/E160, D78/E90/E160, relative to SEQ ID NO: 3.

14. The engineered Cas protein of any of claims 11-13, wherein the one or more substitutions comprise: E90R/D78R, D78R/E160R, E90R/E160R, D78R/E90R/E160R, relative to SEQ ID NO: 3.

15. The engineered Cas protein of any of claims 1-4, wherein the wild-type protein is a Cas11 protein comprising an amino acid sequence of SEQ ID NO: 4 and the one or more substitutions are selected from residues: E22, E27, D38, E67, E69, or combinations thereof, relative to SEQ ID NO: 4.

16. The engineered Cas protein of claim 15, wherein the one or more substitutions comprise: E22R, E27R, D38R, E67R, E69R, or combinations thereof, relative to SEQ ID NO: 4.

17. A fusion protein comprising a Type I Cas protein of any of claims 1-16 and at least one effector domain.

18. The fusion protein of claim 17, wherein the at least one effector domain comprises a transcription activator, a transcription repressor, a base editor, an epigenetic modifier, or a combination thereof.

19. A system comprising:

Cas3, or a nucleic acid encoding thereof;

one or more engineered Cas protein of any of claims 1-16, a fusion protein of claims 17-18, and/or one or more nucleic acids encoding thereof; and

at least one guide RNA (gRNA), wherein each gRNA is configured to hybridize to a portion of a target nucleic acid sequence.

20. The system of claim 19, wherein the one or more Type I Cas protein comprises a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1.

21. The system of claim 19 or 20, wherein the one or more Type I Cas protein comprises: a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2.

22. The system of any of claims 19-21, wherein the one or more type I Cas protein comprises a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3.

23. The system of any of claims 19-22, wherein the one or more Type I Cas protein comprises:

i) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2;

ii) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1 and a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3;

iii) a Cas7 protein comprising an amino acid sequence having an E90R substitution relative to SEQ ID NO: 3 and a Cas5 protein comprising an amino acid sequence having an E76R substitution relative to SEQ ID NO: 2; or

iv) a Cas8 protein comprising an amino acid sequence having one or more substitutions selected from: E235R, D387R, E391R, or combinations thereof, relative to SEQ ID NO: 1, a Cas7 protein comprising an amino acid sequence with an E90R substitution relative to SEQ ID NO: 3, and a Cas5 protein comprising an amino acid sequence with an E76R substitution relative to SEQ ID NO: 2.

24. The system of any of claims 19-23, wherein the Cas3 is fully or partially catalytically inactive.

25. A method of altering a target nucleic acid sequence comprising contacting a target nucleic acid sequence with a system of any of claims 19-24.

26. The method of claim 25, wherein the altering comprises a deletion.

27. The method of claim 25 or 26, wherein the target nucleic acid sequence encodes a gene product.

28. The method of any of claims 25-27, wherein the target nucleic acid sequence is a genomic DNA sequence.

29. The method of any of claims 25-28, wherein contacting a target nucleic acid sequence comprises introducing the system into the cell.

30. A method for recruiting one or more effector domains to a target nucleic acid in a cell comprising:
introducing into a cell a fusion protein of any of claims 17-18 or a system of any of claims 19-24.

31. A method for modulating expression of a target gene in a cell comprising:
introducing into a cell a fusion protein of any of claims 17-18 or a system of any of claims 19-24.

32. The method of claim 31, wherein the Cas3 is fully or partially catalytically inactive.

33. The method of claim 31 or 32, wherein the target nucleic acid comprises the promoter region or the upstream activator sequence of the target gene.

34. The method of any of claims 29-33, wherein the cell is a prokaryotic cell.

35. The method of any of claims 29-33, wherein the cell is a eukaryotic cell.

36. The method of claim 35, wherein the cell is a mammalian cell.

37. The method of claim 35 or 36, wherein the cell is a human cell.

38. The method of any of claims 29-37, wherein the introducing into the cell comprises administering the system to a subject.

39. The method of claim 38, wherein the administering comprises in vivo administration.

40. A system of any of claims 19-24, or a composition comprising thereof, for use in altering a target nucleic acid sequence comprising contacting a target nucleic acid sequence with the system.

41. A fusion protein of any of claims 17-18 or a system of any of claims 19-24, or a composition comprising thereof, for use in recruiting one or more effector domains to a target nucleic acid in a cell and/or modulating expression of a target gene in a cell.

42. An engineered Cas protein of any of claims 1-16, a fusion protein of any of claims 17-18, or a system of any of claims 19-24, for use in altering a target nucleic acid sequence.

43. An engineered Cas protein of any of claims 1-16, a fusion protein of any of claims 17-18, or a system of any of claims 19-24, for use in recruiting one or more effector domains to a target nucleic acid in a cell and/or modulating expression of a target gene in a cell.

FIG. 1A

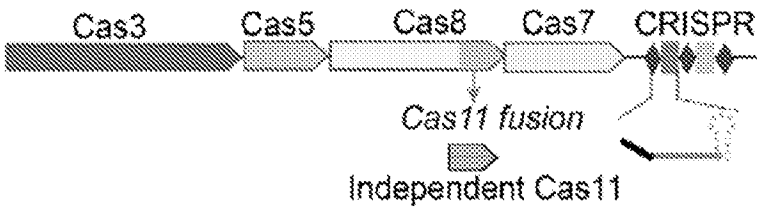


FIG. 1B



FIG. 1C

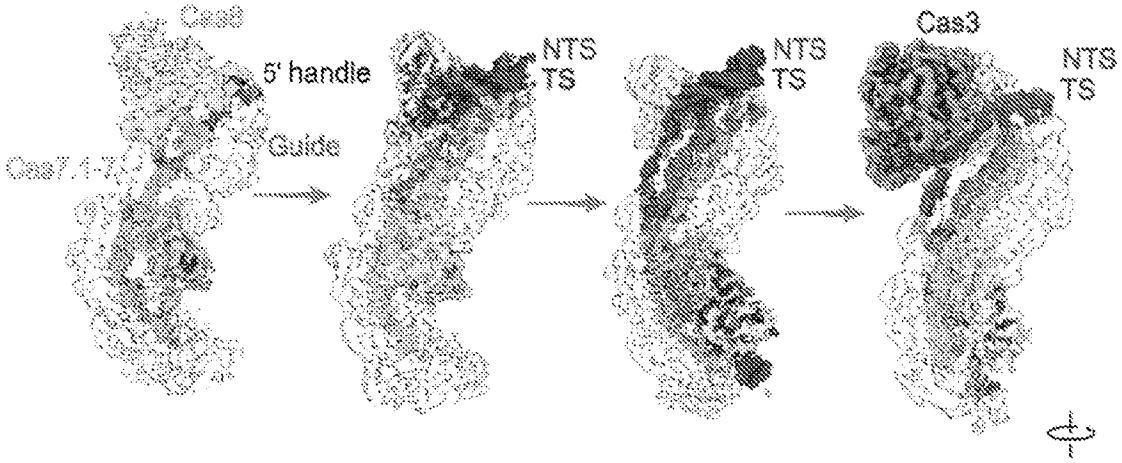


FIG. 1D

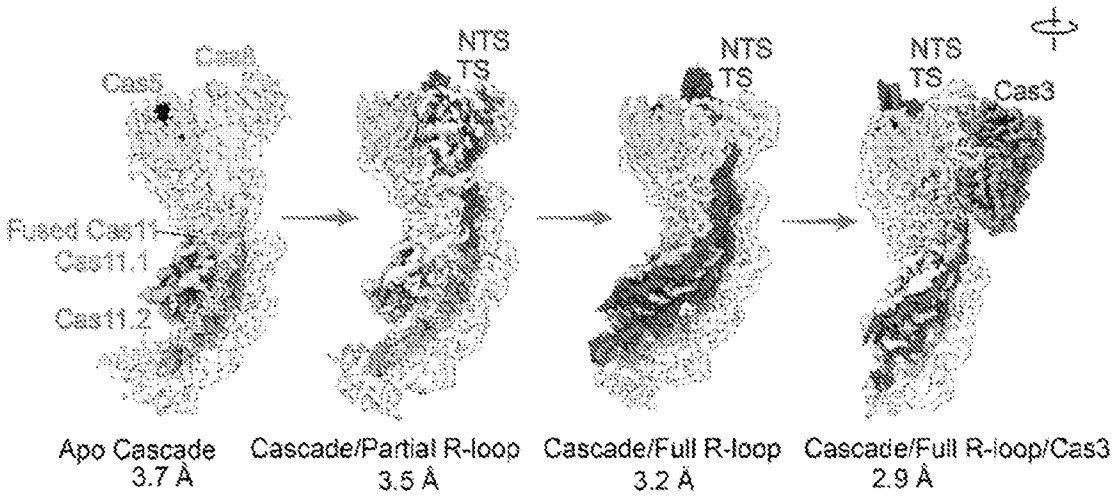


FIG. 2

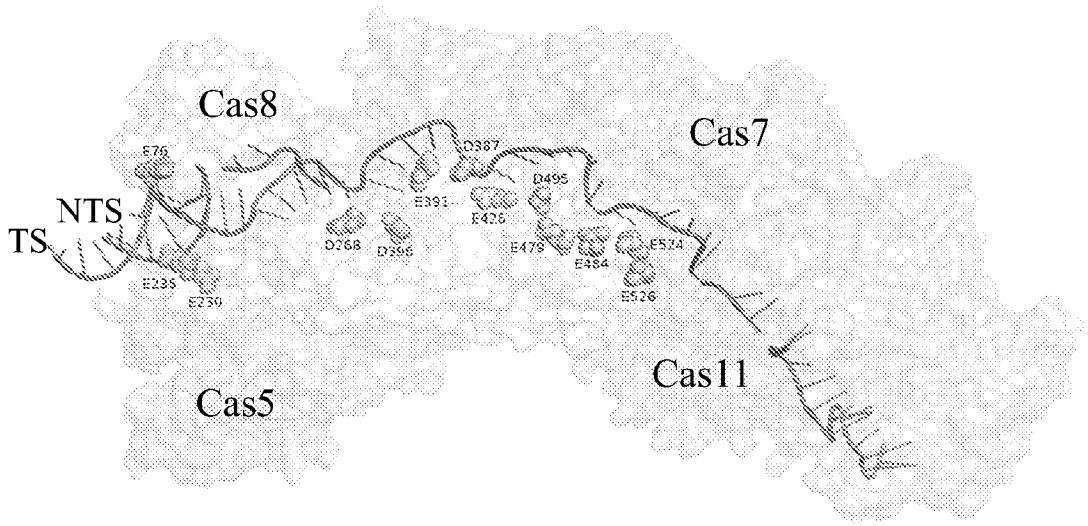

Cas8
Cas7
Cas5
Cas11
NTS
TS
E76
D387
D495
E391
E426
D268
D396
E478
E484
E524
E526
E235
E239

FIG. 3

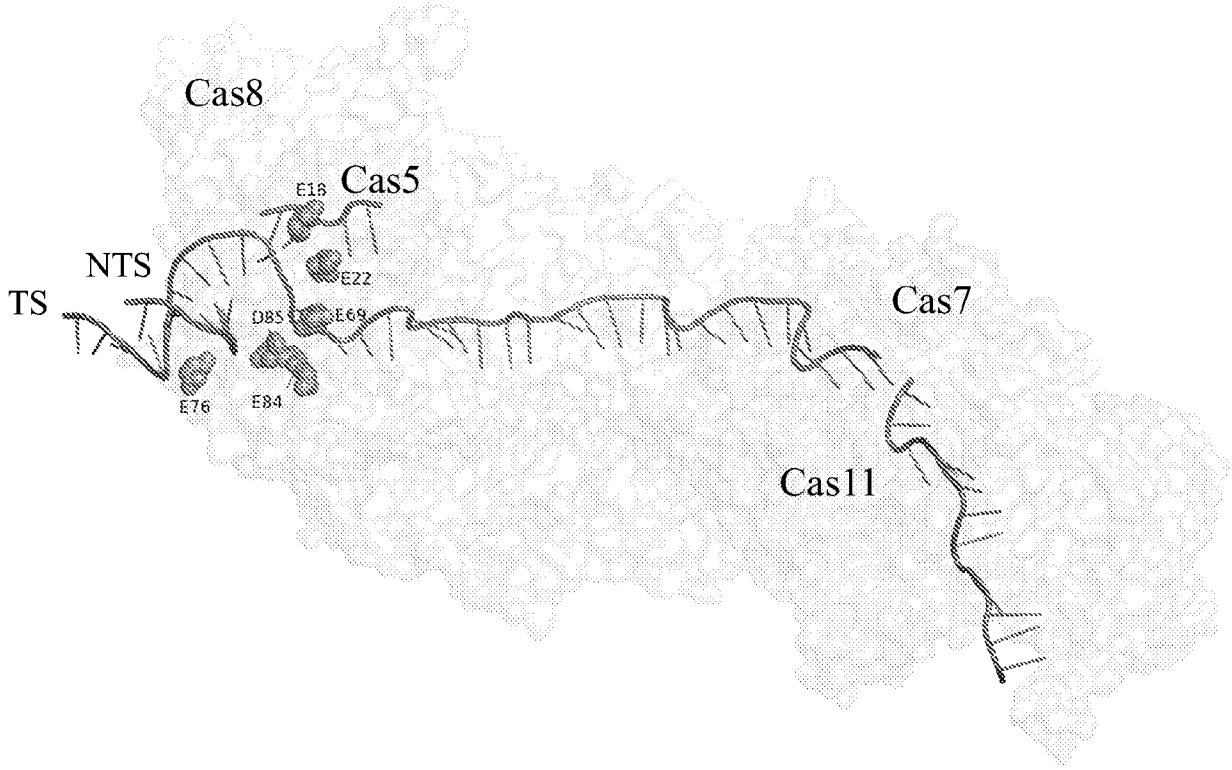


FIG. 4

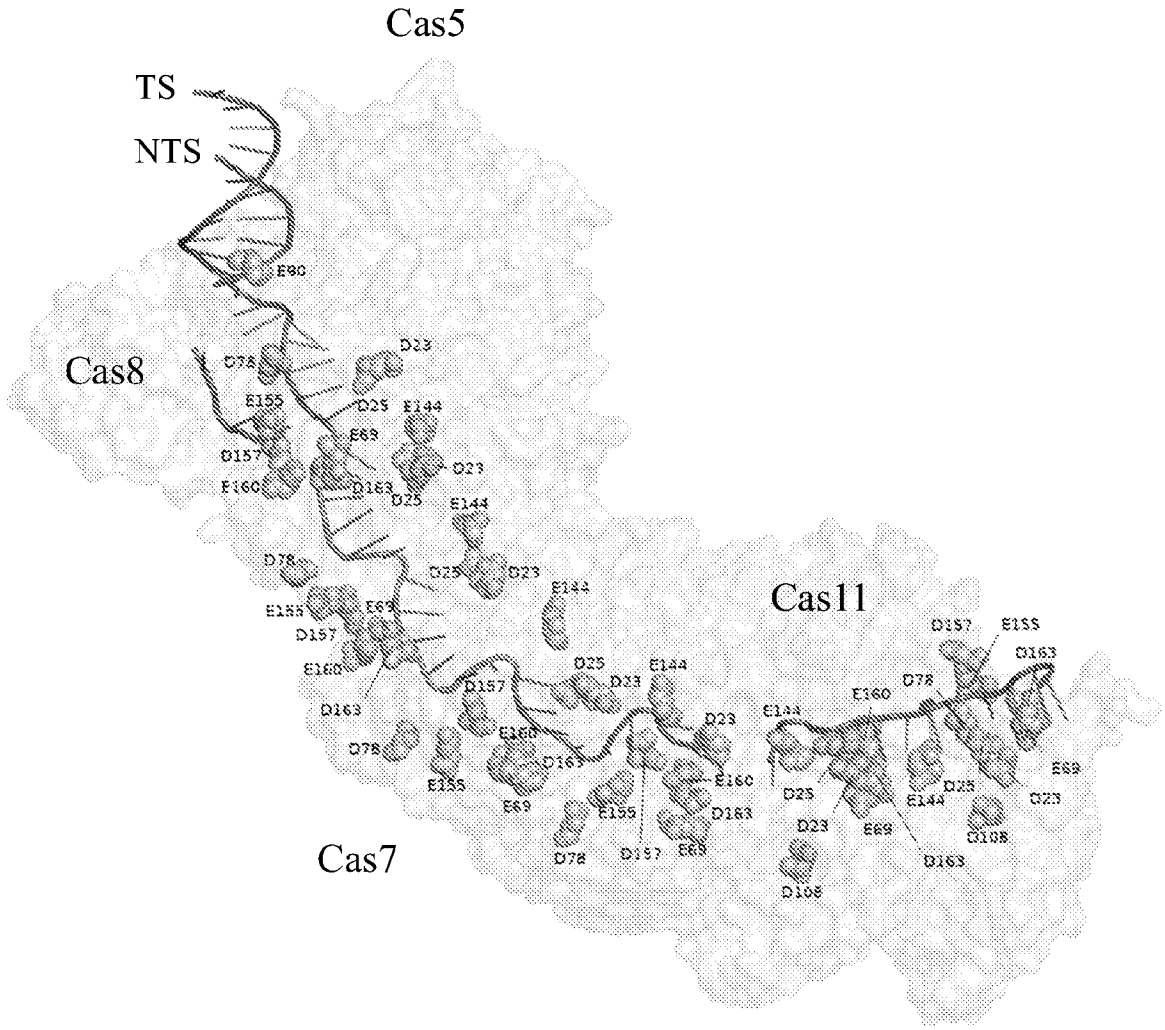


FIG. 5

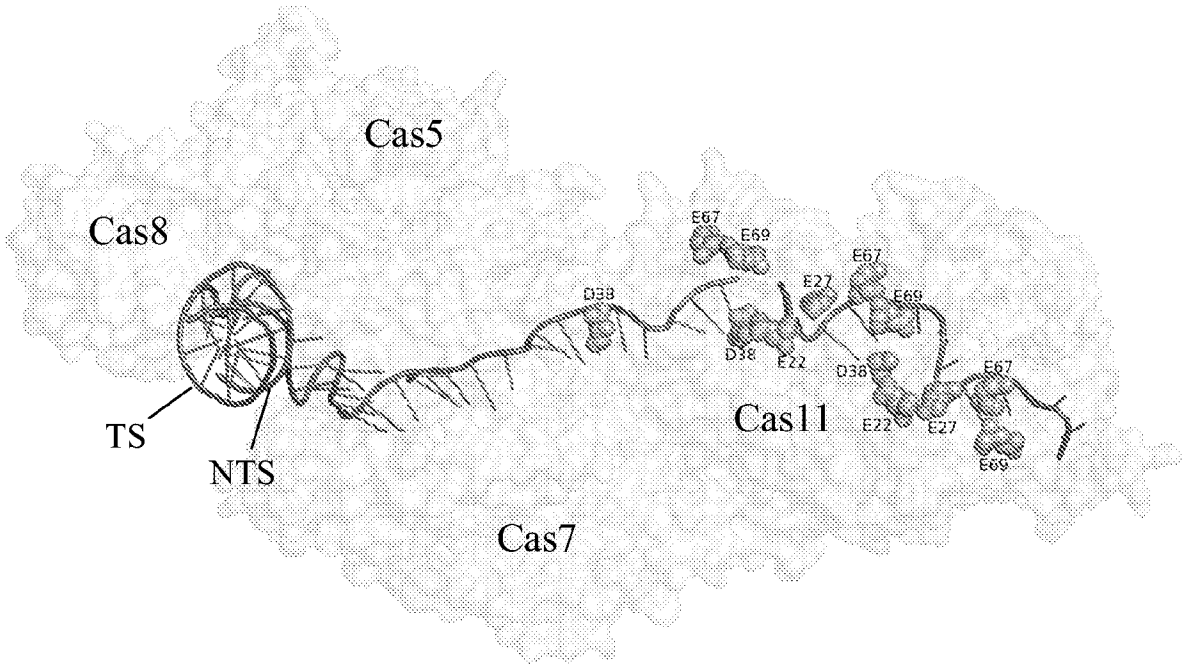

Cas5
Cas8
E67
E69
E67
E27
E69
D38
D38
E22
D38
E67
Cas11
E22
E27
E69
TS
NTS
Cas7

FIG. 6

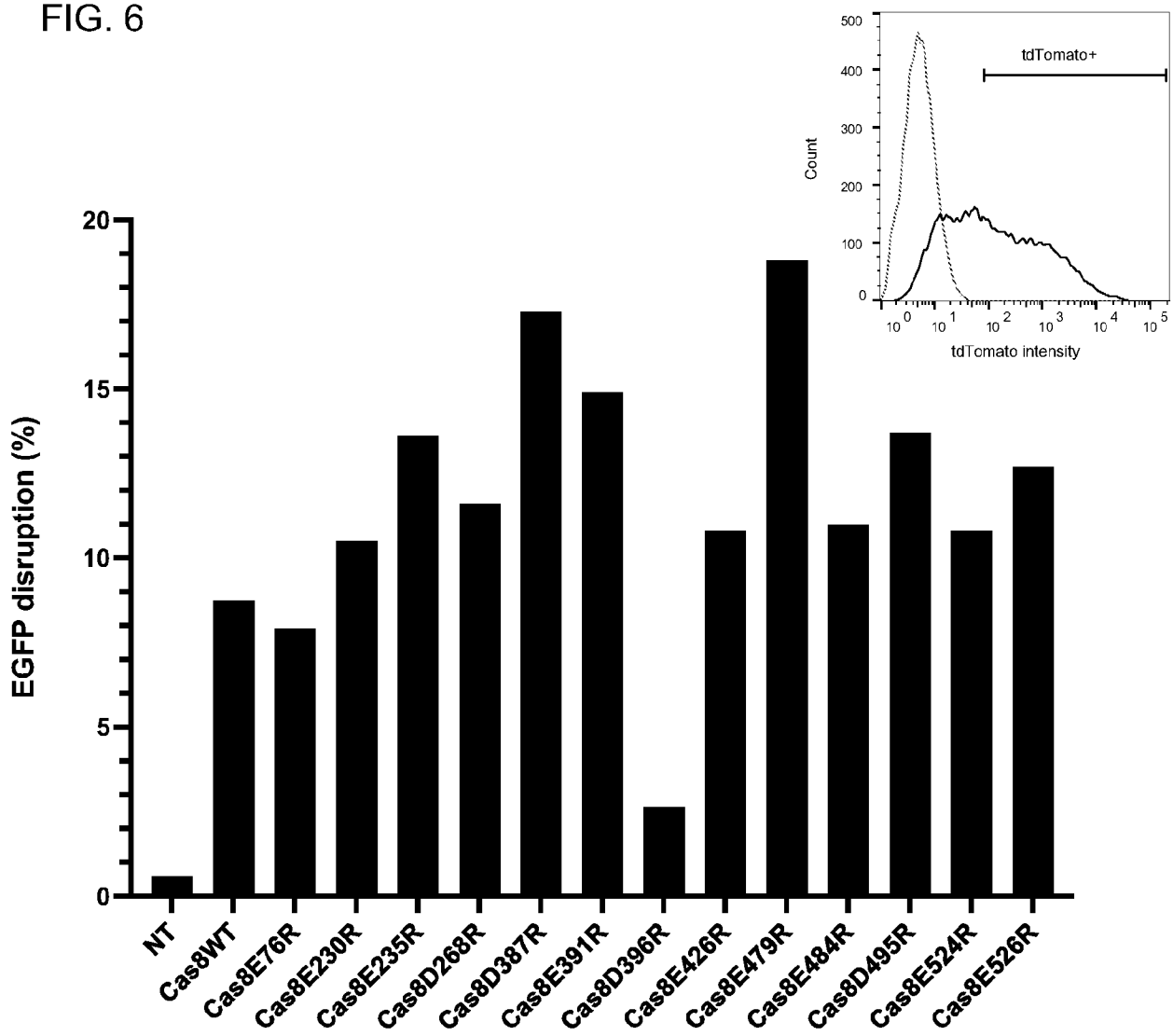

EGFP disruption (%)
0
5
10
15
20
NT
Cas8WT
Cas8E76R
Cas8E230R
Cas8E235R
Cas8D268R
Cas8D387R
Cas8E391R
Cas8D396R
Cas8E426R
Cas8E479R
Cas8E484R
Cas8D495R
Cas8E524R
Cas8E526R
Count
tdTomato+
tdTomato intensity

FIG. 7

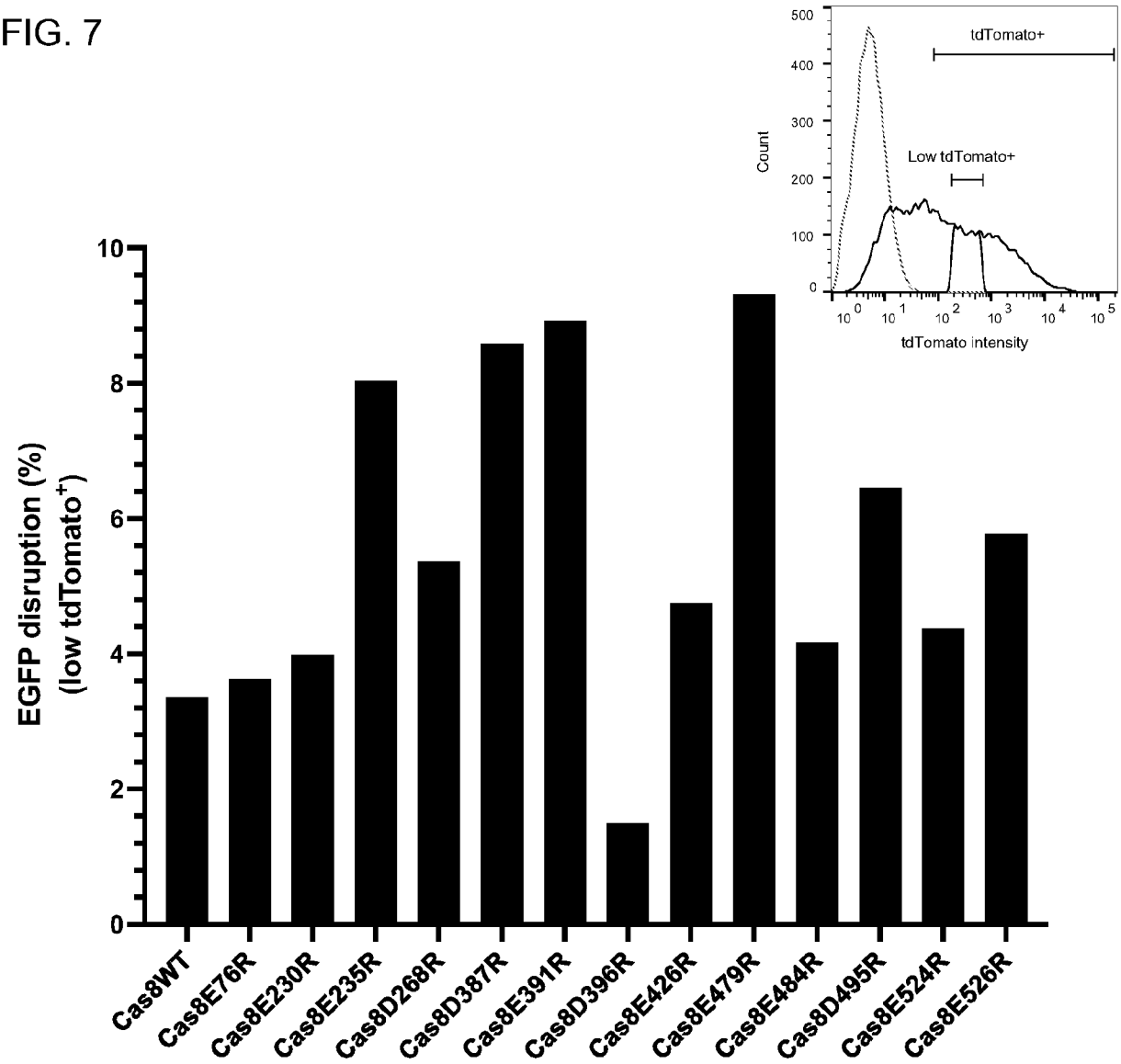

EGFP disruption (%)
(low tdTomato+)
0
2
4
6
8
10
Cas8WT
Cas8E76R
Cas8E230R
Cas8E235R
Cas8D268R
Cas8D387R
Cas8E391R
Cas8D396R
Cas8E426R
Cas8E479R
Cas8E484R
Cas8D495R
Cas8E524R
Cas8E526R
Count
tdTomato+
Low tdTomato+
tdTomato intensity

FIG. 8

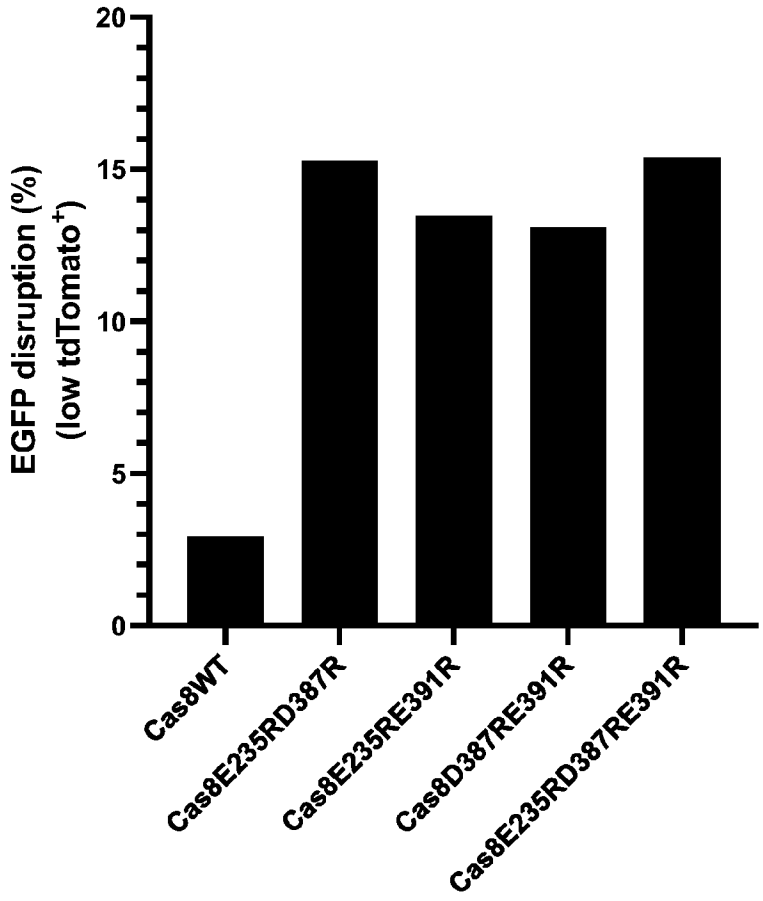

EGFP disruption (%)
(low tdTomato+)
20
15
10
5
0
Cas8WT
Cas8E235RD387R
Cas8E235RE391R
Cas8D387RE391R
Cas8E235RD387RE391R

FIG. 9

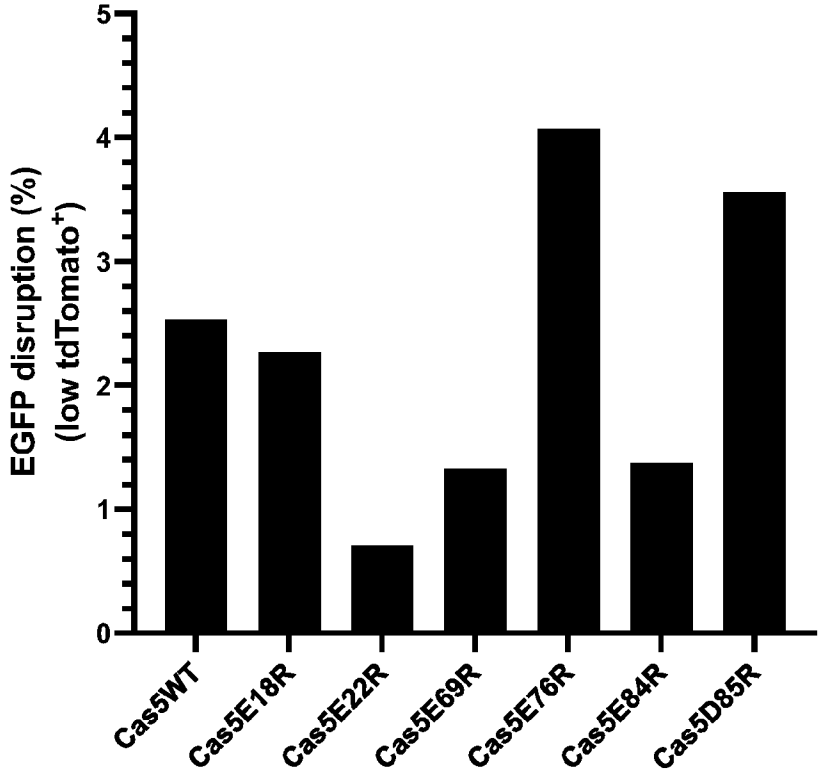

EGFP disruption (%)
(low tdTomato+)
0
1
2
3
4
5
Cas5WT
Cas5E18R
Cas5E22R
Cas5E69R
Cas5E76R
Cas5E84R
Cas5D85R

FIG. 10

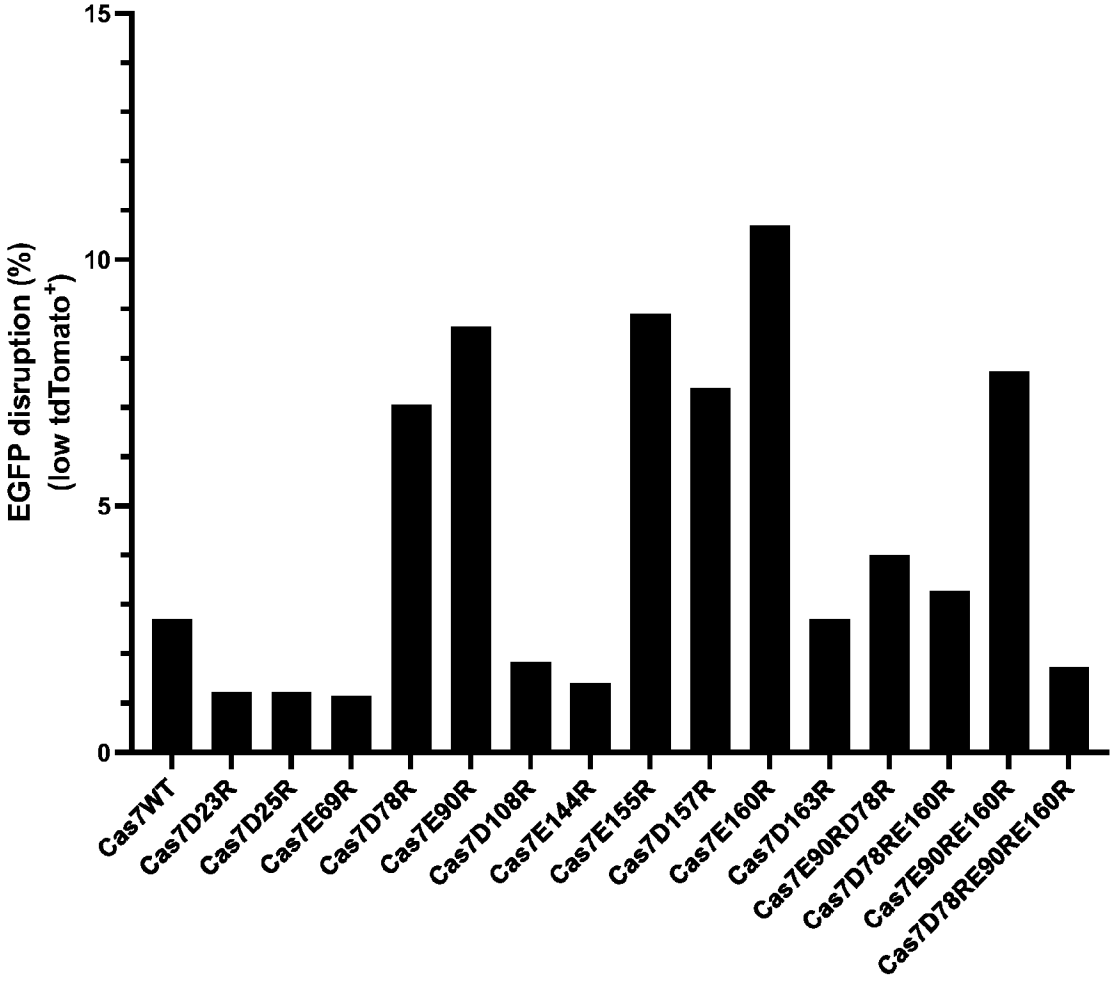

EGFP disruption (%)
(low tdTomato+)
15
10
5
0
Cas7WT
Cas7D23R
Cas7D25R
Cas7E69R
Cas7D78R
Cas7E90R
Cas7D108R
Cas7E144R
Cas7E155R
Cas7D157R
Cas7E160R
Cas7D163R
Cas7E90RD78R
Cas7D78RE160R
Cas7E90RE160R
Cas7D78RE90RE160R

FIG. 11A

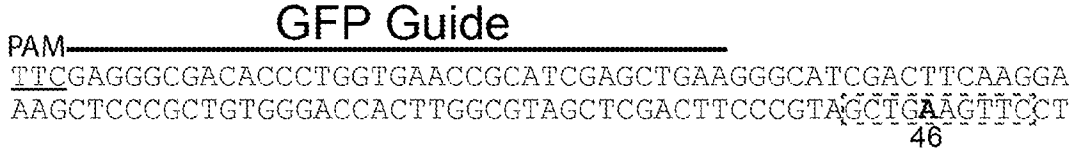


FIG. 11B

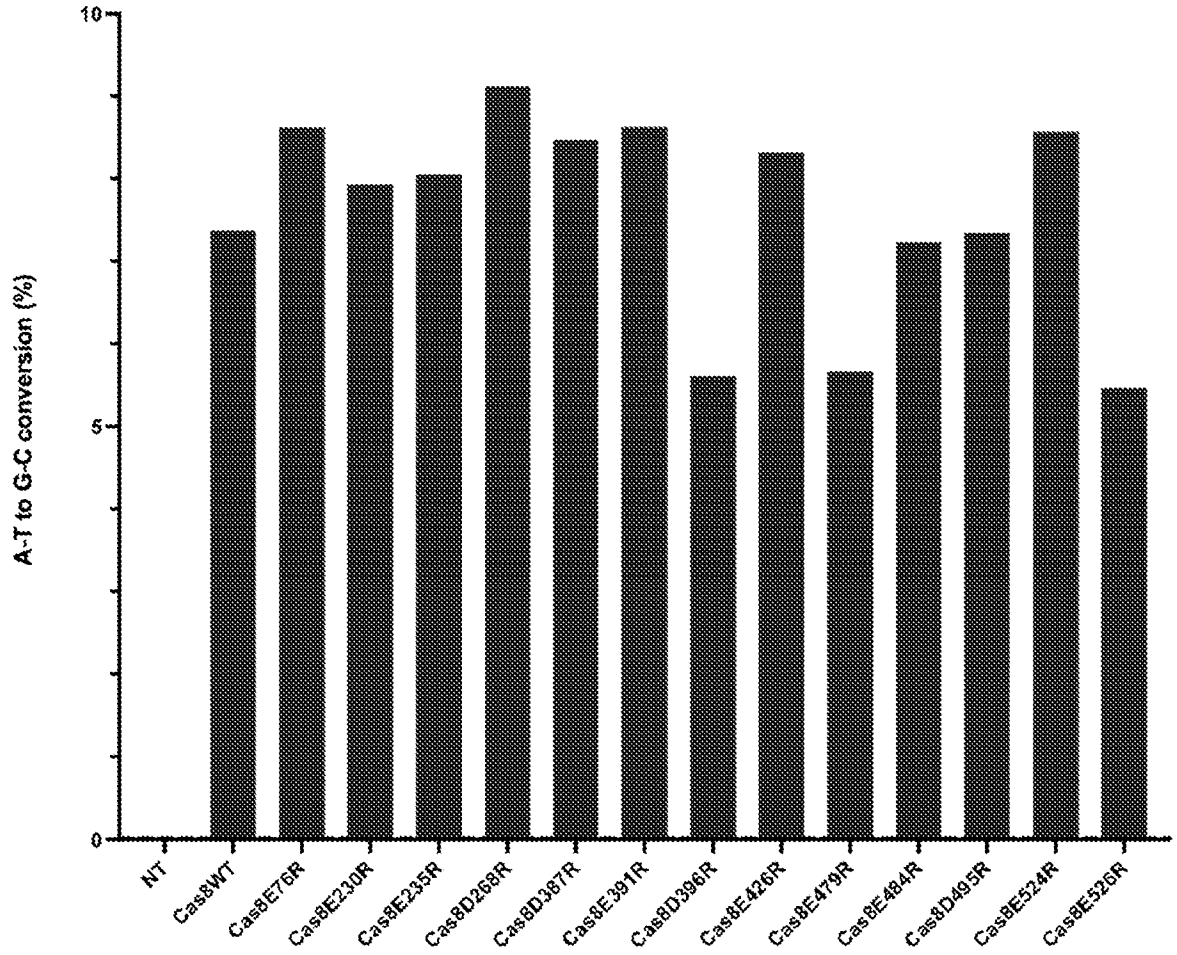


FIG. 12A

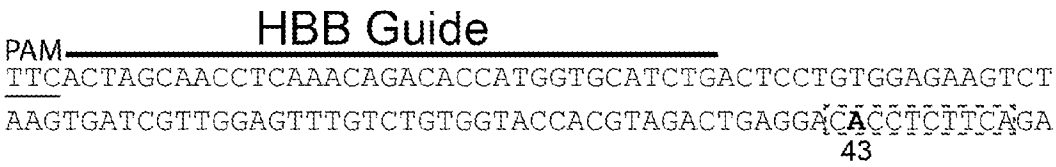


FIG. 12B

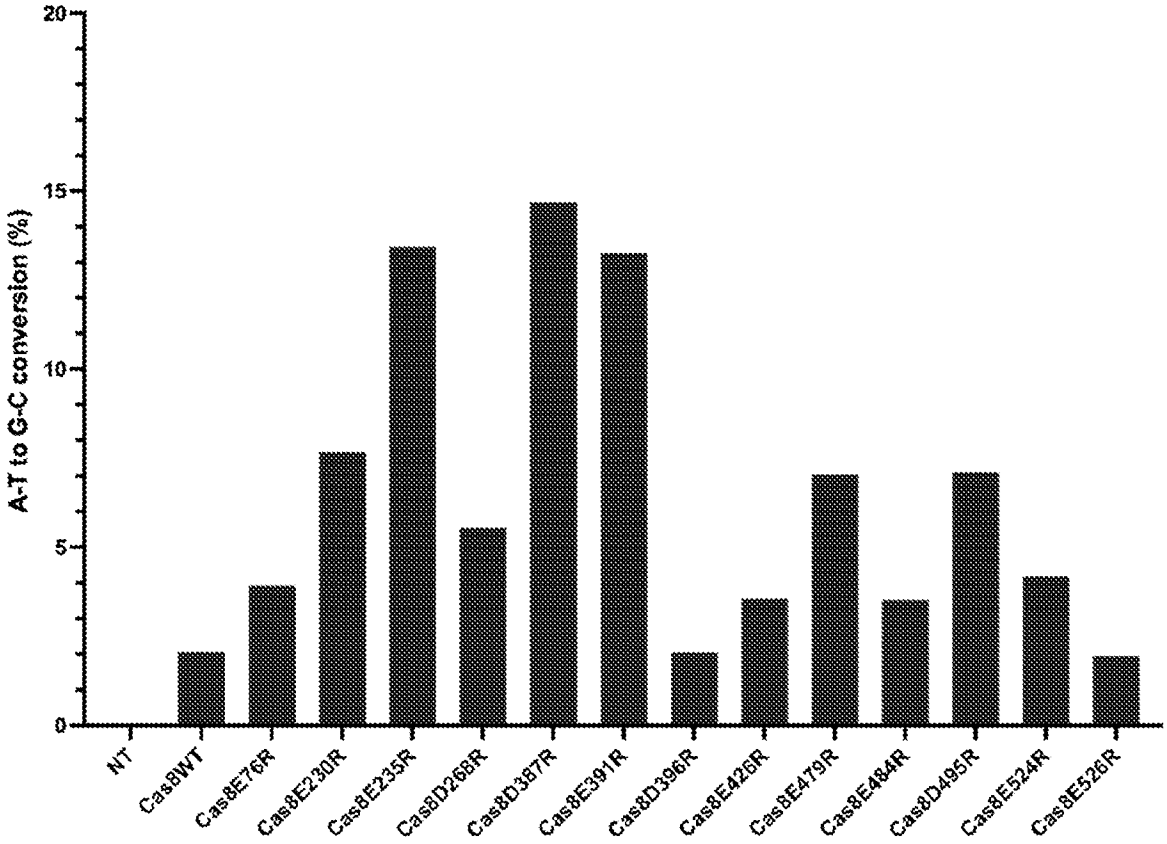


FIG. 13A

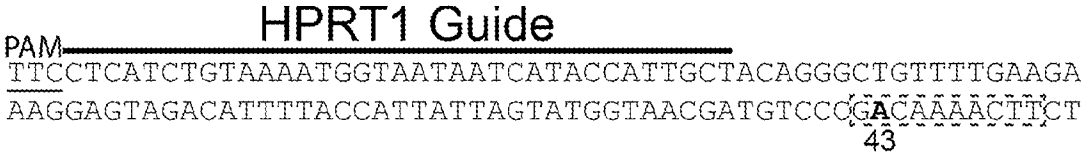


FIG. 13B

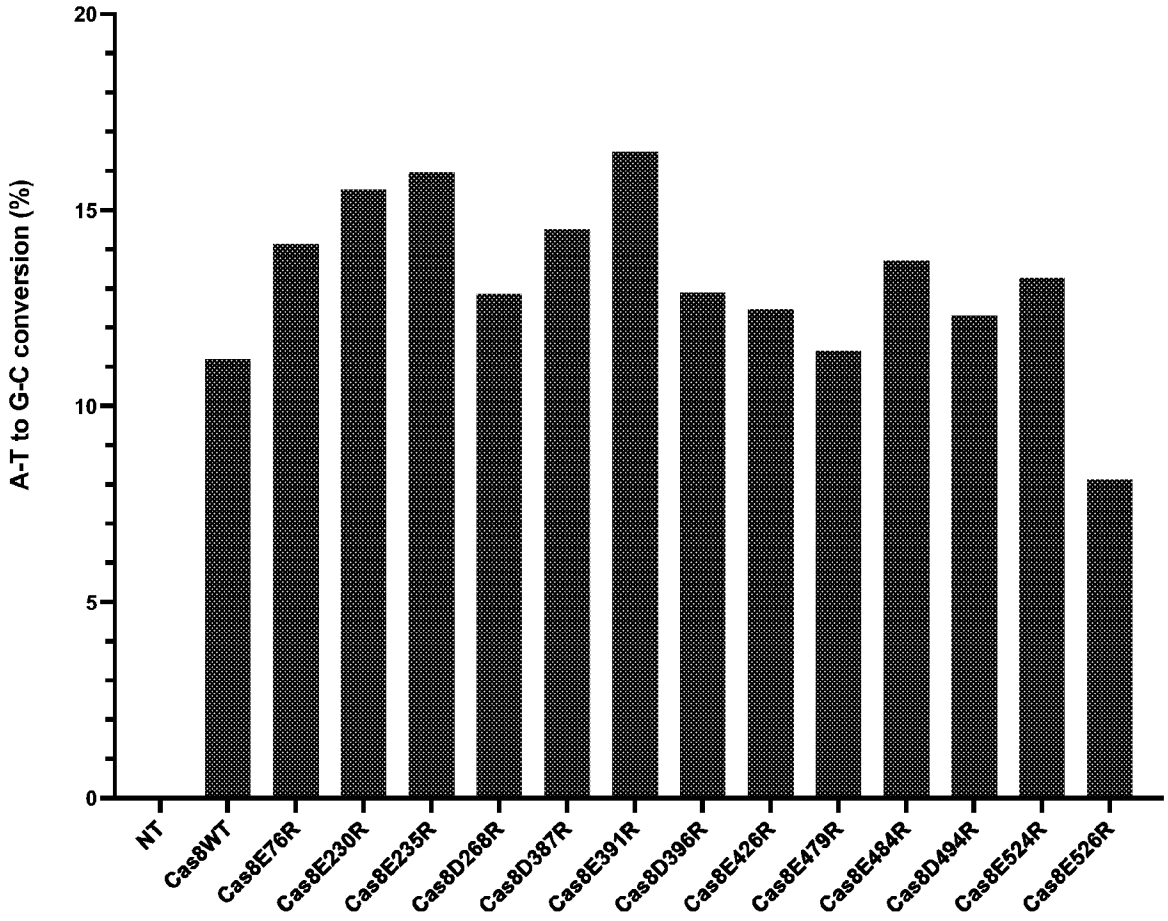


FIG. 14

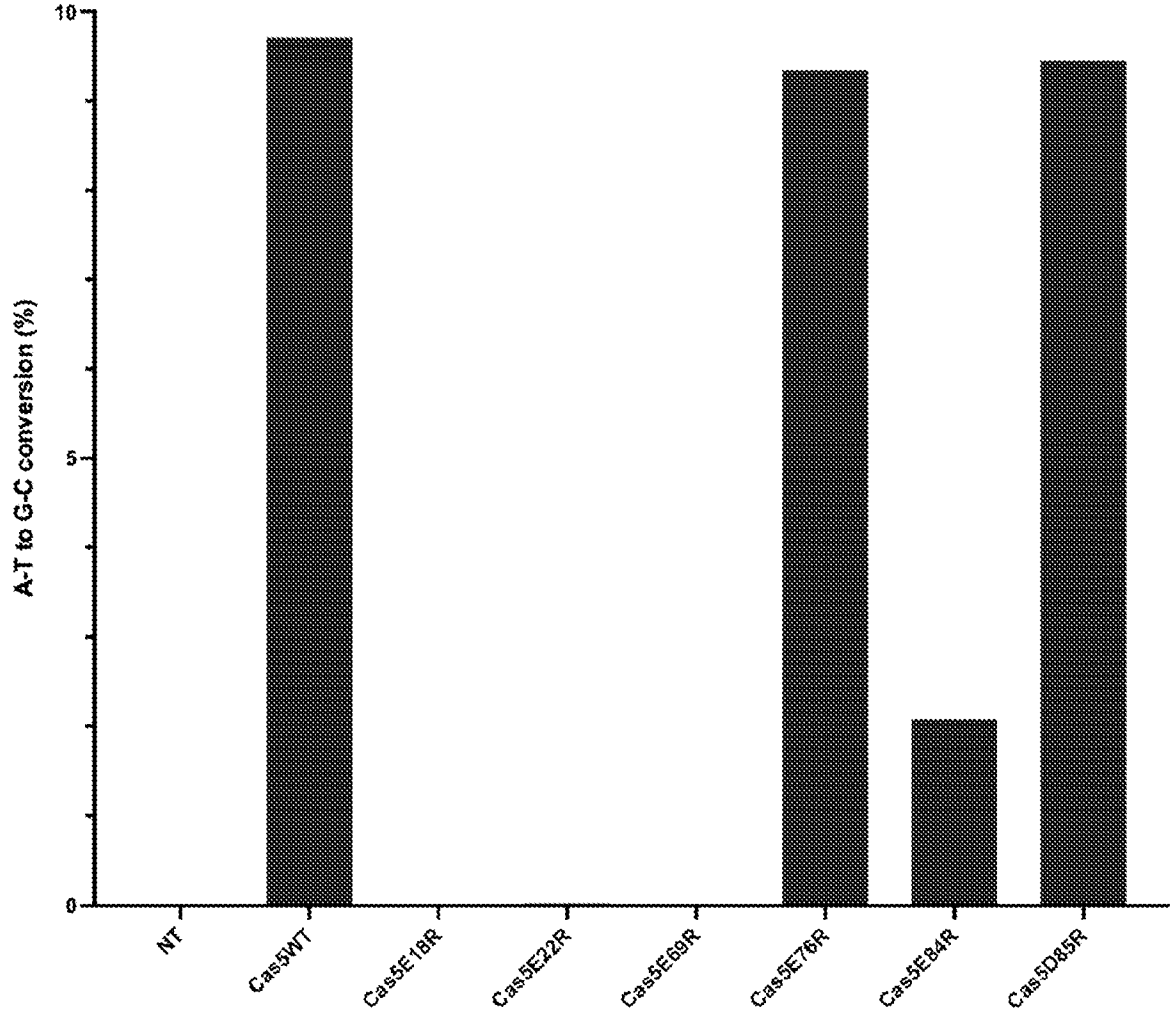

A-T to G-C conversion (%)
10
5
0
NT
Cas5WT
Cas5E18R
Cas5E22R
Cas5E69R
Cas5E76R
Cas5E84R
Cas5D85R

FIG. 15

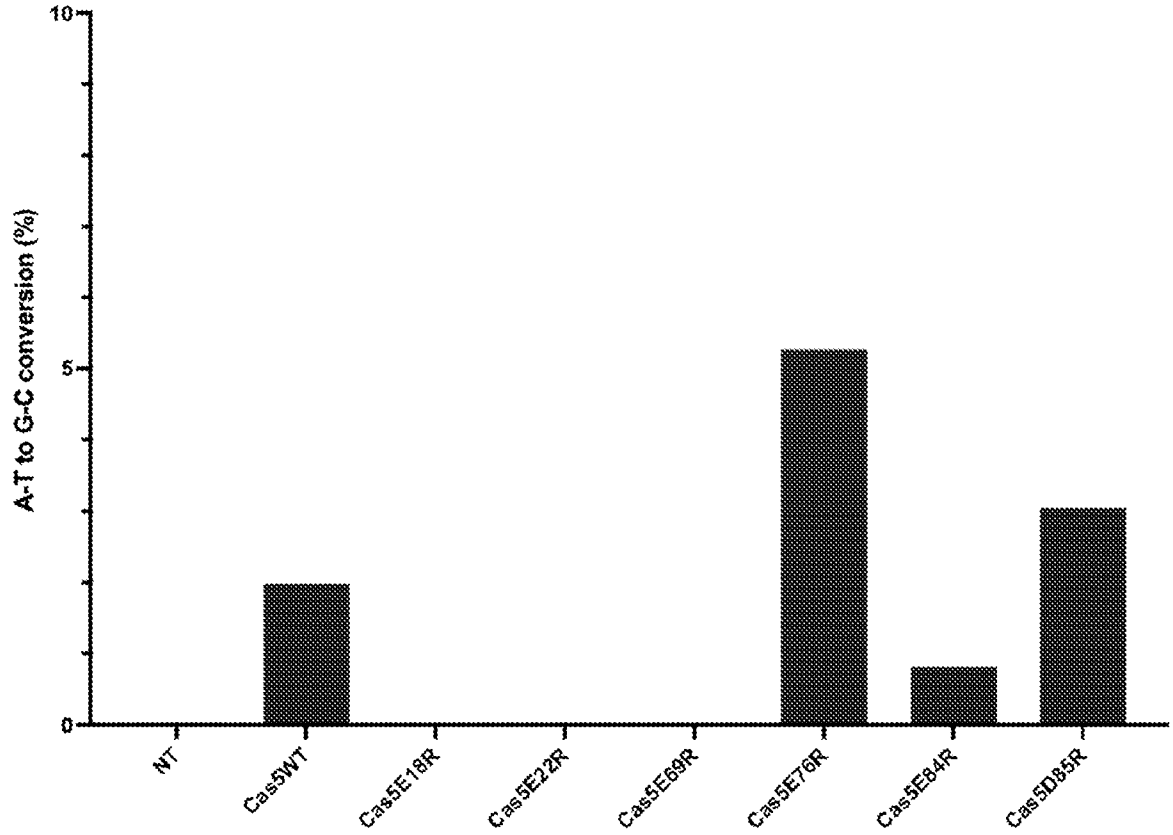

A-T to G-C conversion (%)
10
5
0
NT
Cas5WT
Cas5E18R
Cas5E22R
Cas5E69R
Cas5E76R
Cas5E84R
Cas5D85R

FIG. 16

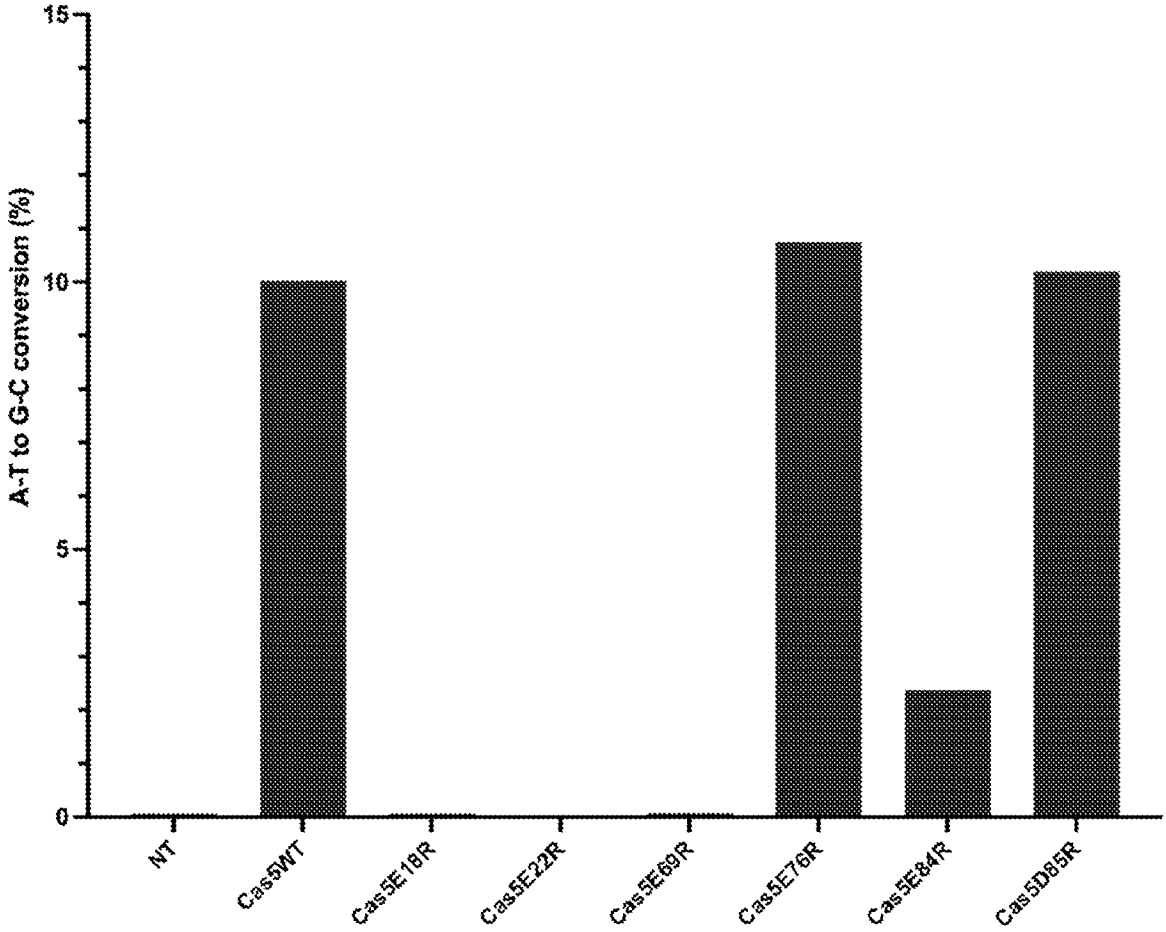

A-T to G-C conversion (%)
15
10
5
0
NT
Cas5WT
Cas5E18R
Cas5E22R
Cas5E69R
Cas5E76R
Cas5E84R
Cas5D85R

FIG. 17

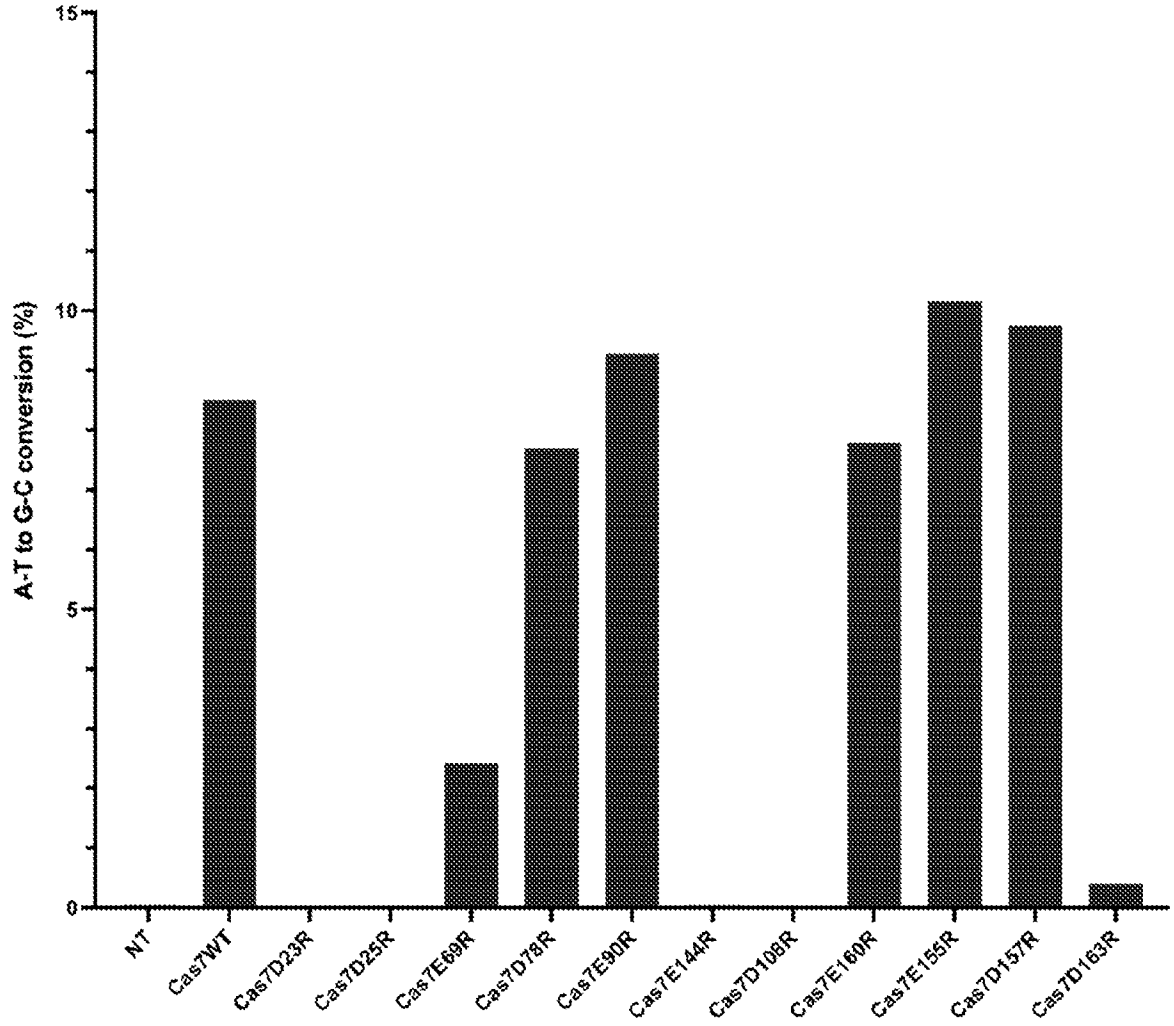

A-T to G-C conversion (%)
15
10
5
0
NT
Cas7WT
Cas7D23R
Cas7D25R
Cas7E69R
Cas7D78R
Cas7E90R
Cas7E144R
Cas7D108R
Cas7E160R
Cas7E155R
Cas7D157R
Cas7D163R

FIG. 18

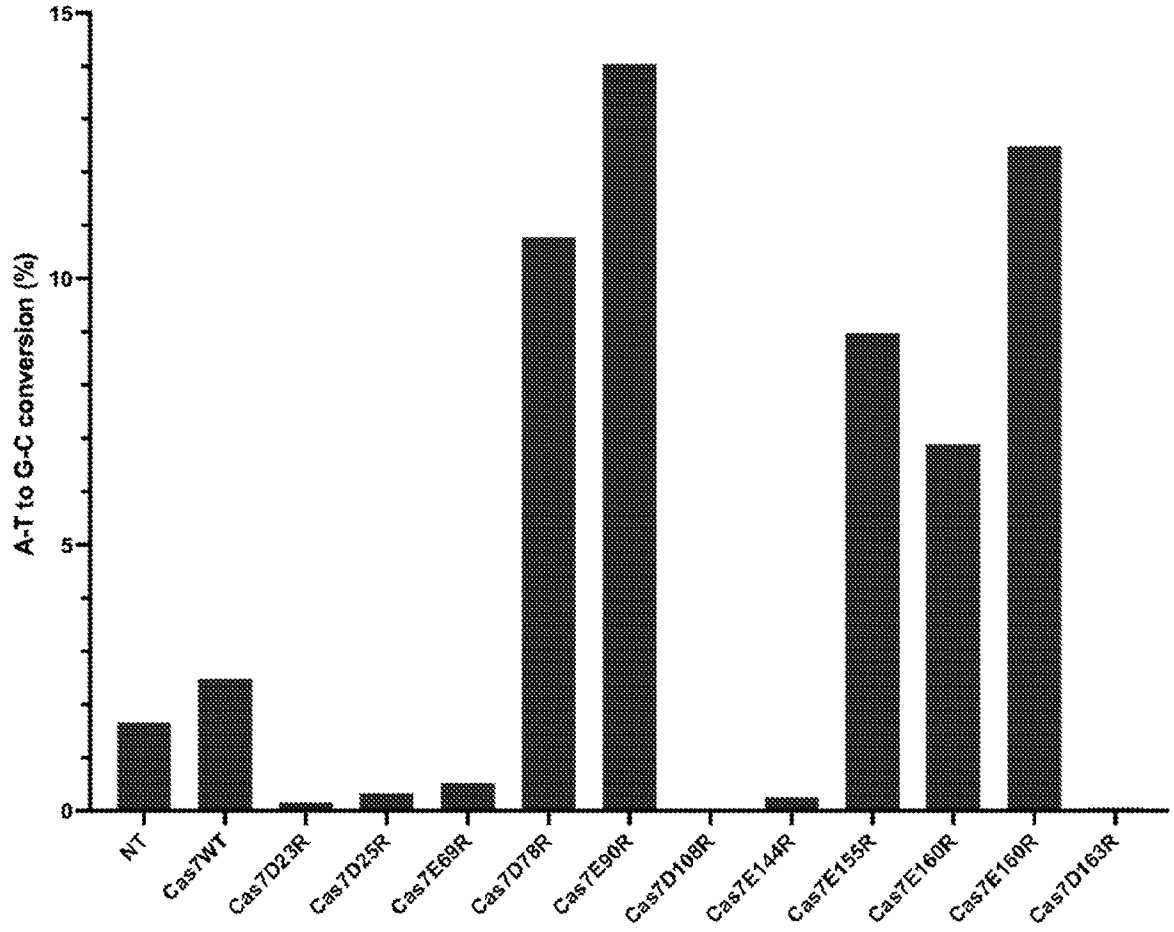

A-T to G-C conversion (%)
15
10
5
0
NT
Cas7WT
Cas7D23R
Cas7D25R
Cas7E69R
Cas7D78R
Cas7E90R
Cas7D108R
Cas7E144R
Cas7E155R
Cas7E160R
Cas7E160R
Cas7D163R

FIG. 19

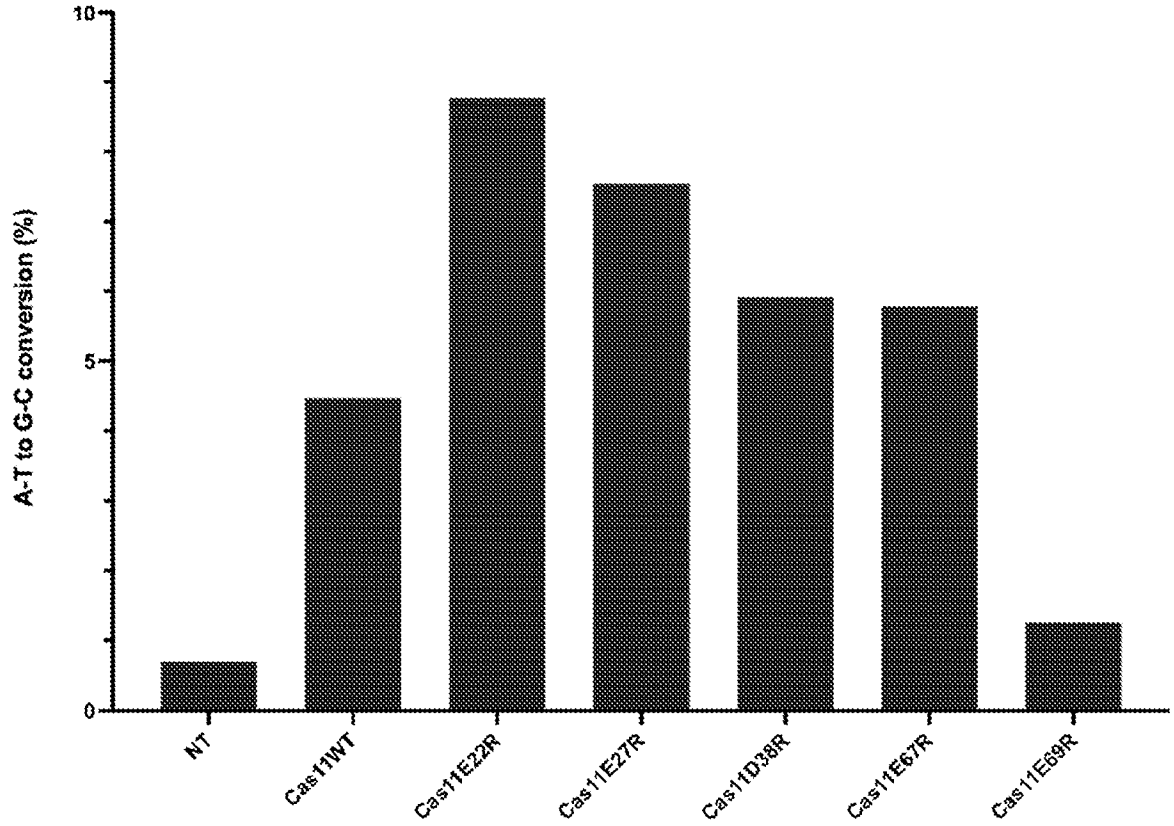

A-T to G-C conversion (%)
10
5
0
NT
Cas11WT
Cas11E22R
Cas11E27R
Cas11D38R
Cas11E67R
Cas11E69R

FIG. 20

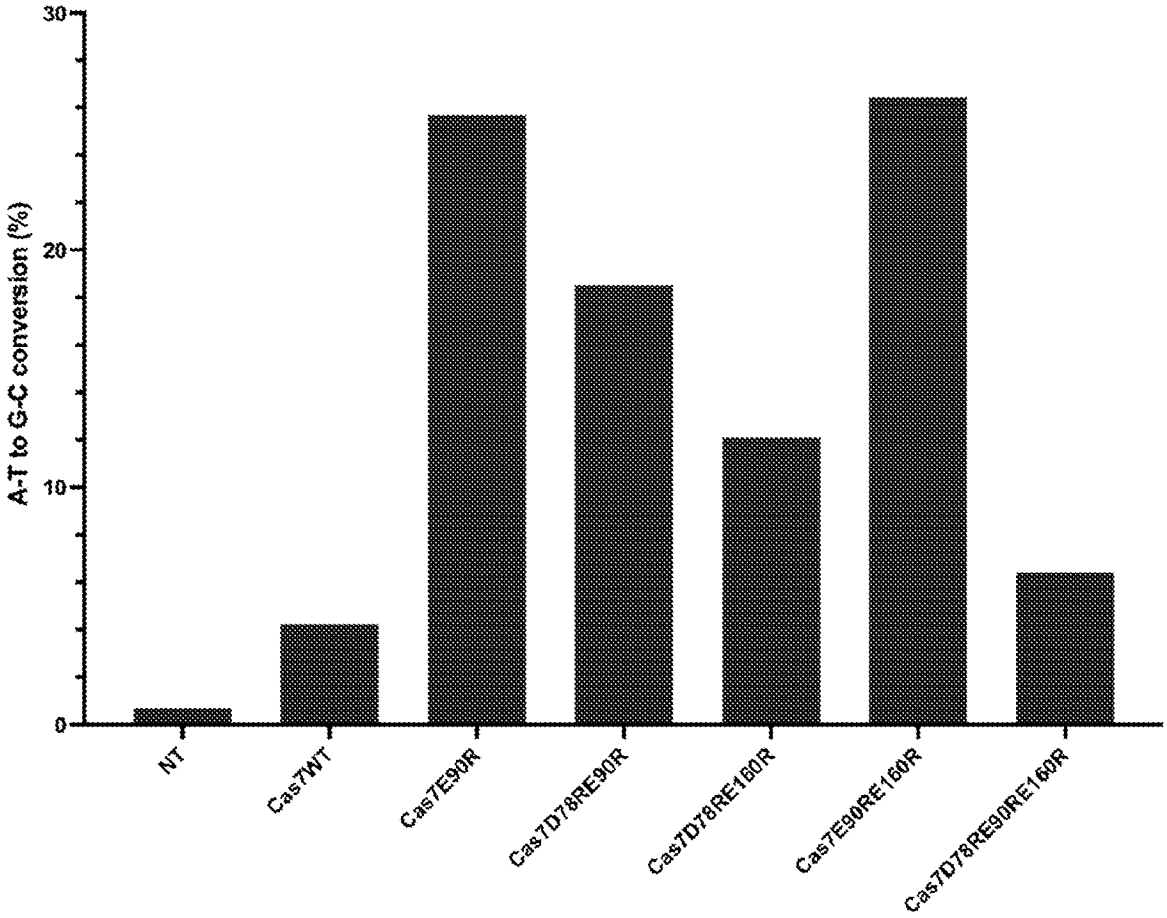

A-T to G-C conversion (%)
30
20
10
0
NT
Cas7WT
Cas7E90R
Cas7D78RE90R
Cas7D78RE160R
Cas7E90RE160R
Cas7D78RE90RE160R

FIG. 21

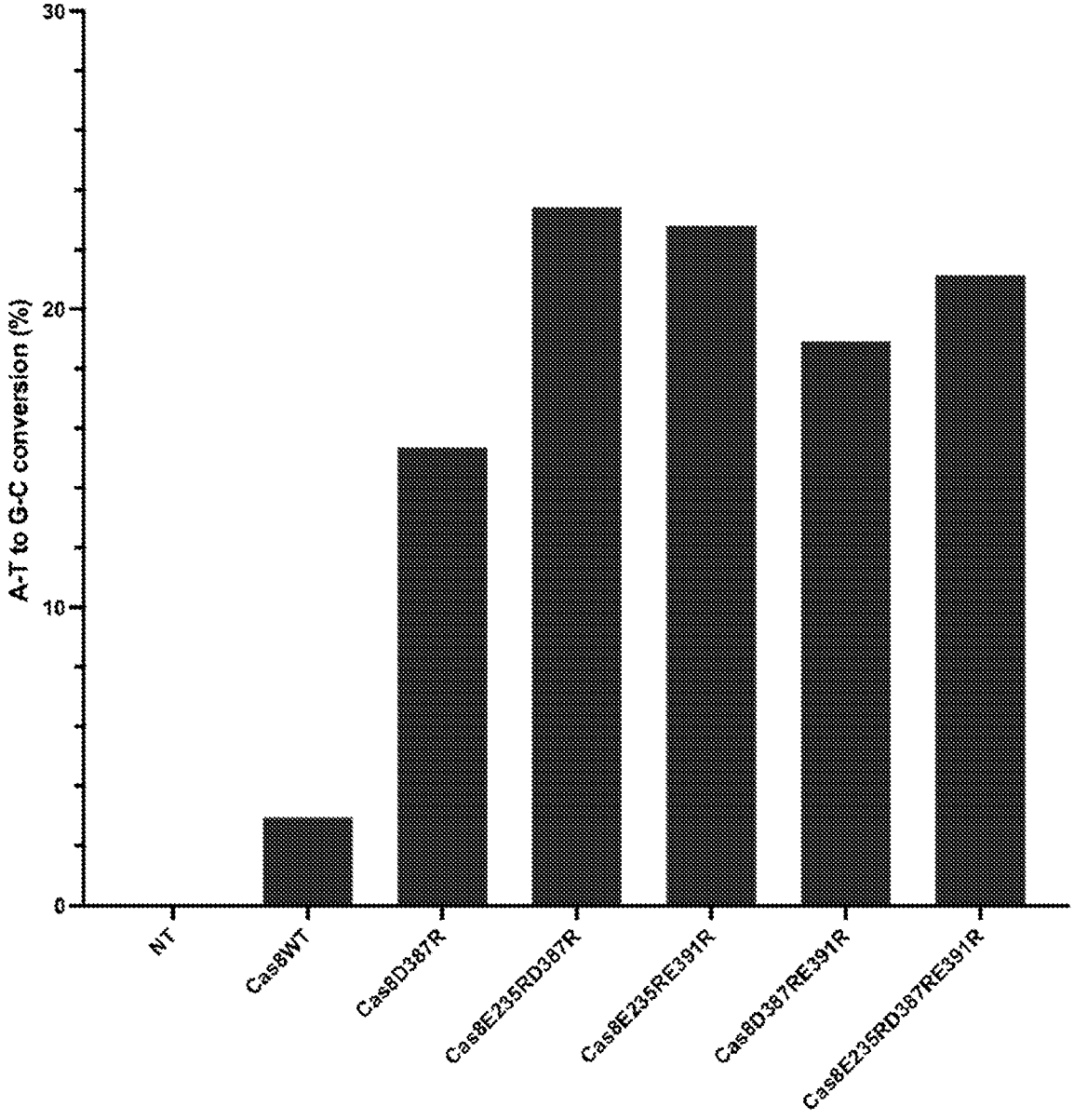

A-T to G-C conversion (%)
30
20
10
0
NT
Cas8WT
Cas8D387R
Cas8E235RD387R
Cas8E235RE391R
Cas8D387RE391R
Cas8E235RD387RE391R

FIG. 22

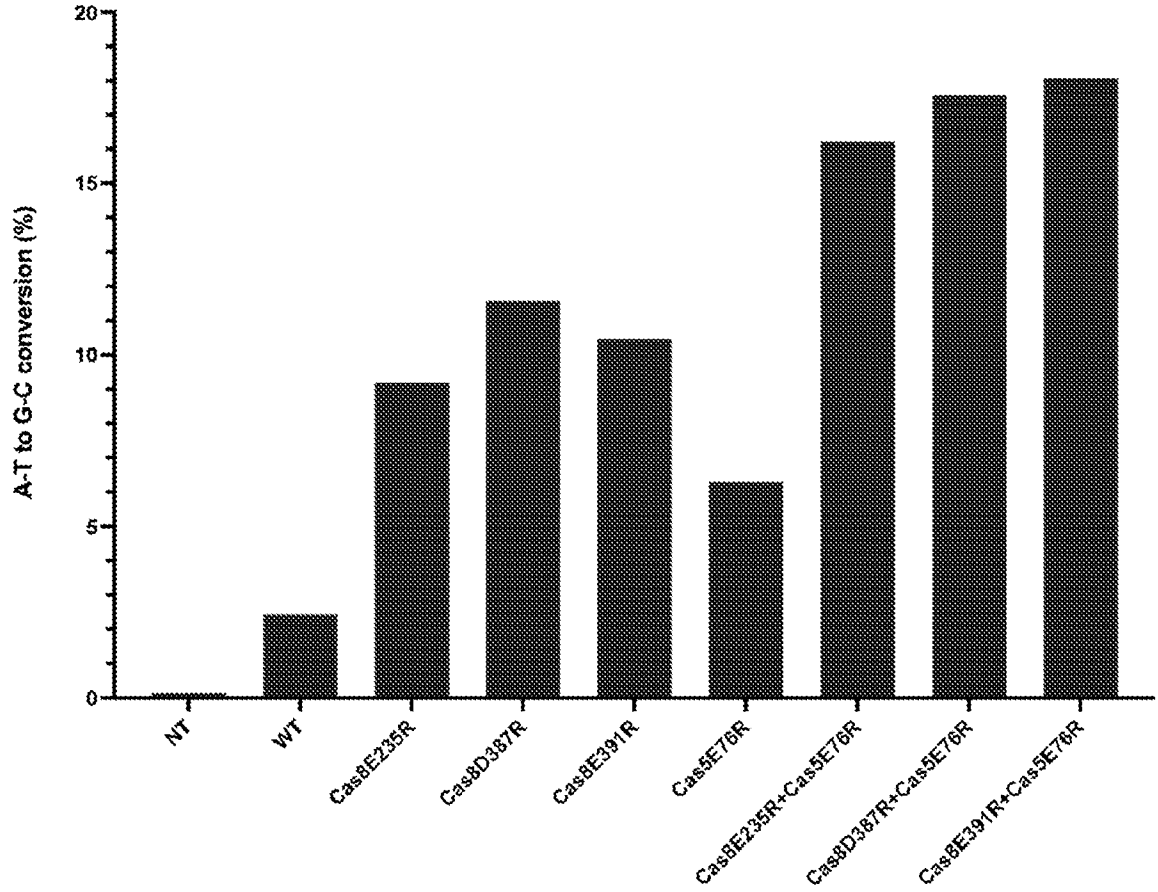

A-T to G-C conversion (%)
20
15
10
5
0
NT
WT
Cas8E235R
Cas8D387R
Cas8E391R
Cas5E76R
Cas8E235R+Cas5E76R
Cas8D387R+Cas5E76R
Cas8E391R+Cas5E76R

FIG. 23

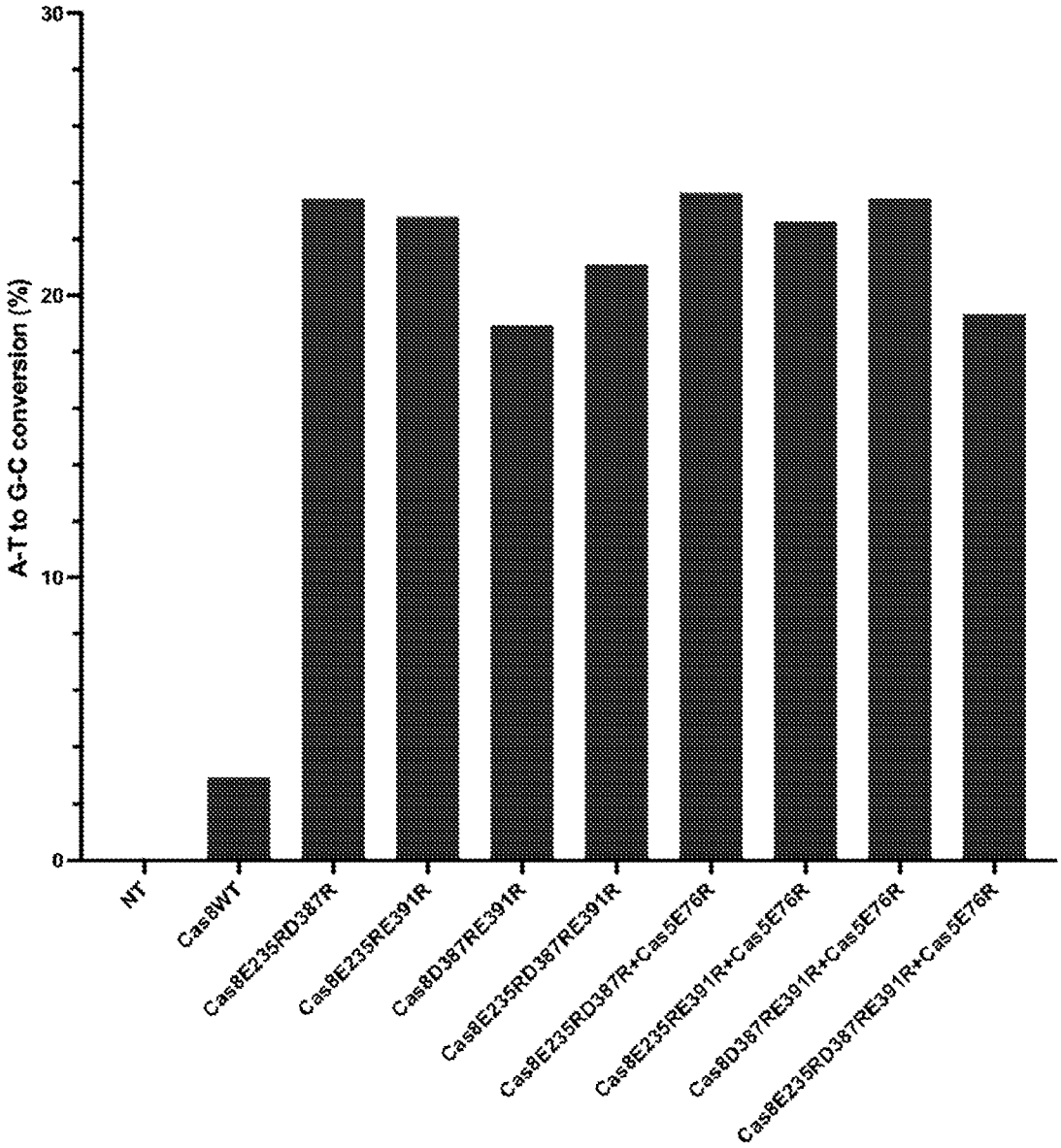

A-T to G-C conversion (%)
30
20
10
0
NT
Cas8WT
Cas8E235RD387R
Cas8E235RE391R
Cas8D387RE391R
Cas8E235RD387RE391R
Cas8E235RD387R+Cas5E76R
Cas8E235RE391R+Cas5E76R
Cas8D387RE391R+Cas5E76R
Cas8E235RD387RE391R+Cas5E76R

FIG. 24

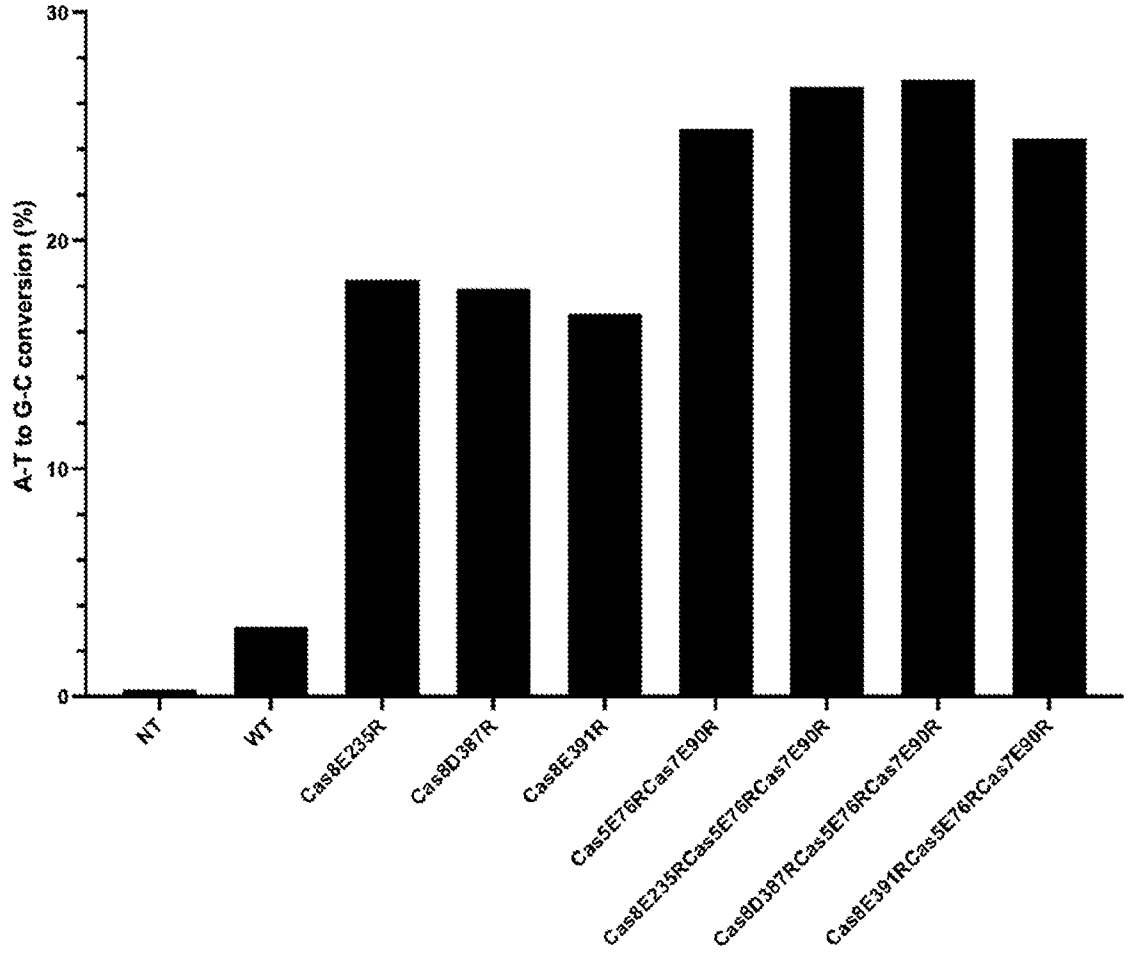

A-T to G-C conversion (%)
30
20
10
0
NT
WT
Cas8E235R
Cas8D387R
Cas8E391R
Cas5E76RCas7E90R
Cas8E235RCas5E76RCas7E90R
Cas8D387RCas5E76RCas7E90R
Cas8E391RCas5E76RCas7E90R

FIG. 25

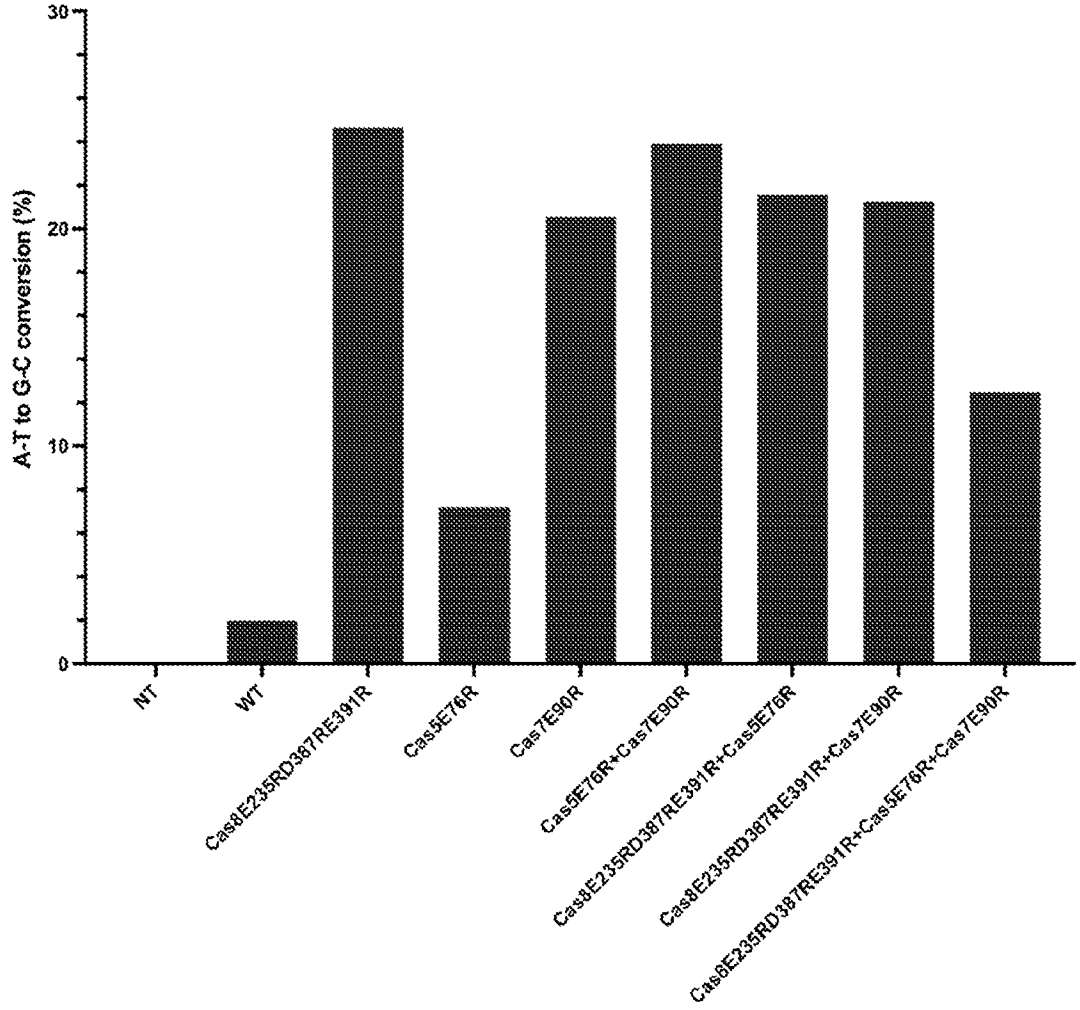

A-T to G-C conversion (%)
30
20
10
0
NT
WT
Cas8E235RD387RE391R
Cas5E76R
Cas7E90R
Cas5E76R+Cas7E90R
Cas8E235RD387RE391R+Cas5E76R
Cas8E235RD387RE391R+Cas7E90R
Cas8E235RD387RE391R+Cas5E76R+Cas7E90R

FIG. 26A

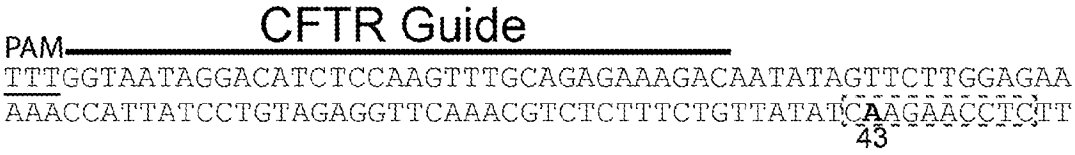


FIG. 26B

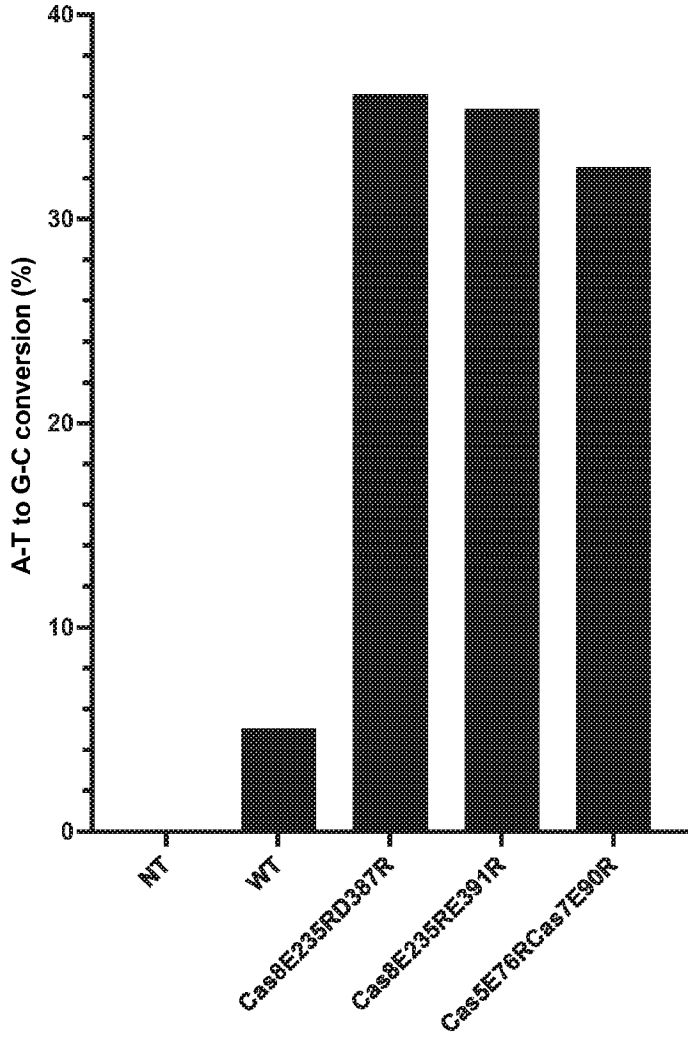


FIG. 27A

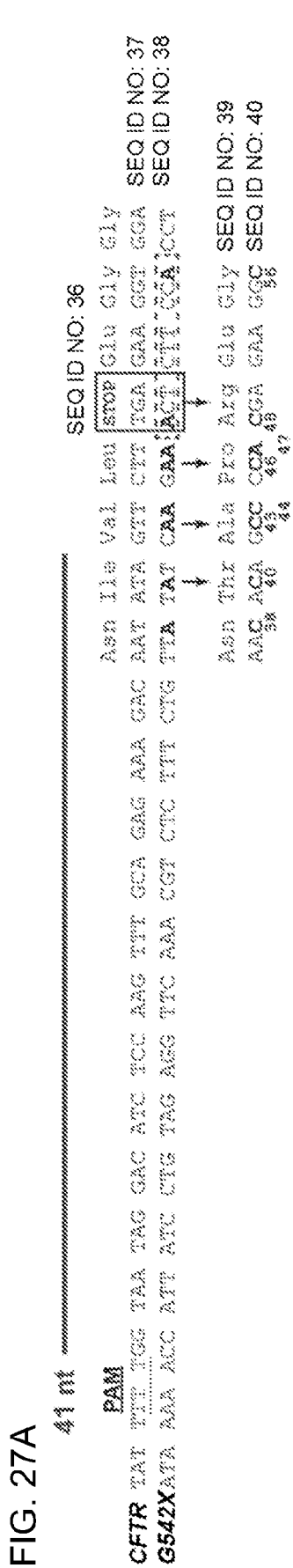

41 nt
PAM
CFTR TAT TTT TGG TAA TAG GAC ATC TCC AAG TTT GCA GAG AAA GAC AAT ATA GTT CTT TGA GAA GGT GGA SEQ ID NO: 37
G542X ATA AAA ACC ATT ATC CTG TAG AGG TTC AAA CGT CTC TTT CTG TTA TAT CAA GAA ACT CTT CCA CCT SEQ ID NO: 38
SEQ ID NO: 36
Asn Ile Val Leu STOP Glu Gly Gly
Asn Thr Ala Pro Arg Glu Gly SEQ ID NO: 39
AAC ACA GCC CCA CGA GAA GGC SEQ ID NO: 40

FIG. 27C

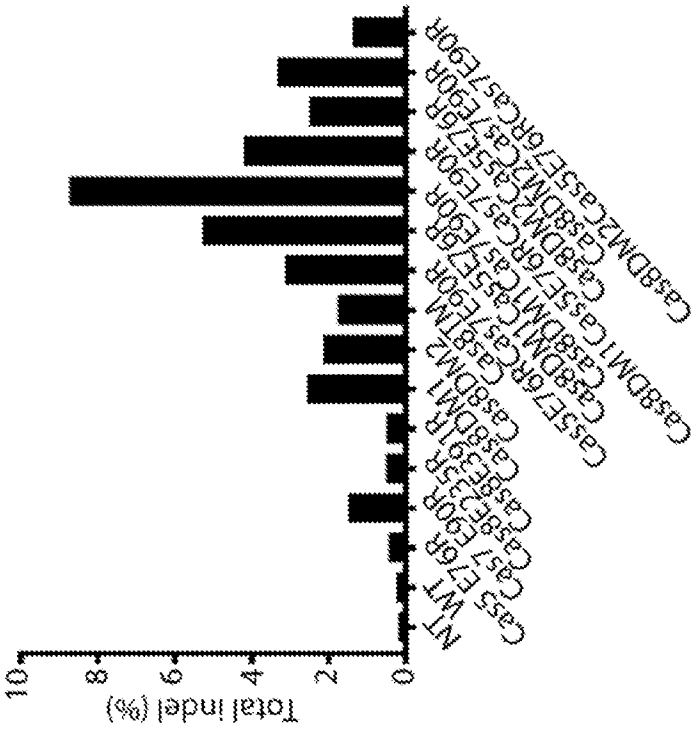
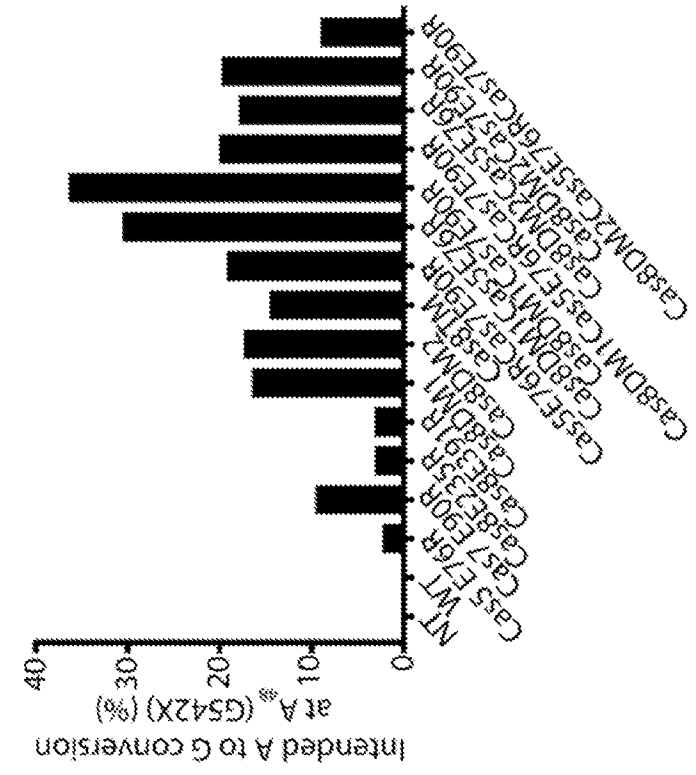


FIG. 27B



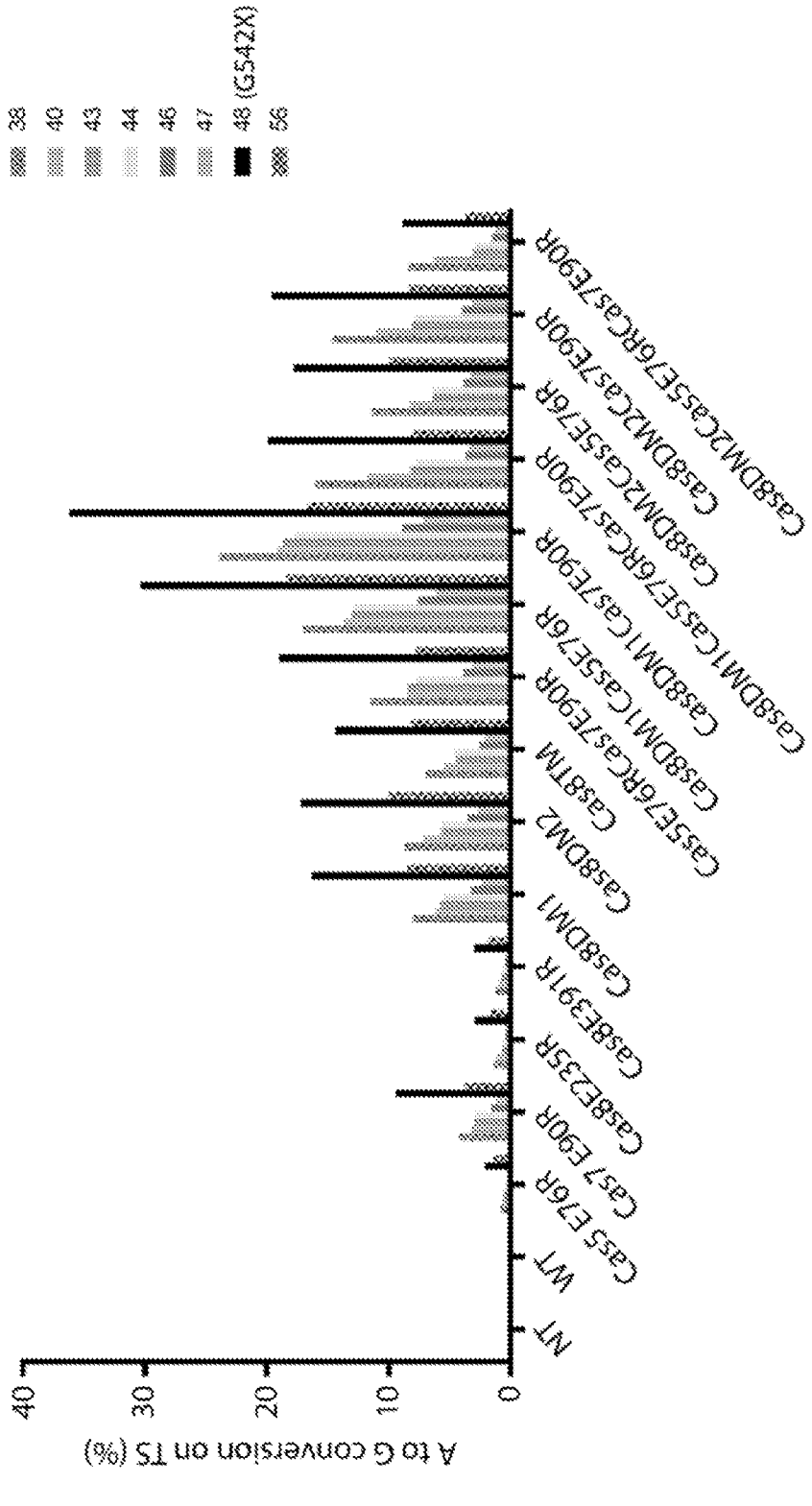
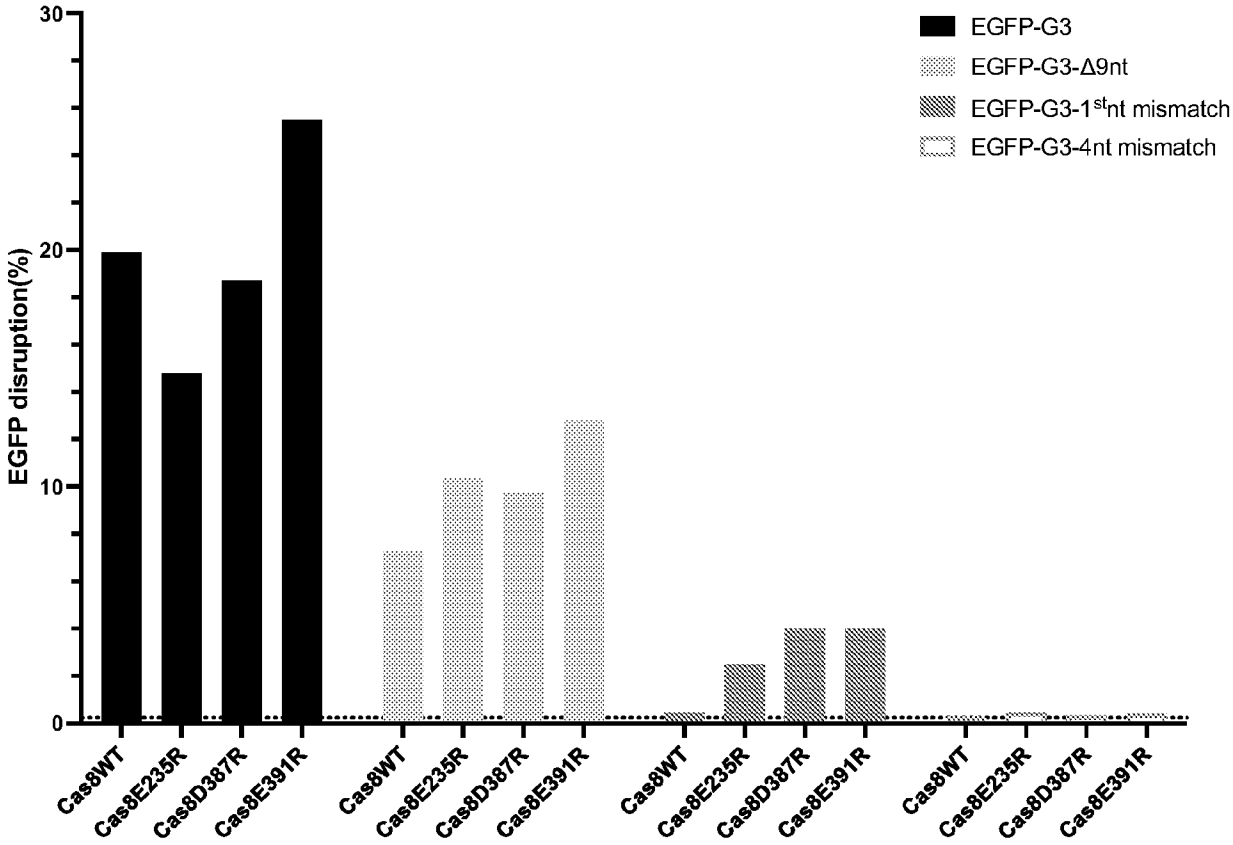

38
40
43
44
46
47
48 (G542X)
56
A to G conversion on TS (%)
0
10
20
30
40
NT
WT
Cas5 E76R
Cas7 E90R
Cas8E235R
Cas8E391R
Cas8DM1
Cas8DM2
Cas8TM
Cas5E76RCas7E90R
Cas8DM1Cas5E76R
Cas8DM1Cas7E90R
Cas8DM1Cas5E76RCas7E90R
Cas8DM2Cas5E76R
Cas8DM2Cas7E90R
Cas8DM2Cas5E76RCas7E90R

FIG. 27D

FIG. 28A

GTGACCGCCGCCGGGATCACTCTCGGCATGGACGA	EGFP-G3
GTGACCGCCGCCGGGATCACTCTCGG	EGFP-G3-Δ9nt
CTGACCGCCGCCGGGATCACTCTCGGCATGGACGA	EGFP-G3-1stnt mismatch
GTGACCGCCGCCCCCTTCACTCTCGGCATGGACGA	EGFP-G3-4nt mismatch

FIG. 28B



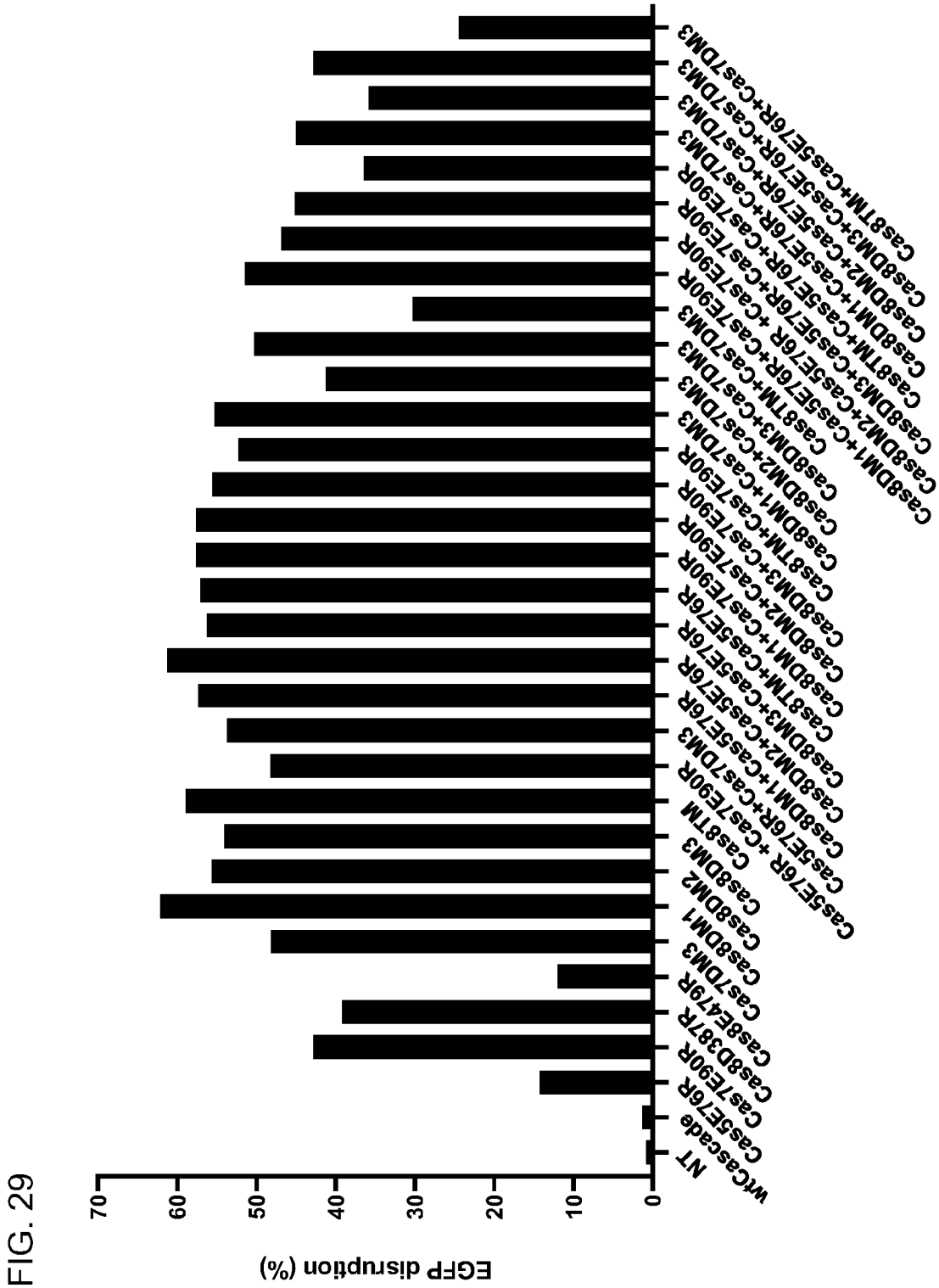

EGFP disruption (%)
0
10
20
30
40
50
60
70
NT
wtCascade
Cas5E7 6R
Cas7E90R
Cas8D387R
Cas8E479R
Cas7DM3
Cas8DM1
Cas8DM2
Cas8DM3
Cas8TM
Cas5E7 6R +Cas7E90R
Cas5E7 6R+Cas7DM3
Cas8DM1+Cas5E7 6R
Cas8DM2+Cas5E7 6R
Cas8DM3+Cas5E7 6R
Cas8TM+Cas5E7 6R
Cas8DM1+Cas7E90R
Cas8DM2+Cas7E90R
Cas8DM3+Cas7E90R
Cas8TM+Cas7E90R
Cas8DM1+Cas7DM3
Cas8DM2+Cas7DM3
Cas8DM3+Cas7DM3
Cas8TM+Cas7DM3
Cas8DM1+Cas5E7 6R +Cas7E90R
Cas8DM2+Cas5E7 6R +Cas7E90R
Cas8DM3+Cas5E7 6R +Cas7E90R
Cas8TM+Cas5E7 6R +Cas7E90R
Cas8DM1+Cas5E7 6R+Cas7DM3
Cas8DM2+Cas5E7 6R+Cas7DM3
Cas8DM3+Cas5E7 6R+Cas7DM3
Cas8TM+Cas5E7 6R+Cas7DM3

FIG. 29

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/035845

A. CLASSIFICATION OF SUBJECT MATTER

IPC: ***C12N 9/22*** (2024.01); ***C07K 14/22*** (2024.01); ***C12N 15/113*** (2024.01); ***C12N 15/90*** (2024.01); *C12N 15/11* (2024.01)

CPC: ***C12N 9/22; C12N 15/113; C12N 15/90; C07K 14/22;*** *C12N 2310/20; C12N 2800/80; C12N 15/11*

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2021/247301 A1 (THE REGENTS OF THE UNIVERSITY OF MICHIGAN et al.) 09 December 2021 (09.12.2021) entire document	1-3
A	US 11,439,712 B2 (NORTH CAROLINA STATE UNIVERSITY) 13 September 2022 (13.09.2022) entire document	1-3

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
22 August 2024 (22.08.2024)	**03 September 2024 (03.09.2024)**

Name and mailing address of the ISA/US	Authorized officer
Mail Stop PCT, Attn: ISA/US **Commissioner for Patents** **P.O. Box 1450, Alexandria, VA 22313-1450**	**MATOS** **TAINA**
Facsimile No. **571-273-8300**	Telephone No. **571-272-4300**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/035845

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
   a. ☑ forming part of the international application as filed.
   b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)),
      ☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/035845

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☑ Claims Nos.: **4-43**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).