



(51) International Patent Classification:

CI2N 9/22 (2006.01)

CI2N 15/10 (2006.01)

CI2N 9/50 (2006.01)

(21) International Application Number:

PCT/US2021/040160

(22) International Filing Date:

01 July 2021 (01.07.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/073,898

02 September 2020 (02.09.2020) US

63/208,606

09 June 2021 (09.06.2021) US

(71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US).

(72) Inventors: ABUDAYYEH, Omar; 49 Bristol Street, Cambridge, MA 02141 (US). GOOTENBERG, Jonathan; 300 Third Street, Apt. 510, Cambridge, MA 02142 (US).

(74) Agent: LABEOTS, Laura, A. et al.; Lathrop GPM LLP, 155 North Wacker Drive, Suite 300, Chicago, IL 60606-1787 (US).

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

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: SYSTEMS, METHODS, AND COMPOSITIONS FOR RNA-GUIDED RNA-TARGETING CRISPR EFFECTORS

(57) Abstract: This disclosure provides systems, methods, and compositions for RNA-guided RNA-targeting CRISPR effectors for the treatment of diseases as well as diagnostics. In some embodiments, nucleotide deaminase functionalized CRISPR systems for RNA editing RNA knockdown, viral resistance, splicing modulation, RNA tracking, translation modulation, and epi-transcriptomic modifications are disclosed.



SYSTEMS, METHODS, AND COMPOSITIONS FOR RNA-GUIDED RNA-TARGETING CRISPR EFFECTORS

RELATED APPLICATIONS

- 5 [0001] This application claims priority to U.S. Provisional Application Nos. 63/073,898 filed September 2, 2020 and 63/208,606, filed June 9, 2021, the entire contents of which are hereby incorporated by reference in their entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR 10 DEVELOPMENT

[0002] This invention was made with government support under Grant No. R21 AI149694 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

TECHNOLOGY FIELD

- 15 [0003] The subject matter disclosed herein is generally related to systems, methods, and compositions for RNA-guided RNA-targeting CRISPR effectors for the treatment of diseases and diagnostics.

BACKGROUND

- 20 [0004] RNA targeting tools for genetic engineering are important to study RNA biology and develop RNA therapeutics. These tools can regulate intracellular and intercellular target-gene functions and expressions and manipulate specific target-genomic information. In contrast to DNA targeting tools, only a few RNA targeting tools for CRISPR applications have been developed, and these RNA specific tools present challenges. For instance, protein complexes for CRISPR RNA editing can have a weak activity in mammalian cells and present collateral
25 effects which can be toxic in some cell types. Also, the size of the RNA editing tools can be a significant barrier to their use, as technologies such as programmable activation cannot be delivered in mouse models using common methods such as adeno-associated vectors (AAV). Development of effective gene and cell therapies requires genome editing tools and delivery technologies that can meet the demands for cell type specificity, large payload sizes, and
30 efficient integration of diverse and large sequences. Many applications, including single-vector homology directed repair (HDR), CRISPR base editing, gene activation, and large gene delivery, are limited by AAV size restrictions. Thus, there remains a need for more effective tools for gene correction and delivery.

SUMMARY

[0005] The present invention is directed to systems, methods, and compositions for RNA-guided RNA-targeting CRISPR effectors for the treatment of diseases and diagnostics. It also pertains to nucleotide deaminase functionalized CRISPR systems for RNA editing RNA knockdown, viral resistance, splicing modulation, RNA tracking, translation modulation, and epi-transcriptomic modifications.

[0006] In one aspect, a composition that cleaves an RNA target is discussed, which comprises a guide RNA that specifically hybridizes to the RNA target, and a polypeptide comprising an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-32 and 632. The amino acid sequence can be SEQ ID NO: 1. The polypeptide can comprise a glutamate at a position corresponding to position 70 of SEQ ID NO:1, an aspartate at a position corresponding to position 429 of SEQ ID NO:1, an aspartate at a position corresponding to position 487 of SEQ ID NO:1, and/or an aspartate at a position corresponding to position 654 of SEQ ID NO:1. If desired, the polypeptide can comprise an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO:1. If desired, the polypeptide can comprise an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO:1. If desired, the polypeptide can comprise an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO:1. If desired, the polypeptide can be the amino acid sequence of SEQ ID NO:1.

[0007] Yet in another aspect, the guide RNA comprises a mismatch distance that is 20-65% of the length of the guide. If desired, the guide RNA can comprise a mismatch that is about 20 to about 30 nucleotides from the non-pairing C of the guide RNA. If desired, the guide RNA can have a sequence with a length of from about 20 to about 53 nucleotides (nt), preferably from about 25 to about 53 nt, more preferably from about 29 to about 53 nt or from about 40 to about 50 nt. If desired, the guide RNA can be a pre-crRNA. If desired, the guide RNA can be a mature crRNA.

[0008] Yet in another aspect, the RNA target can be a single-strand RNA (ssRNA). If desired, the RNA target can be in a cell. The cell can be a prokaryotic cell. The cell can be a eukaryotic cell. The eukaryotic cell can be a mammalian cell. The mammalian cell can be a human cell.

[0009] Yet in another aspect, the polypeptide can comprise a deaminase domain. The deaminase can be selected from the group consisting of an adenosine deaminase, a cytidine deaminase, and a catalytic domain thereof. If desired, the deaminase can be an adenosine

deaminase. If desired, the deaminase domain can comprise an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 25-31. If desired, the deaminase domain can comprise an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 32-42.

5 [0010] Yet in another aspect, a nucleic acid molecule encoding the polypeptide is discussed. The nucleic acid molecule can encode the guide RNA.

[0011] Yet in another aspect, a vector comprising the nucleic acid molecule is discussed.

[0012] Yet in another aspect, a cell comprising the composition, the nucleic acid molecule or the vector is discussed. If desired, the cell can be a prokaryotic cell. If desired, the cell can be
10 a eukaryotic cell. The eukaryotic cell can be a mammalian cell. The mammalian cell can be a human cell.

[0013] Yet in another aspect, a method of cleaving an RNA target in a cell is discussed, which comprises providing to the cell the composition, the nucleic acid molecule or the vector. If desired, the RNA target can be an ssRNA.

15 [0014] Yet in another aspect, a method of stabilizing an RNA target in a cell is discussed, which comprises providing to the cell the composition, the nucleic acid molecule or the vector. If desired, the RNA target can be an ssRNA.

[0015] Yet in another aspect, a method of affecting translation of an RNA target in a cell is discussed, which comprises providing to the cell the composition, the nucleic acid molecule
20 or the vector. If desired, the RNA target can be an ssRNA.

[0016] Yet in another aspect, a method of treating a genetically inherited disease in a subject in need thereof is discussed, which comprises administering to the subject an effective amount of the composition, the nucleic acid molecule or the vector, wherein the genetically inherited disease involves a guanosine to adenosine change in the genome of the subject. The
25 genetically inherited disease can be selected from the group consisting of Meier-Gorlin syndrome; Seckel syndrome 4; Joubert syndrome 5; Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; leukoencephalopathy; Usher syndrome, type 2C; spinocerebellar ataxia 28; glycogen storage disease type III; primary hyperoxaluria, type I; long QT syndrome 2; Sjögren-Larsson syndrome; hereditary fructosuria; neuroblastoma;
30 amyotrophic lateral sclerosis type 9; Kallmann syndrome 1; limb-girdle muscular dystrophy, type 2L; familial adenomatous polyposis 1; familial type 3 hyperlipoproteinemia; Alzheimer's disease, type 1; metachromatic leukodystrophy; and cancer.

[0017] Yet in another aspect, a method of treating a genetically inherited disease in a subject in need thereof is discussed, which comprises administering to the subject an effective amount of the composition, the nucleic acid molecule or the vector, wherein the genetically inherited disease is a pre-termination disease.

5 [0018] Yet in another aspect, a method of altering splicing of a pre-mRNA in a cell is discussed, which comprises administering to the cell an effective amount of the composition, the nucleic acid molecule or the vector.

[0019] Yet in another aspect, a method of changing microRNA targets in a subject in need is discussed, which comprises administering to the subject an effective amount of the
10 composition, the nucleic acid molecule or the vector.

[0020] Yet in another aspect, a method of increasing RNA stability in a cell is discussed, which comprises administering to the cell an effective amount of the composition, the nucleic acid molecule or the vector.

[0021] Yet in another aspect, a method of modulating translation in a cell is discussed, which
15 comprises administering to the cell an effective amount of the composition, the nucleic acid molecule or the vector.

[0022] Yet in another aspect, a method of detecting a bacterium or derivative thereof in a sample is discussed, the method comprises adding to the sample an effective amount of the composition, the nucleic acid molecule or the vector, and detecting a reporter specific to the
20 bacterium or derivative thereof.

[0023] Yet in another aspect, a method of detecting a virus or derivative thereof in a sample is discussed, the method comprises adding to the sample an effective amount of the composition, the nucleic acid molecule or the vector, and detecting a reporter specific to the virus or derivative thereof.

25 [0024] These and other aspects of the applicants' teaching are set forth herein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] Aspects, features, benefits and advantages of the embodiments described herein will be apparent with regard to the following description, appended claims, and accompanying
30 drawings where:

[0026] **FIG. 1A** is a schematic representation of RNA editing via dDisCas7-11a-NES-ADAR2 fusions that are programmed to target a specific adenosine via a cytidine mismatch for A to I editing in the guide:target RNA duplex according to embodiments of the present teachings;

[0027] FIG. 1B is a schematic of Cas7-11 expression, processing, and interference against ssRNA viruses and other ssRNA targets according to embodiments of the present teachings;

[0028] FIG. 1C illustrates the comparison of Cas7-11 features to characteristics of Cas13 and type III-A/B/C systems according to embodiments of the present teachings;

5 [0029] FIG. 1D illustrates hypothetical scenario for the evolution of a single-protein CRISPR-Cas effector from a multi-subunit effector complex according to embodiments of the present teachings;

[0030] FIG. 1E shows type III-E family members and type III-D2 loci architecture and multiple alignment of representative orthologs, wherein CRISPR array symbols are not

10 representative of the number of spacers in the array according to embodiments of the present teachings;

[0031] FIG. 1F illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (D177) according to embodiments of the present teachings;

15 [0032] FIG. 1G illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (D429) according to embodiments of the present teachings;

[0033] FIG. 1H illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (D654) according to embodiments of the

20 present teachings;

[0034] FIG. 1I illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (D745, D758) according to embodiments of the present teachings;

[0035] FIG. 1J illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (E959) according to embodiments of the

25 [0036] FIG. 1K illustrates multiple alignments of representative Cas7-11 orthologs showing conservation of the residues involved in catalysis (E998) according to embodiments of the present teachings;

30 [0037] FIG. 2A illustrates tree of type III-D2 and III-E effectors based on Cas7-11 and Cas7x3 multiple alignment via a computational discovery pipeline according to embodiments of the present teachings;

[0038] FIG. 2B is a schematic of a typical type III-E Cas7-11 locus and type III-D2 Cas7x3 locus wherein the inset shows the multiple Cas7-like domains of Cas7-11 including a putative small subunit Cas11 domain and an uncharacterized insertion in the Cas7-like domain 5 according to embodiments of the present teachings;

5 [0039] FIG. 2C illustrates a native expression of a CRISPR array associated with DisCas7-11 according to embodiments of the present teachings;

[0040] FIG. 2D is a schematic representation of a DR secondary structure and a scissor denoting the cleavage site for processing and maturation according to embodiments of the present teachings;

10 [0041] FIG. 3A illustrates the CRISPR array screen of RNA-guided RNA-targeting Cas7-11 for all crRNAs targeting the MS2 phage genome according to embodiments of the present teachings;

[0042] FIG. 3B is a diagram showing results of the MS2 interference screen (for the MAS targeting spacers) according to embodiments of the present teachings;

15 [0043] FIG. 3C is a diagram showing results of the MS2 interference screen (for the non-targeting spacers) according to embodiments of the present teachings;

[0044] FIG. 3D are images of plaque dilution assays for MS2 targeting spacer and non-targeting spacer according to embodiments of the present teachings;

20 [0045] FIG. 3E is a diagram of the number of DisCas7-11a spacers that display survival enrichment over a threshold of 1.7 across different phage dilution conditions according to embodiments of the present teachings;

[0046] FIG. 3F is a diagram showing Quantification of resistance conferred by top MS2-targeting DisCas7-11 spacers compared against a panel of 4 non-targeting spacers according to embodiments of the present teachings;

25 [0047] FIG. 3G is a diagram showing the quantification of resistance conferred by two MS2-targeting DisCas7-11 spacers compared against a panel of 6 non-targeting spacers according to embodiments of the present teachings;

30 [0048] FIG. 3H illustrates the RFP knockdown assay containing a DisCas7-11 and spacers expression vector co-transformed with an RFP-expressing vector according to embodiments of the present teachings;

[0049] FIG. 3I is a diagram of the DisCas7-11-mediate RFP knockdown by multiple targeting spacers normalized to a non-targeting condition as determined by flow cytometry according to embodiments of the present teachings;

[0050] FIG. 3J is a schematic of the DisCas7-11 a full locus with TPR-CHAT nearby and DisCas7-11 a effector expressed alone with TPR-CHAT according to embodiments of the present teachings;

[0051] FIG. 3K is a diagram of the MS2 interference activity of DisCas7-11a in the full locus and as a single effector with and without Csx29 expression according to embodiments of the present teachings;

[0052] FIG. 4A illustrates a CRISPR array processing assay for DisCas7-11 according to embodiments of the present teachings;

[0053] FIG. 4B illustrates the *in vitro* cleavage of ssRNA 1 target according to embodiments of the present teachings;

[0054] FIG. 4C illustrates the cleavage of EGFP ssRNA at multiple sites with tiled crRNAs according to embodiments of the present teachings;

[0055] FIG. 4D illustrates the protospacer flanking site (PFS) sequence screen showing *in vitro* cleavage of randomized PFS targets and lack of sequence preference flanking the target site for DisCas7-11 cleavage according to embodiments of the present teachings;

[0056] FIG. 4E is a diagram of the kinetics of fluorescence due to RnaseAlert collateral cleavage by DisCas7-11 or LwaCas13a targeting and non-targeting guides against the MS2 RNA target according to embodiments of the present teachings;

[0057] FIG. 4F illustrates the processing activity of a synthetic DisCas7-11 CRISPR array by DisCas7-11 protein with predicted catalytic processing mutants in the protein insert region with a characteristic KxYxH catalytic according to embodiments of the present teachings;

[0058] FIG. 4G is the graph showing the measurement of bacterial growth (OD600) during target interference by DisCas7-11 and LwaCas13 according to embodiments of the present teachings;

[0059] FIG. 5A is a schematic of the protein domains of DisCas7-11 with putative catalytic residues highlighted in Cas7-like domains 3 and 4 according to embodiments of the present teachings;

[0060] FIG. 5B illustrates that the cleavage of synthetic ssRNA target with two crRNAs and wild type, D429A, D654A, and D429A/D654A DisCas7-11 proteins according to embodiments of the present teachings;

[0061] FIG. 5C illustrates the processing of a synthetic DisCas7-11 CRISPR array by catalytic DisCas7-11 mutants according to embodiments of the present teachings;

[0062] FIG. 5D is a diagram showing the MS2 phage interference in bacteria according to embodiments of the present teachings;

[0063] FIG. 5E is a diagram showing the RFP knockdown in *E. coli* that is prevented by the catalytic mutation D429A/D654A in DisCas7-11 according to embodiments of the present teachings;

[0064] FIG. 5F illustrates the DisCas7-11:crRNA complex binding to a complementary MS2 ssRNA target that is determined by electrophoretic mobility shift assay (EMSA) according to embodiments of the present teachings;

[0065] FIG. 6A illustrates DisCas7-11 guides targeting multiple sites on a synthetic 100nt EGFP ssRNA according to embodiments of the present teachings;

[0066] FIG. 6B are images of cleavage patterns of each EGFP-targeting guide incubated with DisCas7-11 on a 3' labeled EGFP ssRNA according to embodiments of the present teachings;

[0067] FIG. 6C illustrates DisCas7-11 that is incubated with a panel of crRNAs targeting either MS2 ssRNA or ssRNA target 2, two targets of equivalent length (153 nt) according to embodiments of the present teachings;

[0068] FIG. 6D are images of cleavage patterns of each guide incubated with MS2 ssRNA or ssRNA target 2 according to embodiments of the present teachings;

[0069] FIG. 7A tbc is a schematic of three Cas7-11 orthologs (GwCas7-11, CjcCas7-11, and DisCas7-11) according to embodiments of the present teachings;

[0070] FIG. 7B illustrates multiple sequence alignments of the DR sequences for GwCas7-11, CjcCas7-11, and DisCas7-11 according to embodiments of the present teachings;

[0071] FIG. 7C illustrates the activity of Cas7-11 from *Candidatus Jettenia caeni* (CjcCas7-11) according to embodiments of the present teachings;

[0072] FIG. 7D illustrates the activity of Cas7-11 from groundwater metagenomes (GwCas7-11) according to embodiments of the present teachings;

[0073] FIG. 7E illustrates the MS2 ssRNA target cleavage by GwCas7-11 and CjcCas7-11 incubated with their cognate crRNAs complementary to the MS2 ssRNA target according to embodiments of the present teachings;

[0074] FIG. 7F illustrates the incubation of GwCas7x3 and CjcCas7-11 proteins with their cognate crRNAs complementary to the MS2 ssRNA target at varying positions along the target according to embodiments of the present teachings;

[0075] FIG. 8A is a schematic of a luciferase reporter assay for RNA knockdown using a vector expressing human codon-optimized DisCas7-11 and guide RNA according to embodiments of the present teachings;

[0076] FIG. 8B is a diagram of the knockdown activity of DisCas7-11 against the *Gaussia* luciferase (Gluc) transcript normalized to two non-targeting conditions according to

[0077] FIG. 8C is a diagram showing the knockdown activity of DisCas7-11 against the multiple endogenous transcripts normalized to two non-targeting according to embodiments of the present teachings;

[0078] FIG. 8D is a schematic of the DisCas7-11a guide design for programmable A-to-I editing according to embodiments of the present teachings;

[0079] FIG. 8E is a diagram showing the RNA A-to-I editing of *Cypridina* luciferase (Cluc) mRNA W85X mutation in mammalian cells by dead DisCas7-11a-NES-ADAR2 according to embodiments of the present teachings;

[0080] FIG. 8F is a diagram showing the RNA A-to-I editing of *Cypridina* luciferase (Cluc) mRNA W85X mutation in mammalian cells by dead DisCas7-11a-NES-ADAR2 according to embodiments of the present teachings;

[0081] FIG. 8G is a graph illustrating the RFP knockdown by DisCas7-11 or LwaCas13a in *E. coli* bacteria according to embodiments of the present teachings;

[0082] FIG. 8H is a bar graph illustrating the knockdown of Gluc mRNA in mammalian cell by DisCas7-11 with a panel of guides containing the mature DR sequence according to embodiments of the present teachings;

[0083] FIG. 8I is a bar graph illustrating knockdown of Gluc mRNA in mammalian cells by DisCas7-11 with a panel of guides of different lengths according to embodiments of the present teachings;

[0084] FIG. 8J is a bar graph comparing the number of significant off-targets when targeting the Gluc transcript in HEK293FT and U87 cells by DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d using transcriptome-wide RNA-sequencing data according to embodiments of the present teachings;

[0085] FIG. 8K is a bar graph comparing the number of significant off-targets when targeting the Gluc transcript in HEK293FT and U87 cells by DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d using transcriptome-wide RNA sequencing data according to embodiments of the present teachings;

[0086] FIG. 8L is a bar graph showing the knockdown activity in HEK293FT cells of DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d against the Gaussia luciferase (Gluc) transcript normalized to corresponding non-targeting controls according to the embodiments of the present teachings;

5 [0087] FIG. 8M shows the transcriptome wide specificity in HEK293FT cells of RNA knockdown by DisCas7-11, RfxCas13d, and shRNA against the Gaussia luciferase (Gluc) transcript, wherein the analysis off targets is performed via significance testing (y-axis) of differentially expressed transcripts (x-axis) between targeting and non-targeting guide conditions (n = 3) and the number of significant off-targets is displayed according to
10 embodiments of the present teachings;

[0088] FIG. 8N is a bar graph showing the cell viability of HEK293FT cells during RNA knockdown by DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d against the Gaussia luciferase (Gluc) transcript according to embodiments of the present teachings;

[0089] FIG. 8O is a bar graph of the knockdown activity in U87 glioblastoma cells of
15 DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d against the Gaussia luciferase (Gluc) transcript normalized to corresponding non-targeting controls according to embodiments of the present teachings;

[0090] FIG. 8P is showing the transcriptome wide specificity in U87 glioblastoma cells of RNA knockdown by DisCas7-11, PspCas13b, and RfxCas13d against the Gaussia luciferase
20 (Gluc) transcript, wherein the analysis off targets is performed via significance testing (y-axis) of differentially expressed transcripts (x-axis) between the targeting guide condition and cells only expressing EGFP (n = 3) and the number of significant off-targets is displayed according to embodiments of the present teachings;

[0091] FIG. 8Q is a graph of the cell viability of U87 glioblastoma cells during RNA
25 knockdown by DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d against the Gaussia luciferase (Gluc) transcript, wherein the data are mean (n=3) $\pm$ s.e.m. according to embodiments of the present teachings;

[0092] FIG. 8R is a graph showing the knockdown activity in HEK293FT cells of DisCas7-11, LwaCas13a, PspCas13b, and RfxCas13d against the KRAS transcript normalized to
30 corresponding non-targeting controls according to embodiments of the present teachings;

[0093] FIG. 8S is a graph showing the knockdown activity in HEK293FT cells of DisCas7-11, LwaCas13a, PspCas13b, and RfxCas13d against the PPIB transcript normalized to corresponding non-targeting controls according to embodiments of the present teachings;

[0094] FIG. 8T is a graph showing the knockdown activity in HEK293FT cells of DisCas7-11, LwaCas13a, PspCas13b, and RfxCas13d against the MALAT transcript normalized to corresponding non-targeting controls according to embodiments of the present teachings;

[0095] FIG. 8U is a schematic of type III-E loci from Hydrothermal vent microbial

5 metagenome sample, Hydrothermal sediment microbial communities metagenome sample, and Deltaproteobacteria bacterium, containing HvsCas7-11, HsmCas7-11, and DpbaCas7-11 respectively according to embodiments of the present teachings;

[0096] FIG. 8V is a graph showing the knockdown activity in HEK293FT cells of HvsCas7-11, HsmCas7-11, DpbaCas7-11, and DisCas7-11 against the Gaussia luciferase (Gluc)

10 transcript normalized to corresponding non-targeting controls, wherein for HvsCas7-11, HsmCas7-11, and DpbaCas7-11, two DR orientations ("F" and "R") were tested, and wherein significance tests are measured by a Student's t-test (*, p-value < 0.05; **, p-value < 0.01) according to embodiments of the present teachings;

[0097] FIG. 8W is a graph showing the knockdown activity in HEK293FT cells of HvsCas7-

15 11 and DisCas7-11 with two distinct target guides, with and without catalytic mutations in the Cas7-11 proteins according to embodiments of the present teachings;

[0098] FIG. 8X is a graph showing the knockdown activity in HEK293FT cells of HvsCas7-11 against the PPIB transcript normalized to corresponding non-targeting controls according to embodiments of the present teachings;

20 [0099] FIG. 8Y is a graph showing the knockdown activity in HEK293FT, HepG2, and primary neuron cultures of HvsCas7-11 delivered as a two vector AAV system targeting the Gaussia luciferase (Gluc) transcript, wherein values are normalized to corresponding non-targeting controls, and wherein significance tests are measured by a Student's t-test (*, p-value < 0.05; **, p-value < 0.01) according to embodiments of the present teachings;

25 [0100] FIG. 8Z is a graph showing the knockdown activity in mouse embryonic stem cells (ESCs) of DisCas7-11, HvsCas7-11, PspCas13b, and RfxCas13d against the Gaussia luciferase (Gluc) transcript normalized to corresponding non-targeting controls according to embodiments of the present teachings;

[0101] FIG. 8AA is a graph showing the ribosomal RNA integrity analysis measured by

30 BioAnalyzer traces in U87 cells treated with DisCas7-11, shRNA, LwaCas13a, and PspCas13b targeting Gluc transcripts for knockdown (the normalized area under the curves for the 18S and 28S bands) according to embodiments of the present teachings;

[0102] FIG. 8AB are graph showing the comparison of DisCas7-11 and RfCas13d knockdown activity across multiple guides in HepG2 cells along with data on collateral activity as measured by Cluc levels, cell health (live stain), and cell toxicity (dead stain), wherein data are mean (n=3) $\pm$ s.e.m according to embodiments of the present teachings;

5 [0103] FIG. 8AC is a graph showing the collateral activity in mouse embryonic stem cells as measured by Cluc levels during Gluc RNA knockdown by DisCas7-11, HvsCas7-11, PspCas13b, and RfxCas13d, wherein the asterisks indicate significant as measured by a Student's t-test between the targeting and non-targeting conditions of a listed ortholog condition (**, p-value < 0.01), and wherein data are mean (n=3) $\pm$ s.e.m according to
10 embodiments of the present teachings;

[0104] FIG. 9A illustrates the *Desulfohalobium* *ishimotoi* Type III-E full locus matures crRNAs in *E. coli* according to embodiments of the present teachings;

[0105] FIG. 9B illustrates the single effector protein DisCas7-11 matures crRNAs in *E. coli* according to embodiments of the present teachings;

15 [0106] FIG. 9C illustrates the expression of the DisCas7-11 CRISPR array without any effector or accessory proteins according to embodiments of the present teachings;

[0107] FIG. 9D illustrates types III-E and III-D2 loci containing Cas7-11 and Cas7x3 respectively according to embodiments of the present teachings;

[0108] FIG. 10 are images of MS2 plaque dilution assays for MS2 targeting spacers, non-targeting spacers and RFP targeting spacers with DisCas7-11 according to embodiments of
20 the present teachings;

[0109] FIG. 11A illustrates the Type III-E locus of the *Desulfohalobium* *ishimotoi* according to embodiments of the present teachings;

[0110] FIG. 11B is a diagram showing the phage plaque assay of the Type III-E DisCas7-11 CRISPR-Cas locus according to embodiments of the present teachings;

25 [0111] FIG. 11C is a diagram showing the phage plaque assay of the individual gene knock-in together with of the Type III-E DisCas7-11 CRISPR-Cas system according to embodiments of the present teachings;

[0112] FIG. 12A illustrates the *in vitro* transcription of pre-crRNA and processing by DisCas7-11 according to embodiments of the present teachings;

30 [0113] FIG. 12B is a sequence of a single spacer CRISPR array showing locations of mutated bases according to embodiments of the present teachings;

[0114] FIG. 12C is an image of the processing of transcribed pre-crRNA showing the effect of single position mutations on processing according to embodiments of the present teachings;

[0115] FIG. 12D is an image of the processing of transcribed pre-crRNA showing the effect of consecutive double mutations on processing according to embodiments of the present teachings;

[0116] FIG. 13A illustrates the processing of pre-crRNA by DisCas7-11 in the presence of different ions or chelating agents according to embodiments of the present teachings;

[0117] FIG. 13B is a schematic of *in vitro* cleavage of ssRNA target with DisCas7-11 and crRNA 5 according to embodiments of the present teachings;

[0118] FIG. 13C illustrates the incubation of ssRNA target with DisCas7-11 and targeting or non-targeting crRNA in the presence of different ions or chelating agents according to embodiments of the present teachings;

[0119] FIG. 13D illustrates the cleavage of ssRNA target with increasing amounts of DisCas7-11-crRNA complex from 0 nM to 233 nM according to embodiments of the present teachings;

[0120] FIG. 13E illustrates the cleavage of ssRNA target at increasing incubation periods from 0 minutes to 180 minutes according to embodiments of the present teachings;

[0121] FIG. 14A is a schematic showing sequence of DisCas7-11 crRNA 5, targeting the ssRNA target according to embodiments of the present teachings;

[0122] FIG. 14B illustrates the cleavage of ssRNA target with crRNA of varying DR and spacer lengths according to embodiments of the present teachings;

[0123] FIG. 14C is a sequence of DisCas7-11 crRNA 1 targeting the MS2 target according to embodiments of the present teachings;

[0124] FIG. 14D illustrates the cleavage of the MS2 target according to embodiments of the present teachings;

[0125] FIG. 15 is a diagram of the fluorescence of RnaseAlert reporter incubated with either DisCas7-11 or LwaCas13a targeting and non-targeting guides against MS2 ssRNA target according to embodiments of the present teachings;

[0126] FIG. 16A is a schematic showing positions of DisCas7-11 crRNA 1 and crRNA 2 targeting the MS2 target according to embodiments of the present teachings;

[0127] **FIG. 16B** illustrates the cleavage of a synthetic MS2 ssRNA target with two crRNAs and wild type, D429A, D654A, and D429A/D654A DisCas7-11 proteins according to embodiments of the present teachings;

[0128] **FIG. 17A** illustrates an EMSA assay for targeting crRNA in the absence of DisCas7-11 according to embodiments of the present teachings;

[0129] **FIG. 17B** illustrates the quantification of band intensities for EMSA gels according to embodiments of the present teachings;

[0130] **FIG. 17C** illustrates the quantification of band intensities for EMSA gels according to embodiments of the present teachings;

[0131] **FIG. 17D** illustrates the quantification of band intensities for EMSA gels according to embodiments of the present teachings;

[0132] **FIG. 17E** illustrates the quantification of band intensities for EMSA gels according to embodiments of the present teachings;

[0133] **FIG. 17F** illustrates the quantification of band intensities for EMSA gels according to embodiments of the present teachings;

[0134] **FIG. 18A** illustrates DisCas7-11 incubated with a panel of crRNAs targeting either MS2 ssRNA or ssRNA target 2, two targets of equivalent length (153 nt) according to embodiments of the present teachings;

[0135] **FIG. 18B** illustrates the cleavage of ssRNA 1 and MS2 ssRNA at multiple sites with tiled crRNAs, showing a varying cleavage pattern based on target position according to embodiments of the present teachings;

[0136] **FIG. 18C** shows the schematic of the position of the tested crRNAs along the SARS-CoV-2 ssRNA and the results from the cleavage of synthetic SARS-CoV-2 ssRNA at multiple sites with tiled crRNAs showing a varying cleavage pattern based on target position according to embodiments of the present teachings;

[0137] **FIG. 18D** shows the schematic of the position of the tested crRNAs along the Influenza A ssRNA target and the results from the cleavage of synthetic influenza A segment 4 HA gene ssRNA at multiple sites with tiled crRNAs according to embodiments of the present teachings;

[0138] **FIG. 18E** illustrates DisCas7-11a incubated with a crRNA targeting a crRNA site flanked by different sequences according to embodiments of the present teachings;

[0139] FIG. 18F illustrates DisCas7-11a incubated with a crRNA targeting a double-stranded (dsDNA) target with 5' labeling of the bottom strand according to embodiments of the present teachings;

[0140] FIG. 18G illustrates DisCas7-11a incubated with a crRNA targeting a double-stranded (dsDNA) target with 5' labeling of the top strand according to embodiments of the present teachings;

[0141] FIG. 19A shows the schematic of the position of the tested crRNAs along the 100 nt EGFP ssRNA target and the biochemical characterization of programmable DisCas7-11 RNA cleavage of 100 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs according to embodiments of the present teachings;

[0142] FIG. 19B shows the schematic of the position of the tested crRNAs along the 200 nt EGFP ssRNA target and the biochemical characterization of the programmable DisCas7-11 RNA cleavage of 200 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs according to embodiments of the present teachings;

[0143] FIG. 19C shows the schematic of the position of the tested crRNAs along the 600 nt EGFP ssRNA target and the biochemical characterization of the programmable DisCas7-11 RNA cleavage of 600 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs according to embodiments of the present teachings;

[0144] FIG. 19D shows the DNA sequencing analysis of selected EGFP targeting crRNA conditions on an unlabeled 400 nt EGFP ssRNA target according to embodiments of the present teachings;

[0145] FIG. 19E shows the *in vitro* cleavage of ssRNA 1 at 37 °C with DisCas7-11 incubated with a crRNA against the 60 nt MS2 target labeled with a 5' or 3' Cy5 molecule, wherein the target cleavage is assessed with and without crRNA, DisCas7-11 protein, and Mg²⁺, and arrows on the target schematic show estimated cleavage positions of DisCas7-11 cleavage according to embodiments of the present teachings;

[0146] FIG. 19F shows cleavage of MS2 ssRNA at multiple sites with tiled crRNAs via incubation with DisCas7-11 protein, wherein MS2 ssRNA target is 5' labeled with Cy5 according to embodiments of the present teachings; and

[0147] FIG. 19G shows *in vitro* cleavage of MS2 ssRNA at 37 °C with varying concentrations of DisCas7-11 incubated with a crRNA against the target with and without Csx29 protein, wherein MS2 ssRNA target is 5' labeled with Cy5 and cleavage bands are marked by asterisks according to embodiments of the present teachings.

[0148] FIG. 20 illustrates the heterologous expression of the *Candidatus Jettenia caeni* Type III-E CRISPR-Cas system and associated CRISPR array in *E. coli*. according to embodiments of the present teachings;

[0149] FIG. 21A illustrates the processing of DisCas7-11 ortholog according to
5 embodiments of the present teachings;

[0150] FIG. 21B illustrates the processing of GwCas7-11 ortholog according to
embodiments of the present teachings.

[0151] FIG. 21C illustrates the processing of CjcCas7-11 ortholog according to
embodiments of the present teachings;

10 [0152] FIG. 21D illustrates a comparison of target cleavages between active CjcCas7-11 and
dead CjcCas7-11 according to embodiments of the present teachings;

[0153] FIG. 22A illustrates the DisCas7-11 guide design for RNA editing according to
embodiments of the present teachings;

15 [0154] FIG. 22B is a table showing the DisCas7-11 ADAR guide designs for according to
embodiments of the present teachings;

[0155] FIG. 22C is diagram showing the dDisCas7-11 correction of Cluc W85X RNA
according to embodiments of the present teachings;

[0156] FIG. 22D is a diagram showing the dDisCas7-11 correction of RNA as measured by
sequencing according to embodiments of the present teachings;

20 [0157] FIG. 22E is a diagram showing the luciferase correction that correlates to the
correction percent according to embodiments of the present teachings;

[0158] FIG. 23A illustrates the modulation of protein function that requires temporal A to G
modulation according to embodiments of the present teachings;

25 [0159] FIG. 23B illustrates the correction of mendelian disease that requires temporal A to G
modulation according to embodiments of the present teachings;

[0160] FIG. 23C illustrates a slicing modulation that requires temporal A to G modulation
according to embodiments of the present teachings;

[0161] FIG. 24A illustrates the initial and edited codon for protein modifications with A to G
RNA editing according to embodiments of the present teachings;

30 [0162] FIG. 24B illustrates first, second and third position codon modifications in RNA
editing according to embodiments of the present teachings;

[0163] FIG. 25 illustrates the initial codon, final codon and potential function in protein
modifications with C to U RNA editing according to embodiments of the present teachings;

[0164] FIG. 26A illustrates the correction of mendelian disease that requires C to U RNA editing according to embodiments of the present teachings;

[0165] FIG. 26B illustrates the multiplexed creation of disease-protective alleles that requires C to U RNA editing according to embodiments of the present teachings;

5 [0166] FIG. 26C illustrates the modulation of catalytic activity that requires C to U RNA editing according to embodiments of the present teachings;

[0167] FIG. 26D illustrates alter post-translation modifications that requires C to U RNA editing according to embodiments of the present teachings;

10 [0168] FIG. 27A-27C are tables of Cas7-11 effectors according to embodiments of the present teachings;

[0169] FIG. 28A illustrates the targeting of DisCas7-11 RNA target in beta-lactamase (ampicillin resistance gene) and the resulting PFS determined by 20 depleted targets according to embodiments of the present teachings;

15 [0170] FIG. 28B illustrates the targeting of DisCas7-11 DNA target in a plasmid, with a motif corresponding to no significantly depleted PFS sequences according to embodiments of the present teachings;

[0171] FIG. 28C is a diagram showing the PFS analysis of top spacers from the MS2 phage screen (e-5 condition) in the 8 bp flanking the target region to the left according to embodiments of the present teachings;

20 [0172] FIG. 28D is a diagram showing the PFS analysis of top spacers from the MS2 phage screen (e-5 condition) in the 8 bp flanking the target region to the right according to embodiments of the present teachings;

[0173] FIG. 28E is a schematic of the heterologous type III-E locus of the *Desulfonema ishimotonii* according to embodiments of the present teachings;

25 [0174] FIG. 28F is a bar graph showing the phage plaque assay of the Type III-E DisCas7-11 effector alone and as part of it entire locus according to embodiments of the present teachings;

[0175] FIG. 28G is a bar graph showing the phage plaque assay of *E. coli* transformed with the DisCas7-11 locus containing according to embodiments of the present teachings;

30 [0176] FIG. 29 illustrates the DisCas7-11a cleavage of synthetic 31nt MS2 ssRNA with a 31nt crRNA according to embodiments of the present teachings;

[0177] FIG. 30A shows the secondary structure folding prediction of the DisCas7-11a DR according to embodiments of the present teachings;

[0178] FIG. 30B shows the secondary structure folding prediction of the CjcCas7-11b DR according to embodiments of the present teachings;

[0179] FIG. 30C shows the structure folding prediction of the GwCas7-11c DR in accordance with embodiments of the present teachings;

5 [0180] FIG. 31A illustrates the comparison of knockdown activity of Gluc mRNA in mammalian cells between active DisCas7-11a, catalytically inactive D429A/D654A DisCas7-11a, and GFP according to embodiments of the present teachings;

[0181] FIG. 31B illustrate the comparison of knockdown activity of endogenous mRNA in mammalian cells between active DisCas7-11a, catalytically inactive D429A/D654A

10 DisCas7-11a, and GFP according to embodiments of the present teachings;

[0182] FIG. 32A is a diagram of the RNA A-to-I editing of *Cypridinia* luciferase (cluc) mRNA W85X mutation in mammalian cells by active DisCas7-11a-NES-ADAR2 or dead DisCas7-11a-NES-ADAR2 according to embodiments of the present teachings;

15 [0183] FIG. 32B is a diagram of the RNA A-to-I editing of *Cypridinia* luciferase (cluc) mRNA W85X mutation in mammalian cells by active DisCas7-11a-NES-ADAR2 or dead DisCas7-11a-NES-ADAR2 according to embodiments of the present teachings;

[0184] FIG. 33A shows the cleavage of 3' labeled targeted by incubation with RNase H according to embodiments of the present teachings;

20 [0185] FIG. 33B shows the cleavage of 5' labeled targeted by incubation with RNase H according to embodiments of the present teachings;

[0186] FIG. 34A shows domains and section 1 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0187] FIG. 34B shows section 2 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

25 [0188] FIG. 34C shows section 3 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0189] FIG. 34D shows section 4 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

30 [0190] FIG. 34E shows section 5 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0191] FIG. 34F shows section 6 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0192] FIG. 34G shows section 7 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0193] FIG. 34H shows section 8 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

5 [0194] FIG. 34I shows section 9 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0195] FIG. 34J shows section 10 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

10 [0196] FIG. 34K shows section 11 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0197] FIG. 34L shows section 12 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

[0198] FIG. 34M shows section 13 of sequence alignments of Cas nucleases according to embodiments of the present teachings;

15 [0199] FIG. 34N shows section 14 of sequence alignments of Cas nucleases according to embodiments of the present teachings; and

[0200] FIG. 34O shows section 15 of sequence alignments of Cas nucleases according to embodiments of the present teachings.

20

DETAILED DESCRIPTION

[0201] It will be appreciated that for clarity, the following discussion will describe various aspects of embodiments of the applicant's teachings. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment

25

is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment", "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more

30

embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

5 **[0202]** General Definitions

[0203] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and
10 Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel *et al.* eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.); Antibodies A Laboratory Manual, 2nd edition 2013 (E.A.
15 Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN
20 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

25 **[0204]** As used herein, the singular forms "a", "an," and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0205] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

30 **[0206]** The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0207] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to

encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, +/-0.5% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0208] Overview

[0209] The embodiments disclosed herein provide (non-naturally occurring or engineered) constructs, compositions, systems, and methods for site-directed RNA editing of RNA molecules. For example, the present invention provides (non-naturally occurring or engineered) methods for inhibiting intra and inter-cellular signaling pathways by modification of post-translational modification sites on select target RNA molecules. In certain example embodiments, the present invention provides (non-naturally occurring or engineered) methods for inhibiting intracellular phosphorylation of serine, threonine and tyrosine residues by editing the genetic codon of these amino acids by means of site-directed RNA editing or RNA molecules. Embodiments disclosed herein further provide methods of inhibiting pathological activation of cell signaling mediated by post-translational modifications, such as phosphorylation, which are involved in many diseases, including cancer, immunodeficiency, infectious diseases, inflammatory disorders and neurodegenerative disorders. The RNA-editing modification may be aimed at a single post-translational modification site of a single gene, but can also be multiplexed by targeting multiple sites on the same or different genes to increase efficacy. These approaches may be further combined with other treatments such as radiation, chemotherapy, targeted therapy based on antibodies or small molecules, and immunotherapy, which may have a synergistic effect.

[0210] The embodiments disclosed herein provide (non-naturally occurring or engineered) systems, constructs, and methods for targeted base editing. In general, the systems disclosed herein comprise a targeting component and a base editing component. The targeting component may function to specifically target the base editing component to a target nucleotide sequence in which one or more nucleotides are to be edited. The base editing component may then catalyze a chemical reaction to convert a first nucleotide in the target sequence to a second nucleotide. For example, the base editor may catalyze conversion of an adenine such that it is read as guanine by a cell's transcription or translation machinery, or vice versa. Likewise, the base editing component may catalyze conversion of cytidine to an

uracil, or vice versa. In certain example embodiments, the base editor may be derived by starting with a known base editor, such as an adenine deaminase or cytidine deaminase, and modified using methods such as directed evolution to derive new functionalities. Directed evolution techniques are known in the art and may include those described in WO

5 2015/184016 "High-Throughput Assembly of Genetic Permutations."

[0211] Compositions and Systems

[0212] The present disclosure provides (non-naturally occurring or engineered) systems for editing a nucleic acid such as a gene or a product thereof (*e.g.*, the encoded RNA or protein).

In some embodiments, the systems may be an engineered, non-naturally occurring system
 10 suitable for modifying post-translational modification sites on proteins encoded by a target nucleic acid sequence. In certain cases, the target nucleic acid sequence is RNA, *e.g.*, mRNA or a fragment thereof. In certain cases, the target nucleic acid sequence is DNA, *e.g.*, a gene or a fragment thereof. In general, the system may comprise one or more of a catalytic inactive (dead) Cas protein (*e.g.*, dead Cas7-11), a nucleotide deaminase protein or catalytic domain
 15 thereof, and a guide molecule. In certain examples, the nucleotide deaminase protein may be an adenosine deaminase. In certain examples, the nucleotide deaminase protein may be a cytidine deaminase. The guide sequence may be designed to have a degree of complementarity with a target sequence at one or more codons comprising an adenine or cytidine and that is post-translationally modified.

20 **[0213] CRISPR-Cas**

[0214] Some embodiments disclosed herein are directed to CRISPR-Cas (clustered regularly interspaced short palindromic repeats associated proteins) systems. In the conflict between bacterial hosts and their associated viruses, CRISPR-Cas systems provide an adaptive defense mechanism that utilizes programmed immune memory. CRISPR-Cas systems provide their
 25 defense through three stages: adaptation, the integration of short nucleic acid sequences into the CRISPR array that serves as memory of past infections; expression, the transcription of the CRISPR array into a pre-crRNA (CRISPR RNA) transcript and processing of the pre-crRNA into functional crRNA species targeting foreign nucleic acids; and interference, the programming of CRISPR effectors by crRNA to cleave nucleic acid of foreign threats.

30 Across all CRISPR-Cas systems, these fundamental stages display enormous variation, including the identity of the target nucleic acid (either RNA, DNA, or both) and the diverse domains and proteins involved in the effector ribonucleoprotein complex of the system.

[0215] CRISPR-Cas systems can be broadly split into two classes based on the architecture of the effector modules involved in pre-crRNA processing and interference. Class 1 systems have multi-subunit effector complexes composed of many proteins, whereas Class 2 systems rely on single-effector proteins with multi-domain capabilities for crRNA binding and interference; Class 2 effectors often provide pre-crRNA processing activity as well. Class 1 systems contain 3 types (type I, III, and IV) and 33 subtypes, including the RNA and DNA targeting type III-systems. Class 2 CRISPR families encompass 3 types (type II, V, and VI) and 17 subtypes of systems, including the RNA-guided DNases Cas9 and Cas12 and the RNA-guided RNase Cas13. Continual sequencing of novel bacterial genomes and metagenomes uncovers new diversity of CRISPR-Cas systems and their evolutionary relationships, necessitating experimental work that reveals the function of these systems and develops them into new tools.

[0216] Among the currently known CRISPR-Cas systems, only the type III and type VI systems have been demonstrated to bind and target RNA, and these two systems have substantially different properties, the most distinguishing being their membership in Class 1 and Class 2, respectively. Characterized subtypes of type III, which span type III-A, B, and C systems, target both RNA and DNA species through an effector complex containing multiple Cas7 (Csm3/5 or Cmr1/4/6) RNA nuclease units in association with a single Cas10 (Csm1 or Cmr2) DNA nuclease. The RNA nuclease activity of Cas7 is mediated through acidic residues in the repeat-associated mysterious proteins (RAMP) domains, which cut at stereotyped intervals in the guide:target duplex. Type III systems also have a target restriction, and cannot efficiently target protospacers *in vivo* if there is extended homology between the 5' "tag" of the crRNA and the "anti-tag" 3' of the protospacer in the target, although this binding does not block RNA cleavage *in vitro*. In type III systems, pre-crRNA processing is carried out by either host factors or the associated Cas6 family protein, which can physically complex with the effector machinery.

[0217] In contrast to type III systems, type VI systems contain a single CRISPR effector Cas13 that can only effect RNA interference, mediated through basic catalytic residues of dual HEPN domains. This interference requires a protospacer flanking sequence (PFS), although the influence of the PFS varies between orthologs and families. Importantly, the RNA cleavage activity of Cas13, once triggered by crRNA:target duplex formation, is indiscriminate, and activated Cas13 enzymes will cleave other RNA species *in vitro*, in bacterial hosts, and mammalian cells. This activity, termed the collateral effect, has been

applied to CRISPR-based nucleic acid detection technologies. In addition to the RNA interference activity, the Cas13 family members contain pre-crRNA processing activity. Just as single-effector DNA targeting systems have given rise to numerous genome editing applications, Cas13 family members have been applied to a suite of RNA-targeting technologies in both bacterial and eukaryotic cells, including RNA knockdown, RNA editing, RNA tracking, epitranscriptome editing, translational upregulation, epi-transcriptomic reading and writing via N6-Methyladenosine, and isoform modulation.

[0218] The novel type III-E system was recently identified from genomes of 8 bacterial species and is characterized as a fusion of several Cas7 proteins and a putative Cas11 (Csm2)-like small subunit. The domain composition suggests the fusion of multiple type III effector module domains involved in crRNA binding into a single protein effector that is predicted to process pre-crRNA given its homology with Cas5 (Csm4) and conserved aspartates. The lack of other putative effector nucleases in these CRISPR loci raise the additional possibility that this fusion protein is capable of crRNA-directed RNA cleavage. If so, this system would blur the distinction of Class 1 and Class 2 systems, as it would have domains homologous to other Class 1 systems, but possess a single effector module characteristic of Class 2 systems. Beyond the single effector module present in all subtype III-E loci, a majority of type III-E family members contain a putative ancillary gene with a CHAT domain, which is a caspase family protease associated with programmed cell death (PCD), suggesting involvement of PCD-mediated antiviral strategies, as has been observed with type III and VI systems.

[0219] Type III-E system associated effector, interchangeably termed Cas15 and Cas7-11 herein, is a programmable RNase (FIGS. 1A-1D). This system can provide defense against RNA phage and be programmed to target exogenous mRNA species when expressed heterologously in bacteria. Orthologs of Cas7-11 are capable of both processing of pre-crRNA and crRNA-directed cleavage of RNA targets, and determine catalytic residues underlying programmed RNA cleavage. A direct evolutionary path of Cas7-11 can be traced from individual Cas7 and Cas11 effector proteins of subtype III-D1 variant, through an intermediate, a partially fused effector Cas7x3 of the subtype III-D2 variant, to the single-effector architecture of subtype III-E that is so far unique among the Class 1 CRISPR-Cas systems (FIG. 1D). Cas7-11 most likely originated from two type III-D variants. Three Cas7 domains (domains 3, 4 and 5) are derived from subtype III-D2 that contains a the Cas7x3 effector protein along with Cas10 and another Cas7-like domain fused to a Cas5-like domain.

The origin of the N-terminal Cas7 and putative Cas11 domain of Cas7-11 is most likely derived from a III-D1 variant, where both genes are stand-alone.

[0220] Cas7-11 differs substantially from Cas13, in terms of both domain organization and activity (FIG. 1C). Cas13 RNA cleavage is enacted by dual HEPN domains with basic catalytic residues, and this cleavage, once triggered, is indiscriminate. In contrast, Cas7-11 utilizes at least two of four Cas7-like domains with acidic catalytic residues to generate stereotyped cleavage at the target binding site in cis. Furthermore, Cas13 targeting is restricted by the requirement for a PFS, which Cas7-11 does not require, and the DR of Cas7-11-associated crRNA is substantially shorter. Because of these unique features, Cas7-11 may have distinct advantages for RNA targeting and transcriptome engineering biotechnology applications.

[0221] Regulation of interference by accessory proteins has been observed in both type III and type VI systems, and other proteins in the *D. ishimotonii* type III-E locus can regulate activity of DisCas7-11a. Notably, TPR-CHAT had a strong inhibitory effect on DisCas7-11a phage interference, raising the possibility that unrestricted DisCas7-11a activity could be detrimental for the host. Alternatively, as TPR-CHAT is a caspase family protease associated with programmed cell death (PCD), it is possible that TPR-CHAT is activated by DisCas7-11a and leads to host death, which could mimic death due to phage in these assays. TPR-CHAT caspase activity could be activated by DisCas7-11a and cause PCD through general proteolysis, analogous to PCD triggered by Cas13 collateral activity.

[0222] Similar to Class 2 CRISPR effectors such as Cas9, Cas12, and Cas13, Cas7-11 is highly active in mammalian cells, with substantial knockdown activity on both reporter and endogenous transcripts. Moreover, via inactivation of active sites through mutagenesis, the catalytically inactive dCas7-11 enzyme can be used to recruit ADAR2DD for efficient site-specific A-to-I editing on transcripts. These applications establish Cas7-11 as the basis for an RNA-targeting toolbox that has several benefits compared to Cas13, including the lack of sequence preferences and collateral activity, the latter of which has been shown to induce toxicity in certain cell types. A Cas7-11 toolbox may serve as the basis for multiple RNA technologies, including RNA knockdown, RNA editing, translation modulation, RNA recruitment, RNA tracking, splicing control, RNA stabilization, and potentially even diagnostics.

[0223] AD-functionalized CRISPR Systems

[0224] In some embodiments, the systems may be AD-functionalized CRISPR system. The term "AD-functionalized CRISPR system" as used here refers to a nucleic acid targeting and editing system comprising (a) a CRISPR-Cas protein, more particularly a Cas7-11 protein which is catalytically active or inactive; (b) a guide molecule which comprises a guide

sequence; and (c) an adenosine deaminase (AD) protein or catalytic domain thereof; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the CRISPR-Cas protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is substantially complementary to the target sequence but comprises a non-pairing C corresponding to the A being targeted for deamination,

resulting in an A-C mismatch in an RNA duplex formed by the guide sequence and the target sequence. In some embodiments, the CRISPR-Cas protein and/or the adenosine deaminase comprise one or more heterologous nuclear export signal(s) (NES(s)) or nuclear localization signal(s) (NLS(s)). For application in eukaryotic cells, the CRISPR-Cas protein and/or the adenosine deaminase can be NES-tagged or NLS-tagged.

[0225] One skilled in the art would appreciate that the components (a), (b) and (c) can be delivered to the cell as a ribonucleoprotein complex. The ribonucleoprotein complex can be delivered via one or more lipid nanoparticles. One skilled in the art would appreciate that the components (a), (b) and (c) can be delivered to the cell as one or more RNA molecules, such as one or more guide RNAs and one or more mRNA molecules encoding the CRISPR-Cas protein, the adenosine deaminase protein, and optionally the adaptor protein. The RNA molecules can be delivered via one or more lipid nanoparticles. One skilled in the art would appreciate that the components (a), (b) and (c) can be delivered to the cell as one or more DNA molecules. The one or more DNA molecules can be comprised within one or more vectors such as viral vectors (*e.g.*, AAV). The one or more DNA molecules can comprise one or more regulatory elements operably configured to express the CRISPR-Cas protein, the guide molecule, and the adenosine deaminase protein or catalytic domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters.

[0226] In some embodiments, the CRISPR-Cas protein is a dead Cas7-11. In some embodiments, the dead Cas7-11 comprises one or more mutations in the Cas7-like domains, including D429A and D654A as well as many other mutations (see table 1).

[0227] In some embodiments, the guide molecule is capable of hybridizing with a target sequence comprising the Adenine to be deaminated within an RNA sequence to form an RNA duplex which comprises a non-pairing Cytosine opposite to said Adenine. Upon RNA

duplex formation, the guide molecule forms a complex with the Cas7-11 protein and directs the complex to bind the RNA polynucleotide at the target RNA sequence of interest. Details on the aspect of the guide of the AD-functionalized CRISPR-Cas system are provided herein below.

- 5 **[0228]** In at least a first design, the AD-functionalized CRISPR system comprises: (a) an adenosine deaminase fused or linked to a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive; and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence. In some embodiments, the CRISPR-Cas protein and/or the
- 10 adenosine deaminase can be NLS-tagged, on either the N- or C-terminus or both.

- [0229]** In at least a second design, the AD-functionalized CRISPR system comprises: (a) a CRISPR-Cas protein that is catalytically inactive; (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence, and an aptamer sequence (*e.g.*, MS2 RNA motif or
- 15 PP7 RNA motif) capable of binding to an adaptor protein (*e.g.*, MS2 coating protein or PP7 coat protein); and (c) an adenosine deaminase fused or linked to an adaptor protein, wherein the binding of the aptamer and the adaptor protein recruits the adenosine deaminase to the RNA duplex formed between the guide sequence and the target sequence for targeted deamination at the A of the A-C mismatch. In some embodiments, the adaptor protein and/or
- 20 the adenosine deaminase can be NLS-tagged, on either the N- or C-terminus or both. The CRISPR-Cas protein can also be NLS-tagged. The CRISPR-Cas protein can also be NLS-tagged.

- [0230]** The use of different aptamers and corresponding adaptor proteins also allows orthogonal gene editing to be implemented. In one example in which adenosine deaminase
- 25 are used in combination with cytidine deaminase for orthogonal gene editing/deamination, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-adenosine deaminase and PP7-cytidine deaminase (or PP7-adenosine deaminase and MS2-cytidine deaminase), respectively, resulting in orthogonal deamination of A or C at the target loci of interest, respectively. PP7 is the RNA-binding coat protein of the bacteriophage
- 30 *Pseudomonas*. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-adenosine

deaminase, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-cytidine deaminase. In the same cell, orthogonal, locus-specific modifications are thus realized. This principle can be extended to incorporate other orthogonal RNA-binding proteins.

- 5 **[0231]** In at least a third design, the AD-functionalized CRISPR system comprises: (a) an adenosine deaminase inserted into an internal loop or unstructured region of a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive or a nickase; and (b) a guide molecule comprising a guide sequence designed to introduce an A-C mismatch in an RNA duplex formed between the guide sequence and the target sequence.
- 10 **[0232]** The AD-functionalized CRISPR system described herein can be used to target a specific Adenine within an RNA polynucleotide sequence for deamination. For example, the guide molecule can form a complex with the CRISPR-Cas protein and directs the complex to bind a target RNA sequence in the RNA polynucleotide of interest. Because the guide sequence is designed to have a non-pairing C, the RNA duplex formed between the guide
- 15 sequence and the target sequence comprises an A-C mismatch, which directs the adenosine deaminase to contact and deaminate the A opposite to the non-pairing C, converting it to an Inosine (I). Since Inosine (I) base pairs with C and functions like G in cellular processes, the targeted deamination of A described herein are useful for correction of undesirable G-A and C-T mutations, as well as for obtaining desirable A-G and T-C mutations.
- 20 **[0233]** In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in an RNA polynucleotide molecule *in vitro*. In some embodiments, the AD-functionalized CRISPR system is used for targeted deamination in a DNA molecule and/or RNA molecule within a cell. The cell can be an eukaryotic cell such as a bacteria or cyanobacteria. The cell can be a eukaryotic cell, such as an animal cell, a mammalian cell, a
- 25 human, or a plant cell.
- 30 **[0234]** The invention also relates to a (non-naturally occurring or engineered) method for treating or preventing a disease by the targeted deamination using the AD-functionalized CRISPR system, wherein the deamination of the A, which remedies a disease caused by transcripts containing a pathogenic G→A or C→T point mutation. Examples of disease that can be treated or prevented with the present invention include cancer, Meier-Gorlin syndrome, Seckel syndrome 4, Joubert syndrome 5, Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2; Charcot-Marie-Tooth disease, type 2; Usher syndrome, type 2C; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Spinocerebellar ataxia 28; Long QT

syndrome 2; Sjogren-Larsson syndrome; Hereditary fructosuria; Hereditary fructosuria; Neuroblastoma; Neuroblastoma; Kallmann syndrome 1; Kallmann syndrome 1; Kallmann syndrome 1; Metachromatic leukodystrophy.

[0235] AD-functionalized CRISPR system for RNA editing can be used for translation upregulation or downregulation, improving RNA stability and diagnostics. For example, for application in diagnostics, TPR-Chat is an accessory protein that interacts with Cas7-11 interference. Cas7-11 can activate TPR-Chat caspase activity which can then activate a reporter. While this can be used for inducing cell death based on RNA detection (*e.g.* in cancer cells), it also can be useful for general RNA diagnostics (*i.e.* molecular diagnostics for bacteria, viruses, and derivatives thereof) in samples. Furthermore, Cas7-11 can re-constitute a split protein like GFP on a specific transcript.

[0236] AD-functionalized CRISPR system for RNA editing can be used to treat or prevent premature termination diseases. Pre-termination diseases are characterized by mutations in early stop codons, either through single nucleotide polymorphisms that introduce termination, indels that change the translational frame of the protein and generate new stop codons, or alternative splicing that preferentially introduces exons that have early termination. By removing stop codons generated in these ways via A to I editing, RNA editing with ADAR could rescue diseases involving premature termination. In cases where SNPs are not G to A, but generate nonsense mutations, clinical benefit could be derived from changing nonsense mutations into missense mutations.

[0237] AD-functionalized CRISPR system for RNA editing can be used to change fertility mutations without germline editing. One advantage of RNA editing over DNA editing is in cases of SNPs affecting fertility, where correction with genome editing would necessarily result in germline editing, with potential ethical or safety implications. RNA editing could correct these mutations without permanent effects on the genome, thereby circumventing these issues.

[0238] AD-functionalized CRISPR system for RNA editing can be used for splicing alteration. Pre-mRNA requires specific splice donor and acceptor sequences in order to undergo processing by the spliceosome. Splice acceptor sites contain an invariant AG sequence that is necessary for acceptance of the attack by the splice donor sequence and intron removal. By targeting Cas7-11-ADAR fusions to pre-mRNA and editing AG splice acceptor sites to IG, it can be possible to inactivate the splice acceptor site, resulting in skipping of the downstream exon. This approach to splicing alteration has advantages over

the current method of exon skipping with chemically modified anti-sense oligos. Cas7-11-ADAR can be genetically encoded, allowing for long-term exon skipping. Additionally, Cas7-11-ADAR creates a mutation to promote skipping, which can be more robust than masking of the splice donor/acceptor site by a double stranded RNA, as is done with anti-sense oligos.

[0239] AD-functionalized CRISPR system for RNA editing can be used to alter neoantigens. Neoantigens in cancer are novel antigens that are expressed in tumor cells due to mutations that arise because of defective mismatch repair. Engineering T cells against neoantigens is advantageous because the T cells will have no off-target activity and thus toxicity since the antigens are only expressed in the tumor cells. With RNA base editors, the Cas7-11-ADAR fusions can be targeted to cancer cells to introduce mutations in transcripts that would introduce amino acid changes and new antigens that can be targeted using chimeric antigen receptor T cells. This approach is better than DNA base editors because it is transient and thus the risk of editing non-tumor cells permanently due to off-target delivery is minimal.

[0240] AD-functionalized CRISPR system for RNA editing can be used to change microRNA targets for tumor suppressors. ADAR naturally edits mRNA to generate or remove microRNA targets, thereby modulating expression. Programmable RNA editing can be used to up- or down- regulate microRNA targets via altering of targeting regions. Additionally, microRNAs themselves are natural substrates for ADAR, and programmable RNA editing of microRNAs can reduce or enhance the function on their corresponding targets.

[0241] AD-functionalized CRISPR system for RNA editing can be used to make multiple edits along a region. The Cas7-11-ADAR fusions can be precisely targeted to edit specific adenosines by introducing a mismatch in the guide region across from the desired adenosine target and creating a bubble that is favorable for A-to-I editing. By introducing multiple of these mismatches across different adenosine sites in the guide/target duplex, it can be possible to introduce multiple mutations at once.

[0242] AD-functionalized CRISPR system for RNA editing can be used for the reversal of TAA (double A to G) for PTC. Many diseases that involve pretermination codon changes involve a TAA stop codon, which would require A-to-I changes to correct rather than the TAG or TGA stop codons which only need one A-to-I edit. Two approaches can be used to reverse the TAA stop codon. (1) As described in the previous section, two mismatches can be introduced in the guide against the two adenosines in the TAA codon. (2) A two-guide array

can be used to convert each of the adenosines to inosine sequentially. The first guide in the array can contain a mutation against the first adenosine and the second guide can then have complementarity to this change and have a mismatch against the second adenosine in the stop codon.

- 5 [0243] AD-functionalized CRISPR system for RNA editing can be used to treat or prevent cancer (GOF, LOF mutation reversal). Many oncogenic changes in cancer involve G to A mutations that introduce gain of function or loss of function phenotypes to the mutated proteins. The RNA base editors are well positioned to correct these changes and reduce oncogenesis.
- 10 [0244] RNA editing with ADAR can be used for the design of new base preferences. Current ADAR1/2 proteins have been found to have surrounding base preferences for catalytic activity, which may pose constraints for certain applications. Rational mutagenesis or directed evolution of ADAR variants with altered or relaxed base preferences can increase the versatility of programmable RNA editing.
- 15 [0245] AD-functionalized CRISPR system for RNA editing can comprise ADAR mutants with increased activity in human cells. Although ADAR mutants with altered activity *in vitro* or in yeast have been previously reported, screening or rational design of mutants with increased activity in the context of human cells can improve the efficiency or specificity of ADAR-based programmable RNA editing constructs.
- 20 [0246] AD-functionalized CRISPR system for RNA editing can be used in biological applications of inosine generation. The RNA editing with ADAR generates inosine, which, when occurring multiple times in a transcript, can interact with endogenous biological pathways to increase inflammation in cells and tissues. Generation of multiple inosine bases can increase inflammation, especially in cells where inflammation can lead to clearance.
- 25 Additional inosine generation could also be used to destabilize transcripts.
- [0247] AD-functionalized CRISPR system for RNA editing can be used in removing upstream start codons to promote protein expression of downstream ORF (ATG mutation). Anti-sense oligos have been used for blocking upstream start codon sites to promote protein expression at downstream start codons. This allows the boosting of endogenous protein levels
- 30 for therapeutic purposes. Cas7-11-ADAR fusions could accomplish a similar effect by converting ATG sites to ITG (GTG) sites and thus remove upstream codons in endogenous transcripts and thus boost protein translation. So far, most therapeutic applications discussed have been for correcting G to A mutations or removing pre-termination sites. This would be

an application that allows for boosting gene expression. A good example is boosting fetal hemoglobin levels in sickle cell disease and thalassemias.

[0248] AD-functionalized CRISPR system for RNA editing can comprise the mutagenesis of ADAR for C to U or any transition. It is possible through rational mutagenesis or directed evolution that the ADARs listed in the ortholog section could be made into C to U editors or editors of any base transition.

[0249] FIGS. 23-26 illustrates codon changes and their effects when RNA editors are used. FIG. 23A shows that modulation can be used for modifying protein function through amino acid changes that affect enzyme domains, such as kinases. FIG. 23B shows that modulation can be used for correction of Mendelian disease mutations. FIG. 23C shows the splicing modulation of transcripts by modifying the splice acceptor site. FIGS. 24A-B show potential codon transitions enabled by A>I editing. FIG. 25 shows show potential codon transitions and function enabled by C>U editing. Adapted and modified based on J. D. Watson, Molecular biology of the gene. (Pearson, Boston, ed. Seventh edition, 2014), pp. xxxiv, 872 pages (38). FIG. 66C). FIG. 26A shows that C-U modulation can be used for correction of Mendelian disease mutations. FIG 26B shows the multiplexed creation of disease-protective alleles that requires C to U RNA editing. FIG. 26C shows the modulation of catalytic activity that requires C to U RNA editing. FIG. 26D shows alter post-translational modifications that requires C to U RNA editing.

[0250] In particular embodiments, the compositions described herein can be used in therapy. This implies that the methods can be performed *in vivo*, *ex vivo* or *in vitro*. In particular embodiments, the methods can be not methods of treatment of the animal or human body or a method for modifying the germ line genetic identity of a human cell. In particular embodiments, when carrying out the method, the target RNA can be not comprised within a human or animal cell. In particular embodiments, when the target is a human or animal target, the method can be carried out *ex vivo* or *in vitro*.

[0251] CRISPR-Cas Proteins and Guides

[0252] In some embodiments, the system comprises one or more components of a CRISPR-Cas system. For example, the system may comprise a Cas protein, a guide molecule, or a combination thereof.

[0253] In the methods and systems of the present invention use is made of a CRISPR-Cas protein and corresponding guide molecule. More particularly, the CRISPR-Cas protein is a class 2 CRISPR-Cas protein. In certain embodiments, said CRISPR-Cas protein is a Cas7-11.

The Cas7-11 may be Cas7-11a, Cas7-11b, Cas7-11c, or Cas7-11d. The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas protein can be programmed by guide molecule to recognize a specific nucleic acid target, in other words the Cas enzyme protein can be recruited to a specific nucleic acid target locus of interest using said guide molecule.

[0254] CRISPR-Cas proteins

[0255] In some embodiments, the systems may comprise a CRISPR-Cas protein. In certain examples, the CRISPR-Cas protein may be a catalytically inactive (dead) Cas protein. The catalytically inactive (dead) Cas protein may have impaired (*e.g.*, reduced or no) nuclease activity. In some cases, the dead Cas protein may have nickase activity. In some cases, the dead Cas protein may be dead Cas 15 protein. For example, the dead Cas 15 may be dead Cas7-11a, dead Cas7-11b, dead Cas7-11c, or dead Cas7-11d. In some embodiments, the system may comprise a nucleotide sequence encoding the dead Cas protein.

[0256] In its unmodified form, a CRISPR-Cas protein is a catalytically active protein. This implies that upon formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence) one or both DNA strands in or near (*e.g.*, within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence is modified (*e.g.* cleaved). As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (*e.g.* within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest). The unmodified catalytically active Cas7-11 protein generates a staggered cut, whereby the cut sites are typically within the target sequence. More particularly, the staggered cut is typically 13–23 nucleotides distal to the PAM. In particular embodiments, the cut on the non-target strand is 17 nucleotides downstream of the PAM (*i.e.* between nucleotide 17 and 18 downstream of the PAM), while the cut on the target strand (*i.e.* strand hybridizing with the guide sequence) occurs a further 4 nucleotides further from the sequence complementary to the PAM (this is 21 nucleotides upstream of the complement of the PAM on the 3' strand or between nucleotide 21 and 22 upstream of the complement of the PAM).

[0257] In the methods according to the present invention, the CRISPR-Cas protein is preferably mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks the ability to cleave one or both DNA strands of a target locus containing a target sequence. In particular embodiments, one or more catalytic domains of the

Cas7-11 protein are mutated to produce a mutated Cas protein which cleaves only one DNA strand of a target sequence.

[0258] In particular embodiments, the CRISPR-Cas protein may be mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks

5 substantially all DNA cleavage activity. In some embodiments, a CRISPR-Cas protein may be considered to substantially lack all DNA and/or RNA cleavage activity when the cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as
10 compared with the non-mutated form.

[0259] In certain embodiments of the methods provided herein the CRISPR-Cas protein is a mutated CRISPR-Cas protein which cleaves only one DNA strand, *i.e.* a nickase. More particularly, in the context of the present invention, the nickase ensures cleavage within the non-target sequence, *i.e.* the sequence which is on the opposite DNA strand of the target
15 sequence and which is 3' of the PAM sequence.

[0260] In some embodiments, a CRISPR-Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the
20 mutated form is nil or negligible as compared with the non-mutated form. In these embodiments, the CRISPR-Cas protein is used as a generic DNA binding protein. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations.

[0261] In addition to the mutations described above, the CRISPR-Cas protein may be additionally modified. As used herein, the term "modified" with regard to a CRISPR-Cas
25 protein generally refers to a CRISPR-Cas protein having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type Cas protein from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some
30 way as known in the art or as described herein. The modification can be fusions of effectors like fluorophore, proteins involved in translation modulation (e.g. eIF4E, eIF4A, and eIF4G) and proteins involved with epitranscriptomic modulation (e.g. pseudouridine synthase and m6a writer/readers), and splicing factors involved with changing splicing. Cas7-11 could also

be used for sensing RNA for diagnostic purposes. In some embodiments, the C-terminus of the Cas7-11 effector can be truncated. For example, at least 20 amino acids, at least 40 amino acids, at least 50 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 150 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 250 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Cas 13 effector. For example, up to 120 amino acids, up to 140 amino acids, up to 160 amino acids, up to 180 amino acids, up to 200 amino acids, up to 250 amino acids, up to 300 amino acids, up to 350 amino acids, or up to 400 amino acids may be truncated at the C-terminus of the Cas 15 effector.

[0262] In some embodiments, the N-terminus of the Cas7-11 effector protein may be truncated. For example, at least 20 amino acids, at least 40 amino acids, at least 50 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 150 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 250 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Cas7-11 effector. For examples, up to 120 amino acids, up to 140 amino acids, up to 160 amino acids, up to 180 amino acids, up to 200 amino acids, up to 250 amino acids, up to 300 amino acids, up to 350 amino acids, up to 400 amino acids may be truncated at the N-terminus of the Cas7-11 effector.

[0263] In some embodiments, both the N- and the C- termini of the Cas7-11 effector protein may be truncated. For example, at least 20 amino acids may be truncated at the C-terminus of the Cas7-11 effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Cas7-11 effector. For example, at least 40 amino acids may be truncated at the C-terminus of the Cas7-11 effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least

[illegible]

[illegible]

[illegible]

[illegible]

amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Cas7-11 effector.

For example, at least 200 amino acids may be truncated at the N-terminus of the Cas7-11

[illegible]

of the Cas7-11 effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Cas7-11 effector. For example, at least 350 amino acids may be truncated at the N-terminus of the Cas7-11 effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Cas7-11 effector.

[0264] The additional modifications of the CRISPR-Cas protein may or may not cause an altered functionality. By means of example, and in particular with reference to CRISPR-Cas protein, modifications which do not result in an altered functionality include for instance codon optimization for expression into a particular host, or providing the nuclease with a particular marker (*e.g.* for visualization). Modifications which may result in altered functionality may also include mutations, including point mutations, insertions, deletions, truncations (including split nucleases), etc. Fusion proteins may without limitation include for instance fusions with heterologous domains or functional domains (*e.g.* localization signals, catalytic domains, etc.). In certain embodiments, various modifications may be combined (*e.g.* a mutated nuclease which is catalytically inactive and which further is fused to a functional domain, such as for instance to induce DNA methylation or another nucleic acid modification, such as including without limitation a break (*e.g.* by a different nuclease (domain)), a mutation, a deletion, an insertion, a replacement, a ligation, a digestion, a break or a recombination). As used herein, "altered functionality" includes without limitation an altered specificity (*e.g.* altered target recognition, increased (*e.g.* "enhanced" Cas proteins) or decreased specificity, or altered PAM recognition), altered activity (*e.g.* increased or decreased catalytic activity, including catalytically inactive nucleases or nickases), and/or altered stability (*e.g.* fusions with destabilization domains). Suitable heterologous domains include without limitation a nuclease, a ligase, a repair protein, a methyltransferase, (viral) integrase, a recombinase, a transposase, an argonaute, a cytidine deaminase, a retron, a group II intron, a phosphatase, a phosphorylase, a sulfurylase, a kinase, a polymerase, an exonuclease, etc. Examples of all these modifications are known in the art. It will be

understood that a "modified" nuclease as referred to herein, and in particular a "modified" Cas or "modified" CRISPR-Cas system or complex preferably still has the capacity to interact with or bind to the polynucleic acid (*e.g.* in complex with the guide molecule). Such modified Cas protein can be combined with the deaminase protein or active domain thereof as described herein.

[0265] In certain embodiments, CRISPR-Cas protein may comprise one or more modifications resulting in enhanced activity and/or specificity, such as including mutating residues that stabilize the targeted or non-targeted strand (*e.g.* eCas9; "Rationally engineered Cas9 nucleases with improved specificity", Slaymaker *et al.* (2016), *Science*, 351(6268):84-88, incorporated herewith in its entirety by reference). In certain embodiments, the altered or modified activity of the engineered CRISPR protein comprises increased targeting efficiency or decreased off-target binding. In certain embodiments, the altered activity of the engineered CRISPR protein comprises modified cleavage activity. In certain embodiments, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In certain embodiments, the altered or modified activity of the modified nuclease comprises altered helicase kinetics. In certain embodiments, the modified nuclease comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA (in the case of a Cas protein), or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters formation of the CRISPR complex. In certain embodiments, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in certain embodiments, there is increased specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In certain embodiments, the mutations result in decreased off-target effects (*e.g.* cleavage or binding properties, activity, or kinetics), such as in case for Cas proteins for instance resulting in a lower tolerance for mismatches between target and guide RNA. Other mutations may lead to increased off-target effects (*e.g.* cleavage or binding properties, activity, or kinetics). Other mutations may lead to increased or decreased on-target effects (*e.g.* cleavage or binding properties, activity, or kinetics). In certain embodiments, the mutations result in altered (*e.g.*

increased or decreased) helicase activity, association or formation of the functional nuclease complex (*e.g.* CRISPR-Cas complex). In certain embodiments, as described above, the mutations result in an altered PAM recognition, *i.e.* a different PAM may be (in addition or in the alternative) be recognized, compared to the unmodified Cas protein. Particularly preferred mutations include positively charged residues and/or (evolutionary) conserved residues, such as conserved positively charged residues, in order to enhance specificity. In certain embodiments, such residues may be mutated to uncharged residues, such as alanine.

[0266] Type-III CRISPR-Cas Proteins

[0267] The application describes methods using Type-III CRISPR-Cas proteins. This is exemplified herein with Cas7-11, whereby a number of orthologs or homologs have been identified. It will be apparent to the skilled person that further orthologs or homologs can be identified and that any of the functionalities described herein may be engineered into other orthologs, including chimeric enzymes comprising fragments from multiple orthologs.

[0268] Computational methods of identifying novel CRISPR-Cas loci are described in EP3009511 or US2016208243 and may comprise the following steps: detecting all contigs encoding the Cas1 protein; identifying all predicted protein coding genes within 20kB of the cas1 gene; comparing the identified genes with Cas protein-specific profiles and predicting CRISPR arrays; selecting unclassified candidate CRISPR-Cas loci containing proteins larger than 500 amino acids (>500 aa); analyzing selected candidates using methods such as PSI-BLAST and HHPred to screen for known protein domains, thereby identifying novel Class 2 CRISPR-Cas loci (see also Schmakov *et al.* 2015, Mol Cell. 60(3):385-97). In addition to the above-mentioned steps, additional analysis of the candidates may be conducted by searching metagenomics databases for additional homologs. Additionally or alternatively, to expand the search to non-autonomous CRISPR-Cas systems, the same procedure can be performed with the CRISPR array used as the seed.

[0269] In one aspect the detecting all contigs encoding the Cas1 protein is performed by GenemarkS, a gene prediction program as further described in "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." John Besemer, Alexandre Lomsadze and Mark Borodovsky, Nucleic Acids Research (2001) 29, pp 2607-2618, herein incorporated by reference.

[0270] In one aspect the identifying all predicted protein coding genes is carried out by comparing the identified genes with Cas protein-specific profiles and annotating them according to NCBI Conserved Domain Database (CDD) which is a protein annotation

resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). In a further aspect, CRISPR arrays were predicted using a PILER-CR program which is a public domain software for finding CRISPR repeats as described in "PILER-CR: fast and accurate identification of CRISPR repeats," Edgar, R.C.,
5 BMC Bioinformatics, Jan 20;8:18(2007), herein incorporated by reference.

[0271] In a further aspect, the case by case analysis is performed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). PSI-BLAST derives a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein-protein BLAST. This PSSM
15 is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences. Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

[0272] In another aspect, the case by case analysis is performed using HHpred, a method for sequence database searching and structure prediction that is as easy to use as BLAST or PSI-
20 BLAST and that is at the same time much more sensitive in finding remote homologs. In fact, HHpred's sensitivity is competitive with the most powerful servers for structure prediction currently available. HHpred is the first server that is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment
25 databases, like Pfam or SMART. This greatly simplifies the list of hits to a number of sequence families instead of a clutter of single sequences. All major publicly available profile and alignment databases are available through HHpred. HHpred accepts a single query sequence or a multiple alignment as input. Within only a few minutes it returns the search results in an easy-to-read format similar to that of PSI-BLAST. Search options include local
30 or global alignment and scoring secondary structure similarity. HHpred can produce pairwise query-template sequence alignments, merged query-template multiple alignments (*e.g.* for transitive searches), as well as 3D structural models calculated by the MODELLER software from HHpred alignments.

[0273] Deactivated / inactivated Cas7-11 proteins

[0274] Where the Cas7-11 protein has nuclease activity, the Cas7-11 protein may be modified to have diminished nuclease activity *e.g.*, nuclease inactivation of at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, or 100% as compared with the wild type enzyme; or to put in another way, a Cas7-11 enzyme having advantageously about 0% of the nuclease activity of the non-mutated or wild type Cas7-11 enzyme or CRISPR-Cas protein, or no more than about 3% or about 5% or about 10% of the nuclease activity of the non-mutated or wild type Cas7-11 enzyme.

[0275] Modified Cas7-11 enzymes

[0276] In particular embodiments, it is of interest to make use of an engineered Cas7-11 protein as defined herein, such as Cas7-11, wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex, wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci, the protein comprises at least one modification compared to unmodified Cas7-11 protein, and wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified Cas7-11 protein. It is to be understood that when referring herein to CRISPR "protein," the Cas7-11 protein is an unmodified or modified CRISPR-Cas protein (*e.g.* having increased or decreased or the same (or no) enzymatic activity, such as without limitation including Cas7-11. The term "CRISPR protein" may be used interchangeably with "CRISPR-Cas protein", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein.

[0277] Computational analysis of the primary structure of Cas7-11 nucleases reveals 5 distinct domain regions.

[0278] Based on the above information, mutants can be generated which lead to inactivation of the enzyme or which modify the double strand nuclease to nickase activity. In alternative embodiments, this information is used to develop enzymes with reduced off-target effects.

[0279] In certain of the above-described Cas7-11 enzymes, the enzyme is modified by mutation of one or more residues (in the Cas7-like domains as well as the small subunit) including but not limited to positions listed in table 1.

[0280] Orthologs of Cas7-11

[0281] The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a

"homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. Homologs and orthologs may be identified by homology modelling (see, *e.g.*, Greer, Science vol. 228 (1985) 1055, and Blundell *et al.* Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):359-66. doi: 10.1002/pro.2225.). See also Shmakov *et al.* (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related, or are only partially structurally related.

[0282] The present invention encompasses the use of a Cas7-11 effector protein, derived from a Cas7-11 locus denoted as subtype III-E. Herein such effector proteins are also referred to as "Cas7-11p", *e.g.*, a Cas7-11 protein (and such effector protein or Cas7-11 protein or protein derived from a Cas7-11 locus is also called "CRISPR-Cas protein").

[0283] In particular embodiments, the effector protein is a Cas7-11 effector protein from an organism from a genus comprising *Candidatus Jettenia caeni*, *Candidatus Scalindua brodae*, *Desulfobacteraceae*, *Candidatus Magnetomorum*, *Desulfonema Ishimotonii*, *Candidatus Brocadia*, *Deltaproteobacteria*, *Syntrophorhabdaceae*, or *Nitrospirae*.

[0284] Example Cas proteins

[0285] In certain example embodiments, the CRISPR effector protein is a Cas7-11 type III-D/III-E ortholog selected from Table 1 and FIG. 27A-27C.

Table 1.

SEQ ID NO & Protein ID/Contig	Sequence
SEQ ID NO: 1 WP_007220849	MHTILPIHLTFLEPYRLAEWHAKADRKKNKRYLRGMSFAQWHKDKDGIGKPYITGTL LRS AVLNA AEELISLNQGMWAKEPCCNGKFETEDKPAVLRKRPTIQWKTGRPAICD PEKQEKD DACPLCMLLGRFDKAGKRHRDNKYDKHDYDIHFDNLNLITDKKFSHPDDI ASERILNRVDYTTGKAHDYFKVWEVDDDDQWWQFTGTITMHDDCSKAKGLLLASLCF VDKLCGALCRIEVTGNNSQDENKEYAHPDTGIITSLNLKYQNNSTIHQDAVPLSGSAH DNDEPPVHDNDSSLDNDTITLLSMKAKEIVGAFRESGKIEKARTLADVIRAMRLQKPD IWEKLPKGINDKHHLWDREVNGKKLRNILEELWRLMNKRNAWRTFCEVLGNELYRC YKEKTGGIVLRFRTLGETEYYPEPEKTEPCLISDNSIPITPLGGVKEWIIIGRLKAETPFY FGVQSSFDSTQDDLDLVPDIVNTDEKLEANEQTSFRILMDKKGRYRIPRSLIRGVLRD LRTAFGGSGCIVELGRMIPCDCKVCAIMRKITVMDSRSENIELPDYRIRLNPYTATV DEGALFDMEIGPEGITFPFVFRYRGEDALPRELWSVIRYWMGMALWGGSGSTGKG RFALIDIKVFEWDL CNEEGLKAYICSRGLRGIEKEVLENKTIAEITNLFKTEEVKFFES YSKHIKQLCHECHNQISFLWGLRSYYEYLGPLWTEVKYEIKIASPLLSSDTISALLNKD NIDCIAYEKRKWEENGKIFVPTIKGETIRGIVRMAVGKRSGLGMDDHEDCSCTLCTI

SEQ ID NO & Protein ID/Contig	Sequence
	FGNEHEAGKLRFEDELEVVEEKLPSEQNSDSNKPFGPVPQDGDGNREKECVTAVKSYK KKLIDHVAIDRFHGAEDKMKFNTLPLAGSFEPKILKGRFWIKKDIVKDYKKKIEDA MVDIRDGLYPIGKGTGIGYGWVTDLTILNPQSGFQIPVKKDISPEPGTYSTYPSHSTPSL NKGHIYPHYFLAPANTVHREQEMIGHEQFHKEQKGELLVSGKIVCTLKTVTPLIIPDT ENEDAFGLQNTYSGHKNYQFFHINDEIMVPGSEIRGMISSVYEAITNSCFRVYDETKYI TRRLSPEKKDESNDKNKSQDDASQKIRKGLVKKTDEGFSIIEVERYSMKTKGGTKLV DKVYRLPLYDSEAVIASIQFEQYGEKNEKRNAKIRAAIKRNEVIAEVARKNLIFLRSLT PEELKKVLQGEILVKFSLKSGKNPNDYLAELHENGTERGLIKFTGLNMVNINKNVNEED KDFNDTWDWEKLNIFHNAHEKRNLSLKQGYPRPVLKFIKDRVEYTIPKRCERIFCIPVK NTIEYKVSSKVCKQYKDVLSDEYKNFNGHINKIFTTKIQKRELTDGDLVYFIPNEGADK TVQAIMPVPLSRITDSRTLGERLPHKNLLPCVHEVNEGLLSGILDSLDKLLSHIPEGLC PTCRLFGTTYKGRVRFGFANLMNPKWLTERENGCGGYVTLPPLERPRLTWSVPSD KCDVPGRKFYIHNGWQEVLRNNDITPKTENNRTVEPLAADNRFTFDVYFENLREWE LGLLCYCLELEPGMGHKLGMGKPMGFGSVKIAIERLQFTTVHQDGINWKPSENEIGV YVQKGREKLVEWFTPSAPHKNMENWNGVHKIKDLRSLLSIPGDKPTVKYPTLNKDAE GAISDYTYERLSDTKLLPHDKRVEYLRTPWSPWNAFVKEAEYSPSEKSDEKGRETIRT KPKSLPSVKSIGKVKWFDEGKGFGILIMDDGKEVSISKNSIRGNILLKKKGQKVTFHIVQ GLIPKAEDIEIAK
SEQ ID NO: 2 KHE91659	MNITVELTFFEPYRLVEWFDWDARKKSHSAMRQQAFAQWTWKGKGRTAGKSFITGT LVRS AVIKAVEELLSLNGKWEVGPCCNGSFQTDSEKGGKPSFLRKRHTLQWQANN KNICDKEEACPFICILLGRFDNAGKVHERNKDYDIHFSNFDLDHKQEKNDLRLVDIASG RILNRVDFDTGKAKDYFRTWEADYETYGTYTGRITLRNEHAKKLLLASLGFVDKLCG ALCRIEVIKSESPLPSDTKEQSYTKDDTVEVLSEDHNDDELKQAEVIVEAFKQNDKL EKIRILADAIRTLRLHGEGVIEKDELDPGKEERDKGHHLWDIKVQGTALRTKLKELWQ SNKDIGWRKFTEMLGSNLYLIYKKETGGVSTRFRILGDTEYYSKAHDSEGSDFIPVTP PEGIETKEWIIVGRLLKAATPFYFGVQQPSDSIPGKEKKSEDSLVINEHTSFNILLDKENR YRIPRSALRGALRRDLRTAFGSGCNVSLGGQILCNCKVCIEMRRITLKDSVSDFSEPPEI RYRIAKNPGTATVEDGSLFDIEVGPEGLTFPFVLRVYRGHKFPEQLSSVIRYWEENDGK NGMAWLGGLDSTGKGRFALKDIKIFEWDLNQKINEYIKERGMRGKEKELLEMGESSL PDGLIPYKFFEERECLFPYKENLKPQWSEVQYTIEVGSPLL TADTISALTEPGNRDAIAY KKRVYNDGNNAIEPEPRFAVKSETHRGIFRTAVGRRTGDLGKEDHEDCTCDMCIIFGN HESSKIRFEDLELINGNEFEKLEKHIDHVAIDRFTGGALDKAKFDYPLTAGSPKKPLK LKGRFWIKKGFGSDHKLLITTALSDIRDGLYPLGSKGGVGVWAGISIDNVPDDFK EMINKTEMLPEEVEESNNGPINNDYVHPGHQSPKQDHKNKNIYYPHYFLDSGSKVY REKDIITHEEFTEELLSGKINCKLETLTPLIIPDTSDENGLKLQGNKPGHKNYKFFNING ELMIPGSELRGMLRTHFEALTKSCFAIFGEDSTLSWRMNADEKDYKIDNSIRKMESE RNPKYRIPDELQKELRNSGNLGNLYTSERRFWSDVSNKFENSIDYKREILRCAGRP KNYKGGIIRQRKDSLMAEELKVHRLPLYDNFDIPDSAYKANDHCRKSATCSTSRGCR ERFTCGIKVRDKNRVFLNAANNRQYLNNIKKSNHDLYLQYLKGEKKIRFNSKVITG SERSPIDVIAELNERGRQTGFIKLSGLNNSNKSQGNTGTTFNSGWDRFELNILLDDLET RPSKSDYPRPRLFTKQYENITKRCERVFEIDKGNKTGYPVDDQIKKNYEDILDSY DGIKDQEVAFERFDTFTRGSKLKVGDLYFHIDGDNKIDSLIPVIRSRKASKTLGGKLD KALHPCTGLSDGLCPGCHLFGTTDYKGRVKFGFAKYENGPEWLITRGNNPERSLTG VLESPRPAFSPDDESEIPGRKFYLLHNGWRIIRQKQLEIRETVQPERNVTTTEVMDKGN VFSFDVRFENLREWELGLLLQSLDPGKNIAHKLKGGKPYGFGSVKIKIDSLHTFKINSN NDKIKRVPQSDIREYINKGYQKLIWESGNNISIQKGNVLPQWHVIPHIDKLYKLLWVPF LNDKLEPDVRYVPLNEESKGYIEGSDYTYKKLGDKNLPHYKTRVKGLTTPWSPWNP FQVIAEHHEQEVNVTGSRPSVTDKIERDGKMGV
SEQ ID NO: 3 OQY58162	MKITLRFLEPFRMLDWIRPEERISGNKAFQRGLTFARWHKSKADDKGKPFITGTLRS AVIRAAEHLLVLSKGVGEKACCPGKFLTETDTETNKAPTMMFLRKRPTLKWTRDKGC DPDFPCPLCELLGPGAVGKKEGEAGINSYVNFNLSFPGDTGYSNAREIAVRRVVNRV DYASGKAHDDFFRIFEVDHIAFPFCFHGEIAFGENVSSQARNLLQDSLRFDRLCGALCVI RYDGDIPKCGKTAPLPETESIQNAAEETARAIVRVFHGGRKDPEQAQIDKAEQIQLLSA AVRELGRDKKKVSALPLNHEGKEDHYLWDKKAGGETIRTILKAAAEKEAVANQWR QFCIELSEELYKEAKKAHGGLEPARRIMGDAEFSDKSVPDVTSHSIGISVEKETIIMGTL KAETPFFFGIESKEKKQTDMLLLDGGQNHYRIPRSALRGILRRDIRSVLGTGCNAEVGG RPCLCPVCRIMKNITVMDTRSSTDTLPEVRPRIRLNPFITGSGVQEKALFNMEMGTGIEF

SEQ ID NO & Protein ID/Contig	Sequence
	PFVLSYRGKKTLPKELRNVLNWWTEGKAFLGGAASTGKSIFQLSDIHAFSSDLSDETA RESYLSNHGWRGIMENSIVHESPLEGGAGGCSFGLSDLPKLGWHAEDLKLSDIEKYKP FHRQKISVKITLNSPFLNGDPVRALTEDVADIVSFKKYTQGGEEKIYAYKSESFRGVVR TALGLRNQGNDDITGKKNVPLIALTHQDCECMLCRFFGSEYEAGRLYFEDLTFESEPE PRRFDHVAIDRFTGGAVNQKKFDDRSLVPGKEGFMTLIGCFWMRKDKELSRNEIEEL GKAFADIRDGLYPLGAKGSMGYGQVAELSIVDDEDSDDENNPALLAESMKNASPSL GTPTSLKKKDAGLSLRFDENADYYPYFLEPEKSVHRDPVPPGHEEAFRGGLLTGRIT CRLTVRTP LIVPNTETDDAFNMKEKAGKKKDAYHKS YRFFTLNRVPMIPGSEIRGMIS SVFEALSNSCFRIFDEKYRLSWRMDADVKELEQFKPGRVADDGKRIEEMKEIRYPFYD RTYPERNAQNGYFRWDARISLTDNSMRKMEKDGVPNRNVIYKLN TLKNKAYKSEKSF LFDLKNKAGGVGRYKKLVLKHA EVRGGEIPYYSHTPTDCKLLSLVGP NRQLCRQDT LVQYRIIKHRRGAKPEEDFMFVGTPSENQKGHKENNDHGGGYLKISGPNKIEKENVLT SGVPSVPENMGAVVHNCPPRLVEVTVRCGRKQEECKRKRLVPEYVCADPEKKVTY TMTKR CERIFLEKSRRIPFTND AVDKFEILVKEYRRNAEQQDTPEAFQTILPENGTVNP GDLLYFEEKGKAAEIVPVIRSRKVDDRHI GKRIDPELRPCHGEWIEDGDL SKLDAYP AEKKLLTRHPKGLCPACRVFGTGSYKSRVRFGAALKGTPKWLKEDPAEPSQKGKIT LPLLERPRPTWAVLHNDKENSEIPGRKFYVHHNGWK GISEGIHPISGENIEPDENNR TV EVLDKGNRFVFE LSFENLEPRELG LLIHSLQLEKGLAHKLGMAKSMGFGSVEIDVESV RVKHRSGEWDYKDGETVDGWIEEGKRGVAAKGKANDLRKLLYLPG EKQNPVHYP TLKKEKKGDPPGYEDLKKSFREKKLNRRKMLTTLWEPWHK
SEQ ID NO: 4 KPA14974	MLKLKVKITYFQPFVRIPWIKEDDRNSDRNYLRGGTFARWHKDKDDIHKPYITGT LLRSALFTEIEKIKIHHSDFIHCCNAIDRTEGKHQPSFLRKRVPY TENKNIQACNKCPLC LIMGRGDDRGEDLKKKKHYNKGHYQNWTVHFSNFDTQATFYWKDIVQKRILNRVD QTCGKAKDFFKVCEVDHIACPTLNGIIRINDEKLSQEEISKIKQLIAVGLAQIESLAGGIC RIDITNQNHDDLKISFFETKPSKILQPNLKESGEERFELAKLELLAEYLTQSFDANQKEQ QLRRLADAIRDLRKYSPDYLDLPKGKKGGRTSIWNKKVADDFTLRDCLKNQKIPNE LWRQFCEGLGREVYKISKNISNRSDAKPRLLGETEYAGLPLRKEDEKEYSPTYQNQES LPKTKWIISGELQAITPFYIGHVNKTSHTSTIFLNMNGQFCIPRSTLRGALRRDLRLVF GDSCNTPVGSRV CYCQVCQIMRCIKFEDALSDVDSPEVRHRIRLNCHTGVVEEGALF DMETGFGQMIFPFRLYYESKNEIMSQHL YEVLNNWNTNGQAFFGGEAGTGFRFKLLN NEVFLWEIDGEEEDYLQYLF SRGYKGIETDEIKKVADPIKWKLTFTKLEIPPEKIPLTQL NYTLTIDSPLISRDP IAAMLNRRNPDAVMVKKTILVYEQDSSTHKNQVILFEPK YFIKSE TIRGLLSIISRTEIKLEDGKKERIFNLDHEDCDCLQCRLFGNVHQQGILRFEDA EITNK NVSDCCIDHVAIDRFTGGGVEKMKFNDYPLSASPKNCLNLKGSWITSALKDSEKEAL SKALSELKYG YASLGGLSAIGYGRVKELTLEENDIIQLTEITESNLNSQSRLSLKPDVK KELSNNHFYYPHYFIKPAPEVVRESRLISHVQGHDTGEFLLTGKIKCRLQTLGPLFI ANNDKGDDYFELQHNNPGHLNYAFFRINDHIAIPGASIRGMISVFTLTHSCFRVMD DKKYLTRRVIPESETTQKRKSGRYQVEESDPDLFPGRVQKKGNKYKIEKMDEIVRLPI YDNFSLVERIREYHYSEECASYVPSVKK AIDYNRMLAQAAADS NREFLYNHPEAKSILQ GKKEVYI LHKQESKNRGKTKEINPNARYACLT DENTPGSRKGFIKFTGPDMMVTVNK ELKSKIAPIYDPEWEKDIPDWERSNQESNHKYSFILHNEIEMRSSQKKKYPRPVFICKK NGVEYRMQKRCERIFDFTKEEEKDKEI VIPQKVVSQYNAILKDNKENTETIPGLFNSK MVNKELEDGDLVYFKYKEGKVTELTPVAISRKT DNKPMGKRFPKISINGKMKPNDSL RSCSHTCTEDCDDCPNLCE SVKDYFKPHPDGLCPACHLFGTTFYKSRLSFG LAWLEN NAKWYISNDFQQKDSKKEKGKGLTLPLLERPRPTWSPMPNNAEVPGRKFYVHHPWS VENIKNNQGNQKDISLKPDSDAIKIKENNR TIEPLGKDNVFNFEISFNLRD WELGLLL Y AIELEDHLAHKLGMAKAFMGMSVKIEIKNLLIKGSINDISK AELIKKGFKKL GIDSLE KDDLSEYLHIKQLREILWFS DKPVGTIEYPKLENK TNSRIPSYTDFVQEKDHETGFKNP KYQNLKSRLHILQNPWNAWWKNEE
SEQ ID NO: 5 WP_12432758 9	MTTMMKISIEFLEPFMRMTKWQESTRRNKNKEFVRGQAFARWHRNKKDNTKGRPYIT GTLLRSAVIRSAENLLTSDGKISEKTCCPGKFDTE DKDRLLQLRQRSTLRWTDKNPC PDNAETYCPFCCELLGRSGNDGKKA EKKDWRFRIFHGNLSLPGKPDFDGPKAIGSQRV LNRVDFKSGKAHDFFKAYEVDHTRFPRFEGEITIDNKVSAEARKLLCDSLKFTDRLCG ALCVIRFDEYTPAADSGKQTENVQAEPNANLAEKTAEQIISILDDNKKTEYTRLLADAI RSLRRSSKL VAGLPKDHGDKDDHYLWDIGKKKKDENSVTIRQILTT SADTKELKNAG KWREFCEKLGEALYLKSKDMSGGLKITRRILGDAEFHGKPDRL EKSRSVSIGSVLKET VVCGLVAKTPFFFGAIDEDAKQTDLQVLLTPDNKYRLPRSAVRGILRRDLQTYF DSP

SEQ ID NO & Protein ID/Contig	Sequence
	CNAELGGRPCMCKTCRIMRGITVMDARSEYNAPPEIRHRTRINPFTGTVAEGALFNME VAPEGIVFPFQLRYRGSEGLPDALKTVLKWWAEGQAFMSGAASTGKGRFRMENAK YETLSDENQRNDYLKNWGWWRDEKGLEELKKRLNSGLPEPGNYRDPKWHEINVSI EMASPFINGDPIRAAVDKRGTDVVTFVKYKAEGEEAKPVCAYKAESFRGVIRSAVARI HMEDGVPLTELTHSDCECLLCQIFGSEYEAGKIRFEDLVFESDPEPVTFDHVAIDRFTG GAADKKKFDDSPSPARPLMLKGSFWIRRDVLEDEEYCKALGKALADVNNGLYP LGGKSAIGYGQVKSGLIKGDDKRISRLMNPADFETDVAVPEPKTDAEVRIEAEKVY YPHYFVEPHKKVEREEKPCGHQKFHEGRLTGKIRCKLITKTPLIVPDTSNDDFFRPADK EARKEKDEYHKSYPFFRLHKKQIMIPGSELRGMVSSVYETVTNSCFRIFDETKRLSWRM DADHQNVLQDFLPGRVTADGKHQKFSETARVPFYDKTQKHFDILDEQEIAGEKPV MWVKRFIKRLSLVDPAKHPQKKQDNKWKRRKEGIATFIEQKNGSYFNVVTNNGCT SFHLWHKPDNFDQEKLEGIQNGEKLDCWVRDSRYQKAFQEIPENDPDGWECKEGL HVVGPSKVEFSDKKGDVINNFQGTLPSPNDWKTIRTNDFKNRKRKNPVCCEDDK GNYYTMAKYCETFFFDLKENEYEIPEKARIKYKELLRVYNNNPQAVPESVFQSRVA RENVEKLKSGDLVYFKHNEKYVEDIVPVRISRTVDDRMIGKRMSADLRPCHGDWVE DGDLSALNAYPEKRLLLRHPKGLCPACRLFGTGSYKGRVRFGFASLENDPEWLIPGK NPGDPFHGGPVMLSLLEPRPTWSIPGSDNKFVPGKIFYVHHAWKTIKDGHNPTT GKAIEQSPNNRTVEALAGGNSFSFEIAFENLKEWELGLLIHSLQLEKGLAHKLGMAKS MGFGSVEIDVESVRLRKDWKQWRNGNSEIPNWLKGKFAKLKEWFRDELDFIENLKK LLWFPEGDQAPRVCPMLRKKDDPNGNSGYEELKDGEFKKEDRQKKLTTPWTPWA
SEQ ID NO: 6 KKO18793	MSKTDDKIDIKLTFLEPYRMVNWLENGLRMTDPRYLRLGSFARWHRNKGKAGRPY ITGTLLRSAVIRAAEELLNLGKWKQLCCPGQFETEREMRKNKTFLLRRPTPAWSA ETKKEICTTHGSACAFCLLLGRRLHGGKEDVNEDAPGSCRKPVGFGLNSLPFOPTKRQ IQDVCKERVNLNRVDFRTGKAQDYFRVFEIDHEDWGVYTGEITITEPRVQEMLEASLKF VDTLCGALCRIEIVGSADETKRTTSSKEGCPASTTTTRDCSSSENDTSPEDPVREDLKK IAHVIANAFQNSGNREKVHALADAIRAMRLEESSIINTLPKGKSEKTTEQIEVNKHLYL WDEIPVNDTSVRHILIEQWRRWQSKKDDPEWWKFCDFLGECLYKEYKKLTSGIQSRA RVMGETEYYGALGMPDKVIPLKSDKTKEWILVGLSKAETPFFFGLETEQTEVEHTS LRLVMDKKGRFIRPSVLRGALRRDMRIAFAFDGCDVKLGSPPCDCSVCQVMRSITIK DSRSEAGKLPIRHRIRLNPFSGTVDEGALFDIEVAPEGVIFPFVMRYRGEEFPALLSV IRYWQDGKAWLGEGATGKGRFALAKDLKMYEWKLEDKSLHAYIDTYGHRGNEH AIGTGQIDGFRSGSLSDLLSDISKESFRDPLASYHNYLDRWKVGYQITIGAPLLSAD PIGALLDPNNVD AIVFEKMKLDGDQVKYLP AIAKGETIRGIVRTALGKRNNLLAKNDH DDCTCSLCAIFGNENETGKIRFEDLEVYDKDIAKKIDHVAIDRFTGGARDQMFKDTLP LIGSPERPLRLKGLFWMRDVPDEKARILLAFLEIREGLYPIGGKTGSGYGWVSDLEF DGDAPFAFKEMNSKRKGQASFKKISFRYPSGAPKHIQNLKATSFYYPHYFLEPGSKV IREQKMIGHEQYYESYPSGASGEKLLSGRIICSMTHHTPLIVPDTGVKIDPENKHATYD FFQMNNAIMIPGSEIRGMISAVYEAMTNSCFRIFHEKQYLTRRISPDKELREFIPGIVRI INGDVYIEKAEREYRLPLYDDVHIITNYEELEYEKYIKKNPGREQKIKNAHRFNKNIA IAESNRNYLCSLDRAVRREILSGRKKVNFRLVKVNNDKNPDKEAELCKTGPLEGLV KFSGLNAVNISNLRPGTAEEGFDKWDWWSLNLNRMDVRNSQKKEYPRPALHFNH DGKEYTIPKRCERVVRAEAGKRAETEGSYKVPKRVQEYQYQNLRDYESNIGHIDNTF RTLIENCGLNNGSLVYFKPDNSRKEVVAITPVKISRKTDRLPQGDRFPHTSSDLRPCVR DCLDTEGDIRMLENSPFKRLFHIHPEGLCPACQLFGTTNYRGRVRFGFASLSDGPKWF RKDEGNETCHITLPLLERPRPTWSMPDDTSTIPGRKIFYVHHMGYETVKKNQRTL VKT ENNRTVKALDKENEFTFEVFFENLREWELGLLLHCLELEPEMGHKLGMGKPLGFGSV KIRIDKLQKCVNVKDGCVLWEPEEDKIQHYIAKGLGKLTTFWFGKEWDRLEHIQGLR SLQRLPL
SEQ ID NO: 7 RLC14096B	FESYARWCKSNSGLWKPYIPGTLLRSAVLESVEYLLALIGSKNKVEICPLYTQSENN PDTKYLRRRRPWYELHAQKEICKTRDTACPLCLLMRTKLDNDGDGETEKNVKFGNLY PTSPLEPLQIRPRILNRMDPGTSKARDYFRVFEIENQLCSQFRGIWLSGDLPMNELV KSLLAAGLSNVATLAGAVCRIRIVSTDNPSMKQDLTTQDLIDDFTNYYLKGDTPPANL AASGKGDAFPRFSPSGDHPDTTGVSHADMASSHEGTALAKDIAEKCKDILSQISASE QLRRLADIMRDLRQDSNREIMYRQVAEENHEKASLLYKKTGKGSIAALIAGKTEGM DAETWRRRLCEFLGQTFYGEAKEAGLVETPVPRILGESERYSLQKKPTVRTDLAELVP DIEFIKGNLIAETPFFFGTDIATETHDLPILLTSDRHFIRPSVLRGILRRDLRLVTGSG

SEQ ID NO & Protein ID/Contig	Sequence
	CSVKLGRSEPCACDVCQIMRSLTMRDCVSSCKVPPEIRHRIRLNPVTETVEEGALFDM EIGPQGISFPFVLRSGVNSSFSTRLKNVLTWWSEGKIFMGGDGKTGKGRFTLAELEA YYFRLTTKRIGKNVWVIGNYLKSQGWRGAELETHFDSLKEWKSLSFSDSDVKVFTW HKITWKVSFEGPVL TNDPIAADIRNESDAVFYQKSVAGEKGPVYALKGEGLRGIVSSS LCKKKNLSSNLHEDCECLRCKIFGSKHQEGNIRFEDMTVSQESEVREKLFDPHVSIDRFT GGAANKLKFDKPLVGNPLVFQGVFWVHQSIGNNEKTQEALSDAFKDVRDGLYPVG AKGSIGYGWIKGIEVVEGPDWLKD ALSAEKTVEAGIASSESEYKLPDLPWISLLPKGR AIYNPHYFLGIPKVTPEREREVGHDRFQTDLHTGRIICTLKTITPLIIPDTENDKAFEVE NASADHERFKFMRMGSQAAIPGSAIRSMTSSVFEAL TNSCFRVLDQKSHLSWRMEAD DAGDYKPGRFEEKDDKA VIRKFKKKARFPFYAGPDTREAF TSDQIMGKEKVTLWVK DFEASLTVPDEIGWKKRGYLVKTGPNKVEIDTENISENNPSPDSDWQDVRINDDGTIP DKKNRKFICQYGTTTYTVDKWCEAFFCDEEKDPYELAPDVERKYRLLMDSYHNNPQ APPQIFRSLPLFSETGPKKTLEHGDLVYFRLSEVNKQSQSKKQVRERVTDIVPVSISRIA NNQPIGKHIAAAFRPCA YVCIEECPD AKTCPIPVYREGYPIKGLCRACHLFGTTGYK GRVRFSAKLNGDAVWAKGAGGKDYFTLPLEKPRPTWTMPNEGAKIPGRKFYVHH NEWKTVQEGKNPIDQKAIRPNPNSSVEVLNLGNEFQFEVSFENLEEWELGLLLYCLE LEPGLAHLKGRGKAFGFGSIEAEVSKIEMRIKSGTWKNETSGKEKFIQSGLSQVPSFFK QDEKQWNKVEQVKNIRKLLQLSWNKGNAVEPEVRYPALREKDDENKRPGYVELKD NGYDAGKKLVSPWAPWHPIK
SEQ ID NO: 8 OGR07205	MTKKPGTEDKATLWGKESASKSVKTILEESIQQFTVEQKRSFFANLADQLVSRAGEQ GAKSVRSQGLIIGRKENYAKPSAQEPTRHHL YRQPSNASAFLATGWLIAETPFFIGSGT EGQKQTDQAEHLRLTLRDGHGRFRIPFTTIRGVMDELRLDQAGCAKGRSLRAPC PCQVCTLMRRIQVRDAIAADILPPDLRMTRIDPSHGTVAHLFSLEMAPQGLKLPFFLK LKGVTIDPDKELLEILNDWSAGQCFLGGLWGTGKGRFRLDDLQWHRLELDNADYY TPLLQDRFFAGETISDLRQGLQSINIQPERIPAQTPSRNMPYCRVDCILEFKSPVLSGDP VAALFESDAPDNVAYKKPVVQYDETGRRLRTDGPVEMLTCLKGEGVRGVVAYLAG KAYDQHDLSHDSNCTFCQAFNGQKAGSLRFDDFMPVQFESDQAGNFSWSPHTPH AMRSDRVALDVFGGAMPEAKFDDRPLAASPGKPLNFKSTIWYREDMGKEAGKALKR ALIDLQNNMAAIGSGGGIGRGWVSRVCFEGDIPDFLEDFPEPITVTEPEQDSQLLKNQA VADETAVSACDTADAPHPLAVTLEPGARYFPRVIIPRAPTVKRDECVTGQRYHTGRLS GKIFCELNTLGLFVPD TDYSAGVPVPISDEQLAECQLQAVFENTSKFNEFFATYPEET VTKLKDLLCAADDKWILAVKDITADLRQEIGEDTFQRIIRKAGHKTRFHFQINDEIGLP GASLRGMVLSNYQILTNSCYRNLKATEEITRRMPADEAKYRKAGRVTVSGDGAQKK YSIQEMEVLRLPIYDNMNTPDNMPDVAKQATTAKRCNNLMNEAAKTSRVELKARW REGQSKIKYQIIDALNKVDPIIQVISSSKQINPNNGKTGWGYVKYTGANVFAKSLVAPI DCLRKKDAGHVCCQVNLNPAWEASNFDILINEKCPVERQSGPRPTLRCKGQDSAWY TLTKRSEIRFTDKKPVDPINIPPREVKRYNELRDSYKKNTAHVPKPLQTFNQNESLAN GDLVYFEVNQFGEASQLTPVISIRTTDLFPIGGRLPQGHKDLFPCTAMCLSECKNCVP ASFCEHRSRHEKLC PACSLAGTTGNRGRIKFSEAWLSGLPKWHSVSQDNVGRGLGV TMPRLERSRRTWHLPTKDAYLLGQSIYLNHPVPAILPSDQVPSENNQTVEPLGPKNIFS FQLAFDNL SIEELGLLLYSLELESMAHRLGRGRALGMGVSQISVKDIQIRDNKSFLFS SNISKKSEWICGKDEFAQEA WFGESWDNIDHIQRLRQAL TIPVKGDVGCIRYPKLEA EGGMPDYIKLRKRLTPLCDREEPVRYRINPVQLARMILPFVPWHGACPALLNEQVMIE AKRLTELXXXDRANWPC
SEQ ID NO: 9 RLC14096	ASEDDDTPTLRKVLKDEINGQEDMWRKFCEALGNSLYDLSKKAKERKRTEALPRLLG ETEIYGLPMRENKEDEPLPSSLT YKFKWLIAGELRAETPFFFGTEVQEGQTSATILLNR DGYFRLPRSVIRGALRRDLRLVMGNDGCNMPIGGQMCECGVCRVMRHIVIEDGLSDC KIPPEVRHRIRLNGHTGTVEEGALFDMETGYQGMTFPFRLYCETENSDDL SYLWEVL NNWQNGQSLFGDGTGTGFRFELTEPKVFLWNFSKKEKHEAYLLNRGFGKGMPVQD VKTKSFKTKTWFIHRELDISPKKLPWYSTDYRFNVTSP LISRPDPIGAMLDPRNTDAIM VRKTVFCPDNNAKNRPAPATVYMIKGESIRGILRSIVVRNEELYDTHEDCDILCRLF GSIHQGSLRFEDA EVQNSVSDKKMDHVAIDRFTGGGVDQMKFDDYPLPGCPAQPLI LEGKFWVKDDIDDESKSALEKAFADFRDGLVSLGGLGAIGYGQIGDFELIGGSADWL NLPKPEENRTDVPCGDRSAQGPEIKISLDADKIYHPHFFLKPSDKNVYRERELVSHAK KKGPDGKSLFTGKITCRLSTEGPVFIPDIDLGEDYFEMQASHKKHKNYGFFRINGNVA IPGSSIRGMISSVFEAL TNSCFRVFDQERYLSRSEKPDPTLTKYYPGKVKRDGNKFFIL KMKDFFRLPLYDFDFEGEAESLRPNYDEDRNEEENKGNKNTQKVKNAVEFNKMA

SEQ ID NO & Protein ID/Contig	Sequence
	GFAKHNRDFLKKYKEQEIKDIFMGKKKVYFTAGKHKPNEAHDNDKIALLTGKSNKK AEKGYFKFTGPGMVNVKAGVEGEECDFHIDESDPDVYWNMSSILPHNQIKWRPSQK KEYPRPVLKCVKDGTEYVMLKRSEHVF AEASSEDSPVPGKVRKQFNSISRDNVQNT DHLSSMFQSRRLHDELSHGDLVYFRHDEKRVTDIAYVRVSRVDDRPMGKRFKNE SLRPCNHVCVEGCDECPDRCKELEDYFSPHPEGLCPACHLFGTTDYKGRVSFGLGWH ESNTPKWYMPEDNSQKGSHTLPLLERPRPTWSMPNKKSEIPGRKFYVHHPWSVDKI RNRQFDPAKEKQPDDEVKPNENNRTVEPLGKGNEFTFEVRFNNLREWELGLLYSLEL EDNMAHKLGMGKALGMGSARIKAEAIELRCESAGQNAELKDKAAAFVRKGFEFLEID KPGENDPMNFDHIRQLRELLWFLPENVSANVRYPMLEKEDDGTGPGYTDIFIKQEEPST GKRNPSTLSSEKRRNILQTPWKHWYLIPPFQASAQSETVFEGTVKWFDDKKGFGFIKI NDGGKDVVFVHHSSIVGTGFKSLNEGDSVAFKMGVGPKGPCAEKVKKIGN
SEQ ID NO: 10 OPY65763	MRRQRLLGDAEYYGGTGREQPASIVISTSDPDHKKVYEWIITGQLKAETGFFFGTKAG AGGHTDLSILLGKDGHYRVPRSVFRGALRRDLRVAFGAGCRVEVGRERPCECPVCKV MRQITVMDTISSYREAPEIRQRIRLNPYTGTVDKALFDMEVGPEGIEFPFVLRFRGSK SFPSELA AVIGSWTKGTAWLGGAATGKGRFSLLGLSIHKWNLSTAEGRKSYLEAAYG LRDAADKTVKRLSIDKGGKGDVGLPAGLERDALPSSVREPLWKKLVCTVDFSSPLLL ADPIAALLGVEGDERIGFDNIAIEKRRYNGE'TNNTESIPAVKGETFRGIVRTALGKRHG NLTRDHEDCRCLCAVFGKEQEAGKIRFEDLMPVGAWTRKHLHDVAIDRFHGGAAE NMKFDTYALAASTNPLRMKGLIWVRSDFLFTGHDGTPPYVKDIIDALADVKRGLY PVGGKTGSGYGWIKDVTIDGLPQGLSLPPAEERVDGVNEVPPYNYSAPPDLPSAAEGE YFFPHVFIKPYDKVDRVSRLTGHDRFRQGRITGRITCTLKTTLPLIIPDSEGIQTDATGH KMCKFFSVAGKPMIPGSEIRGMISVYEALTNCFRVSFDEEKYLTRRVQPKKGAKSSE LVPGIIVWGQNGGLAVQQVKNAYRVPLYDDPAVTSAPTEAQKNKERWESVPSVNL QGALDWNLTANIARDNRTFLNSRPEEKDAILSGTKPISFELEGTNPNDMLVRLVPDG VDGAGHSGYLKFTGLNMVLKANKKTSRKLAPSEEDVRTLAILHNDFDSRRDWRRPPNS QRYFPRSVLRFSLERSTYTIPKRCERVFEGTGCEPYSVPSDVERQYNSIIDISKNYGRIS ETYLTKTANRKLTVGDLVYFIADLDKNMATHILPVFISRISDEKPLGELLPFSGKLIPCE GEPITLKKMAPSLLTEAWRTLSTHLEGFCPACRLFGTTSYKGRIRFGFAEHTGTPKW LREELDWARPFLTPIQERPRPTWSVPDDKSEVPGRKFYLHHHGGNRIVESNLNRPE VNQTKNNSVPEISAGNTFTFDVCFENLEAWELGLLLYCLELSPKLAHKLGRAKAFGF GSVKIHVERIEERTTDGAYQDVTAVKKNGWITTGHDKLREWFHRDDWEDVDHIRNL RTVLRFPDADQEHDRVPELKANNGVSGYVELRDKMTASERQESLRTPWYRWFQPN GTGSGRHEQAATSQEQDTAKDESVLATQRRQAVIDVSDPDERLSGTVESFDRQKG DGYIGCGVRQFYVRLEDIRSRTALCEGQVVTFRARKEWEGHEAYDVEIDQ
SEQ ID NO: 11 RLC02083	MLEKALADFRDGVVSLGGLGAIGYGRIGDFEVAEESGTWLKIPEKKLPEDSVQCGER YRFSSDPATRFEEKEKIYYPHYFLKPSDEDVRRETRLVSHVYQEDTDGKTRLLTGTIRC RLTTGEPFIPDTPDPKEDYFQMEIEGHKSYGFFRINEQVAIPGSSIRGMVSSVFEALTN SCFRVFDQKRYLSRSTKPDPRELEKYLPGKVKRIDNKWVLELEDIFRLPLYDLKDVG PKSLDSAYGLEKFKNEKRFLKKIENAVAFNKKMAGYAKHTREFLNKNYTTETELGKI LRGEMK VWF TIGHKPN SAHDNDKIALLT KKT NKRKSGFIKFTGPSMVNIKADASNS DCEECRFDMKSEDKDGLIFHNAIECRPSQKKEYPRPVLKCVKEAVEYTMIKRCEQVFS EGKKPPRSYSIPDKSRRQYNGILKDNDRDNTTEHIPSFFRNRMKNELSDEDLVYFRYKG KKVTNIAPVRVSR
SEQ ID NO: 12 RLC02082	QYNLPLNPDAFPKFRWIVTGHLRAETPFFFGRGEIKDRTIEEETEQTSTILLNKDNFFR LPRSVIRGALRRDLRLVIGNGCNTPVGGKFCECDVCRIMRHVVVEDTISSCRKPPEIRY NIRLNGHTRTVEEGALFDTETGYQGMRFPRFLCFETRASEFDPDTSEPIKFDPYLSEV MKHWKAGQAVFGGDTGAGFGFRLEGDIFRFFSMDVAKKEEYDPYLLARGFKGMSS QEILEKIGSGRTYDWNVPKIALNIPPNKLPWKEICYTIEVISPLISRDPIRAMMDPRNT DTIMIRKRVFVPDGKGGTLPEPESRYFIKSETLRGILRSLVGGNKTADGEYLCDLHDHED CDCVQCRLFGSIHQGCLRFEDA EVWNSVRDKKMDHVAIDRFTGGALDQMKFDDYP LLGCEPYPVILGGRFWIRDDISDKEIA
SEQ ID NO: 13 OPY65764	MESIPVTLTFLEPYRVVEWYANEDRRSAERYLRGQSFARWHRKKNDKKGRPYITGTL LRSAAIRAAEELLSSGGVWDGQHCKGQFLSGGVKPEYMRKRPTYIWAKEGACS APDYCPFCIFLGDRDQAEKKAESQNGYPDKSYHIRFGNLSLPDPPPLLDLKEVAVERT LNRVDFQTAKAHDYFKVWEISHEDLGVTGQIVIHYTGPWQEKVKSLLGLESLRFVDR LCGALCKIEMAPKPARPLPKSLSDVMTEHAKIIVTAFDDAKKAEKVRGLADAMRSMG

SEQ ID NO & Protein ID/Contig	Sequence
	SKGPTILDKLPAGHDDRDHHTWDVTIVDKTPLRITYLKGVLRADDAASWPALCKALGNALYDVSQG
SEQ ID NO: 14 RME63343	EKQGFRDKGFNIVGSLKDAIGKEIGLREISLRPAKEETMPRWQCVEYTHVNSPLHTAD PIEALLHSGNYDSVVYKKTIVVRNGNIKQIPVFKGETIRGIVRTAFARILRTENVEFDEE HEDCTCPLCQVFGNEHRAGRVRFEDLVIEGYTSEKKFDHVSIDRFTGGAAEKRFDD LSLKGSPPRPVLRGKVWIRNDMDSKGIEKLKQAFMDIRDGLYPLGSRGGIGYGWVT DLKIENTEVEEFRLDKVSTTEGSGPATEEFNFPSPLEIQLNKDAVYHPHYFIRPHEKVN REIRPVGHERFHDDLLTGRIKCTLKTLP LIIPDTEPDFAFLQAEHKGHQNRFFR
SEQ ID NO: 15 Ga0190306_10 003932	MKSIPITLTFLEPYRILPWAEEKGRDKKEYLRGANYVRLHKDKNGKFKPYITGTLIRS AVLSAIEMLLDITNGEWNGKECCLAKFHTEGEKPSFLRKKPIYIRAEKDEICTSRETAC PLCLILGRFDKAEKKEKDEKFDVHFSNLNLYSSKEFSTIEELAPKRALNRIEYQTGKA QDYFTVYEALNKEFWTFKGRIRIKEDIYDKVTDLLFSALRCVEKIAGALCRIEIDKEPS QQKGFVKRQLSKQAKEDIEKIFQVVKDAQKLRLSDCFRELTRMANKDELALPLGPE DDGHYLWDKIKVEGKTLRIFLRNCFSSQYKDNWLCFCDEASKKGYQKYREKRHKLT RELPTATPKHFAEKKDPQISPIYIDKDDKVYEWIIVGRLLIAQTPFFHFGDEEKAEGAILLT PDNRFRPLRPTALRGILRRDLKLAGASACEVEVGRSEPCPCDVCKIMRRVTLLDTVSED LRDFLPELRKRIRINPQSGTVAEALFDTEVGPEGLSFPFLRYKCEKLPDSLTTVLCW WQEGLAFLSGESATGKGRFRLEINGAFVWDLQKGLFNYIKNHGFRGEERLFLEGNEA ELEKMGIQINTELLQPEMIKKEKNFTDFPYDLIKYQLNISSPLLLNDPIRAIALYEGEGK APDAVFFKKYVFENGKIEEKPCFKAESIRGIFRTAVGRIKNVLTKNHEDCICVLCHLFG NVHETGRKFEDLKIVSGQEEKFFDHVAIDRFLGGAKEKYFKDDKPIIAGDPDTPIVLEG KIWVKKDINDEAKETLSQAFSDINTGIYYLGGANGSIGYGWIEEVKALKAPSWLKIKEK PNFEKDTSLNISAIMNEFKKDIQTLNLDKTYLPYGFLKLEKVKRTSSPITHERFYENH LTGFIECSLKVLSPLIIPDTETPEKEENGHKYYHFLKIDNKPIPGAEIRGAVSSIYEALTN SCFRVFGEKKVLSWRMEGKDAKEFMPGRVSKKKGKLYMVKMQALRLPVYDNPALA NEIRSGSIYEKYNSKVEIIFQTVGIRKFLRGFNFNVEWKKVLVTGIDPLAILPSQKIP GNDKWVKNLQSKISPVRGYFKFTGPNKIETKRREEEKDEKLRTKANKVSCLQKDKW YEAMHNHVEYKQDYTPPNSPKTEPLERPNIPCFVCSDEKIKYRMTRKRCERVFVSLGE NAPKYEIPISAIKRYEVILSAYRENWERNKTPELFRTRLPGDGRTLNEDDL VYFRADEN EKVKDIIPVCISRIVDEVPLIKRLSQELWPCVLAECPLLGFCECKCELEGLPEKIWFRIN KDGLCPACRLFGTQIYKSRVRFSAFAYAKNWKFYDGYITLPRLESPRATWLILKEKDKH YIKYKVCGRKFYLHNSTYEDIINNSKKEKEKTENNASFEVLKEGEFTFKVYFENLEN WELGLLLSLTGLGEAIKIGHAKPLGFGSVKIEAKKIYFREEAGKFHPCEKADEYLKK GLNKLTSWFGKNEINEHMRNLLLFMTYYQNLPKVKYPDFDGYAKWRCSYVEQDKV EYFQNRWIVAS
SEQ ID NO: 16 Ga0193932_10 4825	MIINITVKFLGPFRMLEWTDPDNRNRKNREFMRGQAFARWHNSNPQKGSQPYITGTL VRSVIRSAENLLMLSEGKVGKEKCCPGEFRTENRKKRDAMLHLRQRSTLQWKTDK PLCNGKSLCPICELLGRRIGKTDEVKKKGDFRIHFGNLTPLNRYDDPSDIGTQRTLNRV DYATGKAHDFFKVWEIDHSLLSVFQKISIANIGDGATKLLLEDLSLRTDRLCGAICVI SYDCIENS DGKENGKTGEAAHIMGESDAGKTD AENIANAIADMMGTAGEPEKRLILA DAVRALRIGKNTVSQPLDHEGKENHHLWDIGEGKSIRELLLEK AESLPSDQWRKFCE DVGEILYLKSKDPTGGLTVSQRILGDEAFWSKADRQLNPSAVSIPVTTETLICGLISET PFFFGTEIEDAKHTNLKVLLDRQNRYRLPRSAIRGVLRDLRTAFGGKGCNVELGGRP CLCDVCRIMRGITIMDARSEY AEPPEIRHRIRLN PYTGTVAEGALFDMELGPQGLSFDF ILRYRGKGKSIPKALRNVLKWWTKGQAFLSGAASTGKGIFRLDDLKYISFDLSDKDK RKDYLDNYGWRNRIEALSLEKMPLDRMNDYAEPLWQKVSVEIEIGSPFLNGDPIRALI EKDGSDIVSFRKYADDSGKEVYAYKAESFRGVVRAALARQHFDKEGKPLDKEGKPL LTLIHQDCECLICRLFGSEHETGRLRFEDLLFDPQPEPMIFDHVAIDRFTGGAVDKKKF DDCSLPGTPGHPLTLKGCFWIRKELEKPDDEKSEREALSKALADIHNGLYPLGGKGAI GYGQVMNLKIKGAGDVIKAAALQSESSRMSASEPEHKKPDSGLKLSFDDKKAVYYPH YFLKPAAEVNRKPIPTGHETLNSGLLTGKIRCLTTRTPLIVPDTSNDDFFQTGVVEGH ESYAFFSVNGDIMLPGSEIRGMLSSVYEALTNCFRVFDEGYRLSWRMEADRNVLMO FKPGRVTDNGLRIEEMKEYRYPFYDRDCSDKKSQEA YFDEWERSITLTDDSLEKMAE RKGDISPDKLVKSLKGKNYKSTEGLLAAFKDKGGDTGGNIGLIFKYAERIGDVPR YEHPDTDRMMLSLSEYNRNQKSDGKRAYKIIKPASKLKGAYFMFAGTSVENKRIC NPACTDKANKSVKGYLKISGPNKLEKYNISEPELDGVPEDRNCQIIHNRIYLRKIFVAN

SEQ ID NO & Protein ID/Contig	Sequence
	AKKRKERDRLVGEFACYDPEKKVVTYSMTKRCERIFIKDRGRTLPTITHEASELFEILVQE YRENAKRQDTPEVFQTLPLDNGRLNPGDLVYFREEKGKTVEIIPVRISRKIDDSPIGKR LREDLRPCHGEWIEGDDLSQLSEYPEKKLFRNTEGLCPACRLFGTGAYKGRRLRFGA KLENDPKWLMKNSDGPSHGGPLTLPLERPRPTWSMPDDTLNRLKKDGKQEPKKQK GKKGPQVPGRKFYVHHDGWKEINCGCHPTTKENIVQNQNNRTVEPLDKGNTFSFEIC FENLEPYELGLLLYTLELEKGLAHKLGMKPMGFGSIDIEVENVSLRTDSGQWKDAN EQISEWTDKGGKDAGKWFKTDWEAAEHKLNKLLFLPGEEQNPRVIYPALKQKDIP NSRLPGYEELKKNLNMEKRKEMLTTPWAPWHPIKK
SEQ ID NO: 17 Ga0190283_10 011062	MTQITIQTFFHPRVVPWNHRDHRKTDRKYLGGTFKWHCTASEGKSGRPYITGT LLRSALFAEIEKLI AFHDPFKCCRGKDKTENGNAPLFLRRRPRADCDPCGTCPLCLL MGRSDTVRRDAKKQKKDWSVHFCNLREATERSFNWKETAIERIVNRVDPSSGKAKD YMRIWEIDPLVCSQFNGIITINLDTDNAGKVKLLMAAGLAQINILAGSICRADIISEDHD ALIKQFMAIDVREPEVSTSFPLQDDELNNAPAGCGDDEISTDQPVGHNLVDRVRISKIA ESIEDVFSQEQKAQQLRRMADAIRDLRRSKPDETTLDALPKGKTDKDNSVWDKPLKK DILPSRMPASEDDDTPTLRKVLKDEINGQEDMWRKFCEALGNSLYDLSKKAKERKR TEALPRLLGETEYGLPMRENKEDEPLPSSLYKFKWLIAGELRAETPFFFGTEVQEGQ TSATILLNRDGYFRLPRSVIRGALRRDLRLVMGNDGCNMPIGGQMCECGVCRMVHRH VIEDGLSDCKIPPEVRHRIRLNCHTGTVEEGALFDMETGYQGMTPFRLYCETNSDL DSYLWEVLNNWQNGQSLFGGDTGTGFRFELTEPKVFLWNFSKKEKHEAYLLNRGF KGQMPVQDVKTKSFKTKTWFOIHRELDISPKKLPWYSTDYRFNVTSPISRDPGAML DPRNTDAIMVRKTVFCPDPAKNRPAPATVYMIKGESIRGILRSIVVRNEELYDTHDHE DCDCILCRLFGSIHQGSLRFEDAQVQNSVSDKKMDHVAIDRFTGGGVDQMKFDDYP LPGCPAQPLILEGKFWVKDDIDDESKSALEKAFADFRDGLVSLGGLGAIGYGQIGDFE LIGGSADWLNLPKPEENRTDVPCGDRSAQGPEIKISLDADKIYHPHFLLKPSDKNVYRE RELVSHAKKKGPDGKSLFTGKITCRLSTEGPVFIPDTDLGEDYFEMQASHKKHKNYGF FRINGNVAIPGSSIRGMISSVFEALTNSCFRVFDQERYLSRSEKPDPTELTKYYPGKVKR DGNKFFILKMKDFFRLPLYDFDFEGEAESLRPNYDEDRNEEENKGNKNTQKVKNVAV EFNIKMAFGAKHNRDFLKKYKEQEIKDIFMGKKKVYFTAGKHKPNEAHDNDKIALLT KGSNKKAEKGYFKFTGPGMVNVKAGVEGEECDFHIDESDPDVYWNMSSILPHNQIK WRPSQKKEYPRPVLCVKDGT EYVMLKRSEHVFAEASSEDSPVPGKVRKQFNSISR DNVQNTDHLSSMFQSRRLHDEL SHGDLVYFRHDEKRKVTDIAYVRVSRTVDDRPMG KRFKNESLRPCNHVCVEGCDECPDRCKELEYFSPHPEGLCPACHLFGTTDYKGRVSF GLGWHESENTPKWYMPEDNSQKSHLTLPLERPRPTWSMPNKKSEIPGRKFYVHHP WSVDKIRNRQFDPAKEKQPDVVIKPNENNRTVEPLGKGNEFTKEVRFNNLREWELGL LLYSLELEDNMAHKLGMGKALGMGSARIKAEAIELRCESAGQNAELKDKAAFVRKG FEFLEIDKPGENDPMNFDHIRQLRELLWFLPENVSANVRYPMLEKEDDGTGTYDIFI QEEPSTGKRNP SYLSSEKRRNILQTPWKHWYLIPPFQAS AQSETVFEGTVKWFDDKKG FGFIKINDGGKDV FVHHSSIVGTGFKSLNEGDSVAFKMGVGPKGPCAEKVKKIGN
SEQ ID NO: 18 Ga0073580_10 36305	MTKIPISLTFLEPRLVDWVSESERDKSEFLRGLSFARWHRIKNQREDENQGRPYITGT LLRSAVIKAAEELIFLNGGKWQSEECNGQFKGSKAKYRKECPRRRHRATLKWTD NTCSDYHNACPFCLLGLCKPNSKENS DIHFSNLSLPNKQIFKNPPEIGIRRLNRVDFT TGKAQDYFYVWEVEHSMCPKFQGTVKINEDMPKYNVVKDLLISSIQFVDKLCGALC VIEIGKTKNYICQSFSSNIPEEEIKKLAQEIRDILKGEDALDKMRVLADTVLQMRTKGP EIVNELPRGIEKKGGHWLWDLRLRKKFK EIANNYKDSWQELCEKLGNELYISYKEL TGGIAVKKRIIGET EYRKIPEQEISFLPSKAGYSYEWIILGKLISENPPFFGKETKTEEQID MQILLTKDGRYRLPRSVLRGALRRDLRLVIGSGCDVELGSKRPCPCPVCIMRRVTLK DARSDYCKPPEVRKRIRINPLTGT VQKGALFTMEVAPEGISFPFQLRFRGEDKFHDA QNVLVWWEKGLFLGGGASTGKGRFKEIHLVLDLKNFHSYLYQYKGLRDKGD FNSIKEIEGLKVETEEFKVKKPPWSCVEYTI FIESPFVSGDPVEAVLSSNTDLVTFKK YKLEESKEVFAIKGESIRGVFRTAVGKNEGKLT TENEHEDCTCILCRLFGNEHETGKV RFEDLELINDSAPKRLDHVAIDRFTGGAKEQAKFDDSPGSPDPLEFTGIVWVRDDI DEEEKKALKSAFLDIKSGYYPLGGKKG VGYGWVSNLKIESGPEWLRLEVQEKSSQEN VLSPVILSEVM DIEFNPPKIDENG VYFPYAFRLPLNEVKRTREPIGHNEWKKSLSISGYLT CRLELLTPLIPDTSEEVIKEKVNNGEHPVYKFFRLGGHLCIPAAEIRGMISSVYEALTN SCFRVFDEKRLISWRMTAEAAKRDPK KSEEQNRMRFRPGRIKKDKKFYAQEMLEL RIPVYDNKDKRNEISQNDPTRPSEYNHPT EPERIFFSNAEKIRNFKRNSNYLHGSTPLL FRQWSISNRYDKIALIGNKSQGH LKFTGPNKIEVSEGTKCPKYETIPGRDEWDKAVHN

SEQ ID NO & Protein ID/Contig	Sequence
	YVEPGKFVTVISRKKGQKPKAVQRRRNVPFCCYDYNTNRCFVMNKRRCERVFKVSR DKPKYEIPDAIRRYEHVLRKYRENWERYDIPEVFRTRLPGDGETLNEGDLVYFRLDE NNRVLDIIPVSISRISDTQYLGRRLPDHLRSCVRECLYEGWGDCPKKLSLFPEKMWIR INPEGLCPACHLFGTQVYKGRVRFGARAGSNWKFREEQLTLPRFETPRPTWVPIKPK DEYQIPGRKFYLHHNGWEEIYKKNNKNEIKKEKNNATFEVLKQGTIFYFKVFFENLEL WELGLLIFS AELGGEFAHKLGHGKALGFGSVKISVDKIILRRDPGQFEQRGQKFKRD AVDKGFCVLENRFGKTNFKIYLNFLQLLYWPNNKKVKVRYPYLRQEDDPEKLPGY VELKKHQMLKDDNRYSLFARPRAVWLKWTEMVQRDKS
SEQ ID NO: 19 Iso3TCLC_100 1005823	MSVEEFYVRLTFLEPFRVVPWVRNGDERKGDRIYQRRGGTYARWHKINDSHGQPYITG TMLRS AVLREIENTLT LHNTYGCCPGGTRTTEGKLEKPLYLRRRDGFEFENHAEKPCS EEDPCPLCLIQGRFDKLRRDEKKQFVRQGNISFCSVNFNLNISSGIKSFSEWEEIAVSRV VNRVDPNSGKAKDFFRVWEIDHKLCPNFLGKMSISLSEKLEDVKALLAVGLAQVNVL SGALCRVDIIDPETQKDTVHQHLLIQQFVTRIQDKEKGDAADIPAFTLPPAGLSPPSSNEW NDTIKSLAEKIRKIKELEQGQKLRQMADVIRELRRKTPAYLDQLPAGKPEGRESIWEK TPTGETLTLRQLLKSANVPGESWRAFCEELGEQLYRLEKNLYSHARPLRLLGETEFY GQPARKSDDPPMIRASYRAFPSYVWVLDGILRAETPFYFGTETSEGQTSQAAILCPDGS YRLPRSLLRGVIRRDRLAILGTGCNVSLGKVRPCSCPVCCEIMRRITVQQGVSSYREPAE VRQIRSNPHTGTVEEGALFDLETGPQGMTPFPRLYFRTRSPYIDRALWLTINHWQEG KAIFGGDIGVGMGRFLENLQIRSADLVSRDFSLYLRLARGLGLSREEVTRIGLNEE QWEAVMADDPGTHYNPFPWEKISYTLIIHSP LISNDPIAAML DHDNKDAVMVQKTVL FVDESGNYSQMPHHFLKSGGIRGACRFLGRKDAPNENGLTYFEADHEECDCLLCSL FGSKHYQGKLR FEDAE LQDEVEAIKCDHVAIDRFHGGTVHRMKYDDYPLPGSPNRL RIKGNIWVKRDLSDTEKEAVKDVL TELRDGLIPLGANGGAGYGRQRLMIDDGPGWL ALPERKEDERPQPSFSPVSLGPVHVNLKSGSDTADVYYYHPHYFLEPPSQTVSRELDII SHARTRDSGGEALLTGRILCRLITRGPFIPTDNTNNDNAFGLGEGGIGHKNYRFFRINDELA IPGSEL RGMVSSVYEALTNSCFRIMEEGRYLSRRMGADEFKDFHPGIVVDGAKIREMK RYRLPLYDTPDKTSRTKEMTCPELFRKDG RPERAKKFNEEIAKVAVQNRAYLLSLD EKERREVLLGNREVT FDECPDDEYS DDEYSELKYAQKYKDFIAVLKKNQKRGYIKF TGPNTANKKNEDAPDKNYRSDWD PFKLNILLES DPECRVSNIH CYPRPLLVCIKDKAE YRIHKRCEAIFCSIGSPSDLYDIPQKVS NQYRTILQDYNDNTGKIVEIFRTQIKHDQLTT GDLVYFKPAANGQVNAVIPVSIRKTDENPLAKRFKNDLSRPCAGLCVEDCNECPAR CKKVADYFNPHPRGLCPACHLFGTTFYKGRVRFGFAWLTGEDGAPRWYKGPDPDGS GKGRPMTIPLLERPRPTWSIPDNSFDIPGRKFYVHHYPYSDVGIDGETRTPNNRTIEPLAE GNEFVFDIDFENLRDWELGLLLYSLEEDSLAHKLGLGKPLGFGTVQINIRGSLKNGS KGWDTKTGDDKNQWIKKGF AHLGIDIKEANERP YIKQLRELLWVPTGDNLPHVRYPE LESKTKDVPGYTSLLKEKDLADRV SLLKAPWKPKPWSGTAPHPDKGTNRRLRASIVE RDRIQRKTD TAKPEKKEETKVGKSSSSDIEKRYVGT VKW FNDKKGYGFILYGTDEEIF VHRSGVADNSIPKEGQKVGFRIERGARGSHAVEVKAIE
SEQ ID NO: 20 SESD01000293 .1	MPRFQLSLTFFDEPFRLIEWTDKSNRNSANTQWMRGQGFARWHKITLEKGFPFVTGT AVRSKIIREVEALLSRNKGTWNGIPCCSGFFDTKGPSPTHLYRPTLEWEYGKTVCTSE ADVCPCLLLGRFDQAGKKS DTPCQSTDYH VHWENLSAGVAQYRLEDIAQKRTSNR VDFFSKKAHDHYGVWEVTAVKNLLGYIYISDAITESHQKTVISLLKAALSFTDTLCGA NCKLELSDEPVDSIHSNQ SASNFNPHSGAAPSQCSQSMPPFNMDQETKELANTLCKAF TGNMRHLRTLADAVREMRRMSPGISSLPGRNLKEGEITAHYLWDERIDEKTIRQVLE DTIELSPARSIIYKNWISFCNLGQKLYERAKDNDPILERKRPLGEAAFSKVPTSSHAPR HDMNSRVKGGFTREWIIVGTLRALTPFYMGTGSQAGKQTSMP TLQDSNDHFRLPRTA LRGALRRDINQASDGMGCVELGPHNLCSPVCQVLRQIRLLDTKSKFSMPPAIRQKI CKNPVLSIVNEGSLFDVELGIEGETFPFVMRYRGGA KIPDTIITVLSWKNERLFIGGE SGTGRGRFVLECPRIFCWDVEKGQNDYIQYHGF RNKEDELLSVYSTVSGLAEKNDVN LNNARDFSFDKICWEVQFDGPVLTGDPLAALFHGNTDSVFYKKPILKSKEEPSYQW AIKSDTVRGLIRSAFGKRDALLIKSHEDCDCLLCEAFGSKHHEGKLR FEDLTPKSDEIK TYRMDHVAIDRISGGA VDQCKYDDEPLVGTSKHPLVFKGMFWINRDSSVEMQRALIA AFKEIRDGLYPLGSNGGTGYGWISHLAITNGPDWLNLEE VLPQPTADIPVEECTAEPY PKFQKPDLDQNAVYYPHYFLQPGKPAERERHPVSHDHIDDKLLTGRLVCTLT TKTPLI IPDTQTNTMLPPND APEGHK SFRFFRIDDEVLP GSEIRGMVSTVFEALTGSCFRVINQK AHL SWRINADMAKH YRPGRIIQNNEKMF IQPYKMFRLPFYAGFDPRNCLSEKQLLGIE PVKLWVKDFVASLVK PQTDIDIEWKEKIGFVRVTGPNKVEVDSNTDPDPSLPECESDW

SEQ ID NO & Protein ID/Contig	Sequence
	KDIHITEDGSTPSKNDRVYRCQLKGVITYTVAKWCEAFWVKDEGKKPITVNAEAINRY HLIMKSYQDNQSPPIIFRSLPVLNYKQDQKIIGSMIFYRESAKSDKIVNEIIPVKISRTA DTELLAKHLPNNDFLPCAATCLNECDTCNAKTCKFLPLYREGYPVNGLCPSCHLFGTT GYQGRVRFGFAKMNGNAKFCQGGGERPEDRAVTLPLQERPKLTWVMPNENSTIPGRK FFLHHQGWKKIVDEGKNPINDVIEPDANNRTVEPLAAGNDFSFEVFFENLREWELGL LRYTLELESEL AHKLGMGKAFGFGSVKIKIKSVDLRKQGEWEKATNTLVSEDKKSSW YNIHTVNNLRALTALYYVEDDKIQVNYPKLKKDNESDNRPQGYVEMKKTAFVPRDILTP WWPWWPPTPPPMNQSGNQSYARSEEPARITESQPEVYKTGTVKFYKHDKKFGFITM DGRENIHFAGNQICRPETSLSQSGDKVKFIEGENYKGPALKVERLKG
SEQ ID NO: 21 OBJA0100112 7	MRLKINIHFLEPFRLEWHEQDRRNKGNSRWQRGQSFARWHRRKDNDQGRPYITGTL LRSVVIRAVEEELARPDTAWQSCGGLFITPDGQTKPQHLRHRATVRRARQTAADKCAD RQSACPFCLLLGRFDQVGKDGKKGEGLRFDVRFNSLDLPKDFSPRDFDGPQEIGSRR TINRVDDDETGAHDFFSIWEVDAVREFQGEIVLAADLPSRDQVESLLHHALGFVDRLC GARCVISIADQKPAEREERTVAAGDEKATIADYDQVKGLPYTRLRPLADAVRNLRL DLAELNPKPDGKFLPPGRVNDGRRVPHYVWDIPLGKGDTLRKRLFLAASCEGDQA KWRNICESEGQALYEKSKKLKDSAPAPGRHLGAAEQVRPPQPPVSYSEESINSDPLA EWITGTLRAETPFAIGMDAPIDDDQTSRTLVDRDGRYRLPRSLRGILRRDLSASG DQGCQVRLGPERPCTCPVCLILRQVVIADTVSETTVPADIRQIRRNIPITGTAADGGLF DTERGPKGAGFPFSLRYRGHAPMPKALRTVLQWWWSAGKCFAGSDGGVGCGRFALD NLEVYRWDLGTFARQAYSENNGLRSPREEFDLAVIHELAEGLAKEDGQKILKGTEPF TCWQERSWQFSFTGPLLQGDPLAALNSDTADIISFRRTVVDNGEVLREPVLREGELRG LLRTAVGRVAGDDLLTRSHQDCKCEICQLFGSEHRAGILRFEDLPPVSPPTVADKRLD HVAIDRFQSVVEKYDDRPLVGSPPQPLVFKGCFWVQTSGMTHQLTELLAQAWRDI AAGHYPVGGKGGIGYGWINSLVVDGEKITCRPDGDSISLTTVTGDIPPRPALTPPAGAI YYPHYFLPPNPEHKPKRSDKIIGHHTFATDPDSFTGRITCKLEVVTPLIVPDTEGEQPKD QHKNFPFFKINDEIMLPGAPLWAAVSQVYEALTNSCFRVMKQKRFLSWRMEAEYK DFYPGRVLDGGKQIKKMGDKAIRMPYDDSTATGSIKDDQLISDCCPKSDEKLQKAL ATNQKIALAAKHNEQYLAQLSPDEREEALQGLKKVSFWTESLANNEAPPFLIAKLGE RGKPKRAGYLLKITGPNNANIANNTNPDGGYIPSWKDQFDYSFRLLGPPRCLPNTKG NREYPRPGFTCVIDGKEYSLTKRCERIFEDISGGENQVVRAVTERVREQYREILASYRA NAAGIAEGFRTRMYDTEELRENDLVYFKTAKQADGKERVVAISPVCISREADDRPLG KRLPAGFQPCSHVCLEDCNTCSAKNCPVPLYREGWPVNGLCPACRLFGAQMYPKGRV NFGFARLPDDKQPETKTLTLPLLERPRTWVLPKSVKGSNTEDATIPGRKFYLHRHDGW RIVMAGTNPITGESIEKTANNATVEAIMPGATFTFDIVCENLDDQELGLLYSLELEEG MSHTLGRGKPLGFGNVRIKVEKIEKRLSDGSRRMIPPKGAGLFMTDKVQDALRGLT EGGDWHQRPHISGLRRLTRYPEIKARYPKLSQGEDKEPGYIELKSQKDENGVPYINP NRELRVSENGPLPWFLAKK
SEQ ID NO: 22 OBEQ0118074 20	MSNQTRWIEGTLELITPLHIGTGLDEKERDENKETRWLEAVALDHKGQPYIPGASLK GALRALAKRHDFRNLFDNKEVDGDFVRQAEFLSAWCVPDTPDKGRLIQPRVAIDRV GTAQDKKLFQTRLITPGRFAMKIVVQNAVENEIADLLGLLNLPPDDPQFSLGAYANQ GQGRVQWFGKIQTRCFGINEAKAWYEEIRKDESKCWTAFAKPKNVSTPPTPAKEAQL TLPLNLAFHTPFLVKQAGIKADDADAVPRRTHDDKIVLPASSLRGRLRTQSERILRTL CETPQGHAPAYRKQPHDDLAVLLFGAAGWRGIVQTSDCIVEDKSIKTRRHEMLAI DRFTGGGKDGAKFNVYDVECPFLAGLSDLARLKNALKGGKDALLPALGLMTL MLRDLAEGDIPFGYGISKYGQCRASSALGDWAELLKQHLGADSADTTVQALREYL GNPKGQELKLDPPSADATQAGVPAQQNAAKTQAQGAQEFHNPYHFIPLSKPDISQW PEPQKLTEKGHSHDRYASLSGRIVCRLTTQTPLFIGSEQTTPTNPQAPKSLHPFKLNG LAIPATSLRGMISLFSVSNSNFRVLDEKQYSMRKTMQQSLSAMGRIVRHDQKLYLL PLTLPLPQGPFGVYDLGEKWSAVFDWQPPPLRIYFDPPPRRTYQSQQQPCYMKLSTV KYSESNNPNIAGENLGALRFPRGNQNTQFLIGQSNQDECPITQAEYAQKSEDERNEYT SGWVRTLVKPGRDLPRSVKHHVFLPDVFIDAPPPVNDLYPIPDSVIQRFHDLADQVLA SMNLKPEEIVDSTNLLPYTPVGRRSDDCDRDLQAGDLIFFDIDPPLHPGEKSQITEISF SSIWRSGIGKDHLTTDPDLLTNFDVNLQPHGMPGRTQSLSPAELLFGLVGTQNDQATT AYAGKVRIGFGLPEEGHNPRLDARITLKESSPKPPSPALYFRKKSGKDEYVSKANLA DKPEDYILRGRKMYLHAWRKNEQVVELSDTGHDGGVRPPWVSKFDESADDEGNKRR VSIEPIAKDESIFYEVDVFNLSRTELAQLCATLYPNEKFEHRLGMGKPLGLGSIKITPLS LFLVNSRQRYATDGLDKPRYHAVWHTGTASEPRWPDHLQREQQGIAFEGVSTAPT

SEQ ID NO & Protein ID/Contig	Sequence
	MSLA AEAKVSD DVKRALELLGNPDEITVPVHYPQLHNGLMESKQFDWVQNDKSGR DQPANNRQHLLSSFTKDTEKLEPLIRIMRR
SEQ ID NO: 23 OVOO0100010 6	MTTPSAPKSSLPALHWLIRAELEVLTPHLGTGTDQRITPDAPDADPYWQADIALDAD GRPYLP GASLKGALKALARRRQVDAPCLPLFGDLNRGGAPHPDCIPPRRTRAGLAEFR DALQSHATQADGPDTAQPRIADRITGSVVDKCLFHTQTVPVGTRFSVEIILRRADQNL AAQLVALLQHGPDPDFRLGAHANLGFGRVGLYGNIDTRRFGPQQAQHWFAAAQTQ ADARWTDFAEAVTLTAPAPAPAQPAHRLALPLSLTFHTPFLVKQPEHKHRKPQDNA PDGTPRQRGDRALLPGASLRGRLRSQAERILRTLGCCKVAQGHAVPPVKNNTCPDPAT LLFGTAGWRGLLRTDDCTGTAPATLVHDHMLAIDRFTGGGKDGAKFKLRYAECPTL EGQLSLDLSRLRSARLDGANAAADTPWIALGLLTLVLRDLAEGDIPFGHGSAGKYGRC RAQGLPDRWRQALEAHFGPNADARALAALRAWCRTHATAALDAPCSLAGSAPT AAAPSGQAAPADAFHNPYHFIPFSQPDIDRWLSPDAHRKTGGHSRYRGLSGRLVCAL TTVTPLFVGAAARTPASDQHPKPVAGFALQNQPAIPATSLRGLLSSLFESISGSNLRVL HPTPYSIRKTTKEALSAIGRIVERNGELKLYPLTLPTIHQNADNAYPVPARWRKVFW ESPVPLRVYFGSRKQTYDSRQPHYLPQELSYLPNDSDCIAPDQGDRLFRPSRDRDRKFL IGQCPISRYDCPIPETDLPKLSPQERPRYTRGWVRSWTSNREKELPHTVKHQLFIPDP VETPAADDLLPIPGVLDTFHALADLALAGQHWGKDETPADDQLLPFTPAGRQRHD ADRPPRDRQTRLQPGDLVCFDLGDDGAVSEISFSSIWREGLRLAGKPNLATTADLL AQVSPHLLPLGMPGRSARLSPVEQLFGVVEYRPPQTAKGTRKPTDAPAAAYALAGKLQ VGFGPRPARPFEREPAVTLKELSTPKPPSPALYFRPKAGDGYVSKAKLASQPQDYAPAG RKHYLHALRRQGQVARLDNSGHVPSDGSGRPPWQSRFDGQEDSGNKRVRVVEPIPA GETFHFEIDFDNLSPTLEQLCATLLPHPAFEHRLGMGKPIGLGSKVLAVEGLLLVDRP RRYAEDEPNAPRHHRGWRANADAGWPDHLQGDSPAAPLEATEQPAALAERAMARV PADVRRALQLLGNPGAVAAPVHYPQVKDAQIEEKHYLWFWANDDEKTAGGNRHLP RLHANSPGLPTLRLVKREKDHSSNTGKPRRK
SEQ ID NO: 24 PDWI0100592 2.1	MIPDLRSLVVHISFLTPYRQAPWFPPEKRRNNNRDWLRMQSYARWHKVAPEEGHPFI TGTLRLSRVIRAVEEELCLANGIWRGVACCPGEFNSQAKKKPKHLRRRTTLQWYPEG AKSCSKQDRENACPFCLLLDRFGGEKSEGRKKNNNDYDVHFSNLNPFYPGSSPKVW SGPEEIGRLRTLNRIDRLTTKAQDFRIYEVDQVRDFFGTITLAGDLPRKVDVEFLRR GLGFVSTLCGAQCEIKVVDLKKKQNNKEDSILPVSEVPFFLEPEVLAKMCQDVFPSGK LRMLADVILRLREEGPDNLTLPMGSQGLGGRPLPHLWDVPLVSKDRETQTLRSCLEK IAAQCKSEQTQFRLFCQKLGSLLFRINKGVYLAAPNSKISPEPCLDPSKTIRTKGPVPGKQ KHRFSLPPFEWIITGTLKAQTPFFIPDEQGSHDHTSRKILLTRDFYRLPRSLLRGIIR DLHEATDKGGCRVELAPDVPTCQVCRLLGRMLLADTTSTTKVAPDMRHRVGVVDRS CGIVRDGALFDTEYGIEGVCFLPIRYRGNKDLEGPIRQLLSWWQQGLLFLGGDFGIG KGRFRLENMKIHRWDLRDESARADYVQKCGLRGVDGDDTAINEKDLNLNLPESGY PWKKHAWKLSFQVPLLTADPIMAQTRHEEDSVYFQKRIFTSRGRVVLVPALRGEGLR GLLRTAVSRAYGISLINDEHEDCDCLCKIFGNEHHAGMLRFDDMVPGTWNDDKID HVSCSRFDASVVKFDDSLVGSPLHFEFTFWLHRDFQNDVEIKTALQDFADGL YSIGGKGGIGYGWLFDMIPRSLRKLNSGFREASSIQDALLDSAKEIPLSAPLTFTPVKG AVYNPYYYLPFPAEKPERCLVPPSHARLQSDRYTGCLTCELETVSPLLLPDTCREKDG NYKEYPSFRLNNTPMIPGAGLRAAVSQVYEVLTNSCIRIMDQGTLSWRMSTSEHKD YQPGKITDNGRKIQPMGKQAIRLPLYDEVIHHVSTPGDTHDLEKLKAIVLELTRPWKE LPEEQKKRFEKCKNILDGRMLQQKELRALENSGFAYWRDKTSLTFDSFLKDAIEQE YPRYSGDYQRIKALVVNITLPWKLKKEERHKRFDKCRILKGQOPLTKDERKALEES GFANWHGRELLFDRFLKDENSECLIAETTDRIASVAKNNRDYLFIEIKQQDFARYKRI IQGLERVPFSLRSLAKSKETSQIACGLRRGRFLRKGYLKISGPNNANVEISGGSHNS GYSIDIWDDPLDFSRLSGKSELRPNTQKTREYPRPSFTCTVDGKQYTVNKRCEVFE SAAPAIELPRMVREGYKGLTDYEQNAKHIPQGFQTRFSSYRELNDGDLVYYKTDQSQ RVTDLAPVCLSRLLADDRPLGKRLPEEYRPAHVCLCECDPCTGKDPCVPPIYREGYPAR GFCPACQLFGTQMYKGRVRFSGVVPVNSTSPQLKYVTLPQSQRPRPTWVLPESCKG KEKDVPGRKFYL RHDGWREMGDDDKPDSRPSSEECQDIIEGIGPGEKFHFRVAFEN LDKNELGRLLYSLELDAGMNHHLGRGKAFGFGQVKIRVTKLERRLEPGQWRSEKICT DLPVTSSSELVISSLLKVEERRKLLRLVMTYPYKGLTACYPGLERENGRPGYTDLKMLAT YDPYRELVVQIGSNQPLRPWYEPGKSFKPSPGNDCTGRGGSVSKSLISEPKVVPPIAPF CEGVVKWFNSVKGFGIETKEQRDIFVHFSAIRGEGYKILEPGEKVRFEIGERKGPQA INVIRIR

SEQ ID NO & Protein ID/Contig	Sequence
SEQ ID NO: 25 VAPF010013 39.1	MKMNKTWPFREHWEISGYLRTVTPHLHIGSGRTVTRPELTVADRDELVDINAVVTDYT GKPYLPGSTIKGALHAWLQKRLKEESRTCLIQLFGQEEEEAEKKKNNHGGKAFFDA RVIFPHTGPGSLPYWDDCRQTYVAATVAIDNITRTARHRHLHAEMVPPGVTFALTLA GPLDEEDIGLLLAALQGFNESPALVIGAHTANGLGRFSWELSTVRRFGKEHLQGWLE AETRAMRTEAMQPLSREGVEDLLQDGLAQIDHNEDQVRLGLELCFDGPFLVNDPPTK KEQDDKKRRSNTNLRPLRDAIGRPCLPRESSVRGALRAQAERIIRTMEGTCSNPA YKKEIHTDAEIEELSAVCRVFGAPGWKSLEISDFEFVDGEDCDNIQEFVAIDRFTGGA KDKAKFNAEYIGSPRFTGTIALDKRRDLDPWGKGLLYLVLRDLAEGDITLGFGRSKG YGVCRAKIKNLDLLLPEKSAALHKKFAISPADDDKPADTQIAEDTTSNGLGISAGGPA QKTTESEYPPNPSGPGTFHNPYHFVPVVKPSAADRQHWLDKGILPSASENKTQHTHA CYLDDTTNGKKIYHGRIVCRLQAETPMFIGGRHRENTPEILPFTLGGKPAIPATSLRG MLSSIAEAASNSSLRVLEDKTL SYRKSMRANRNEDKPLSALGMIKKIETGDKVEYRLL PLTLPTLVKRGQYYILPEEYQTMFPDGRAKLKVYLNSNYTASDGTNQDFLGKKSWR LPHGEIYYMKLCQDFSLQNGQLTFDSQNQNMLHFPKNRNNFVVGQRSIDNTPPMTKA EWRTNHQTGVPGLRIMQASGRNFPTGRYHEIFIPVPTKKDCKQLYPVDEKAVERFL DLAGEQTKSQNEKNLQYQILPYHPVGTKRNTDPETNDRYMDLNSGDIVYFRPDAT GTKVEEISFSSIWRDRVEDDNHNRAGVHAFFGNIDKELLFPNPKRAEISPAELLFGFVE ERERGVDDGQAPAFAGKVRVSFGRLSSEKKPDTIFQDQVTLKALSSPKPPSPALYFT GNGNGSIAPDLTL SRHSPQGRKFYLHAWQEEENEIIFLSNGKKTSPVINGLYPWES KSNLSRQLPDKHAKLKSAITPIKKGTVFYFHVDFDNLSEWELGGLCYALQPGKEFRHK LGMGKPIGLGSIHIEPAGLYLVHRGNRYSLDGVPDNSRYNGGIWQSEDKRLQEWQEL YPRESTAAASSAAASPADFALFKSGTMTPSVQQALKRLGDPDNNVAPVHYPQVEGA DLEETTYKWFVANDVGSQTKVGNRTTCTEEAARKSMLPLAGRGPLRLKRYKWCP
SEQ ID NO: 26 DRKI01000155 .1	MAAVQDRWTLMDQQGNELKRFRITAELETASSLHIGASETVEHDLIKNDDGTPVQIN ALITGAGGLPIIPGSTIKGRFLARLRERGVDSALLETLFGKGHDRETEDQGRGGRAEFH DAPLCHRLSGARHFPYWRPERQIWVKAQTAVDHRGTALRRSLRYTEMVPPGVFRFR LTITGCMTDAEADVL FALLEDLGDPQACSFAGAGADGNGTMRLFGFRPEVYCLDRS GILGWLASFEGKGGNGGMAMTAAALLQADTVQRRADKVRQAWQPPDVGPRLHVELR FSGPFLVNDPSRNTPDITQAPDMVPLVDEDGNPMLPASSFRGALRAQAERIIRTLGGRC CDTSSPCRPLGSSDKVGELCLACQVFGAPGWGTTLHIQGFCTSVFRREQEQTFVAID RFHGGCKEGALYTIRHAESPRFEGHLVIDPRMPAWGRGLLALVFRDLREGDITFGLGA GKGYGVVDAAVVQDMAELEPYVEAFREQCRQHQMADCHSAPSPQPLRDHDLAEIP PAEEAPGETFLNPHYHFVPIREPDTGSWLARDELSSCCHSHGFYRQQVDDRPL YHGR TCLLETETPLFIGATGDSSVPSRIENYRLGNRIIPAASLRGMLSSLAEEAASNSAMRVL HQGILSYRKKAKNALREIGMIVLRDGRFILPLVPLMEVTKLRHAYTDPAMKHFLDD KNSWSPRCNRVYYLGRDGNQIPAE TRGAGMRPGILRLLGREGRHDALQNKKEHYFIE IPERYVDQDHCDFYRMFIRDARNGTLVPISPAWERYHCLAEERTLSQKNDPELRED KACASLKWLPFHPKGRVRENDPENDVCHLSLRHGDLVFYAEQNRVVSEISFSAIWS RVETSDSYQAVTVDCFVPKELRPFNRDRRAISPAELLFGFVELDESEHSTEKSRYEQM AFAGKVRLSAGLPVEDVEDSALLEPKPIVLKALSSPKPPSPFFYFVMRDGSGAYIAKK DLSPDRHRIKGRKHYLHGLRQRGNPDVRVQSLDRYGHATETAANPPWETCHPEERPQI KVRVQPVRRKTKFFHLD FSNL SRWELGGLCYVLRPTACFRHKLGMGKPLGLG SVRI DIASLQLIDRVRRYGTDDL TAGRYNMGGHFNASCLDLLPQQDSPAPDDSGAAPDPGT LRQDFVKTMDDET VFRALD L LGNPAHVQRPVHYPVQVREMDIEDQTFLWFVDNDKQW KDALQPLTSSSTQLPPLTRNR
SEQ ID NO: 27 DRNY0100054 3.1	MTTVKEKSWAFTGLKRWKIITLTETQTPHLHIGSGEVAEIEINDSQGDRRQVQANAIIRG KDDKPIPGSTLKGKLRSHFESCLDHSKALERVFGKEYQSDEEQGRGGLAEFHDAVCS YVAPGNSYYPNWNENARNTYIEASTTIDCHTGAADATLHYNECVPPGTRFLVTVTGA MSDKDAALIVAALQAFGDENTPIHLGAEEANGKGRMGLFGNVEVSLDHDHDI AWIS QGSDARMATDKFKPLGKEKVNDLAKNITTPTATGGAQRQHFGIELKFDGPFLVNDPS KYSKGDGDQPAVHQPLTDRSNNPILPARSFRGAIRAQAERIIRTMGGACCDTQSPCQN SGQLCIACQMF GTTGWKTTL SISDFTYDGEYRPAKTQQFVAIDRFHGGGKDGALFSIK YFERPVLKGGISLKL RNQNADELSWRKGLLALLFRDLQEGDITFGFGANKGYGGVEE ACITNADVISTADIEAFRAKCHANHADSWCSPVSKPTNRDDKSSLPSINPATGAGHAF HNPYHFIPIKAPDTSTWLDKHKLATPGSPHSHAYYRSCSDDDKPLHHGRITCKLTAET PLFVGSGDAENQLTDSEAKLKEHYQLNNKLAIPATSLRGLISSLAEEAASNSALRVLDN GVLSYRKPAASRALRKIGILFKREEQWRLVQMEGNLANAIKLSAYTNQKMMDFLAN

SEQ ID NO & Protein ID/Contig	Sequence
	KQSWSPHEHNVVYYLSADFRPGDVPQETYLAGRICGILRILGGKDGDRKNELENKKHE LFIRVDEQYVDTEINRFDYEEYVRQGGIPVSPSAVERYTELADQSRSLQKNSRDLKGD NNCCSDKWLPFHLKGAARYKKEKACLLPLREYDLVYFSDGTQVTEISFSAIWRDRV ADKVHAFPEELRPFNQKRKWISPAELLFGFVELNDNKDERDHAQFTGKVRVAAG VLSPDDSIHQDLQEHEPIMLKALSSPKLPSPALYFKQKSGDHRYIAKPDLKKASHQA QGRKIYLHALRDQKDDVQKLNTKGQPANGNGAHLWPWKTADEDERPQLKVRIRPLKP GTSFYFHLDYNNLTEWELGLLCYVLQPSETFRHKLGMGKPIGLGTVKIEIATLQITDR QKRYREAGANEHRHNGSNWVNESLRDELERLPGTVELSPDRQPEAKLRPDELRSQFI ATMDNDIYRAIELLGDPHNIKYPVHYPQVRNKSIEQENFKWVANDSGSGDQRKGTG IDAKEEPMRSIDQISTTIPTLNRYEWNGD
SEQ ID NO: 28 DTXS0100007 0.1	MARNNKQYHFIPRWEIKVNLTTTSFLHIGCDEFTDRPGLEIEQKDGSKVKAENAFIKD SNGKPYLPGSTIKGNIRKWLETNKKADEETCKLFNTLLGFTVKMQDEGCGGSAEFHN AVISSPLEDGNNFPYWDVLDQTSVETSTVIDRVTGTVDGRLFSTEVVPPEVSFTLIITG AMTEQQVSLLTAVIKDGAEDCPTITIGADSGNGFGRFRFDSIHMCKLGTGEVLNWL EDGSQDMAATAMRSLSPDDIEQHIIKGRNYLKSPTSVDVTIEFGFAGPFLVNDPSRKK RKEDIDHQPLRDSAGNARLPKASIRGAMRSQAEKIIRTLGGWCCDPVNPSPSVFVVEI NDRLCLACRVFGATGWKSRISIQKVEYKGTAESTRQETVQDFVAIDRFHGGGKETAK FDASFSWRPQYSILMHIPSDLEGWAKGLLALTRDFKEGDFLGYGRSKGYGRVDSDS VKPGIDTMLTESNLELFRRCDDNPGEYPCPTRQPPNLVQPVERNNLTEAAGEGSFH NPYHFIPTPKPMIESWLAKEDFDETMHDSHALYRDVDENEELPHYGKISCTLTETPV FVGGKHDPNDTEPQQVDHYTENGEIAIPATTLRGLLSSLSEAASNSMRVLDDGMM SYRQPVGSGSLSAIGMVVIRDGKKFIYPLALPIFGERDKLPQEYHIMFPYTQKAPLKVY LERAYLAGNMKSFLDKQNSWNLLNEKIFYLPVPEFSFSRVHTMGAENRDVLKISRGR NLILGARLPVNLCPRSKEKALPGDIPGILRILGKEGRDGEVPVGGKHHELFPVSDGFASN PRSFIDNLTSKELFKIPDEVDRFEELADDRTTQIKHPGNVKNNNQWLPHLKGCTR NDGLTGKDEKRLRVQEGDLMYFRPSPQSPQVAEISFSAVWRGRVKNKTVHNYFPELV QFNKNREKISPAELLFGFVQQDKHEKSLSFAGKVVLSSGKQLRETESVSRENEVTILKIL ASPKLPSPSLYFKRENYIEGGNYIAKNEMNNSNIKPRGRKQYLHALSNSDEPKGVQK ISRTGSVDDGGNYPWQSMNNDNIKQKVCIRPVSKDGCFTFEMEFENCTEWELGMLLY ALRPSQQYRHKIGMGKSIGLGTVRIDINNLFQIHRKNRYNAGIIDVPRYNYEAGHDMD YFHNKFADTIMPEIKNSIELLGDPRNVRFVHYPQVHGADIEDKTYQWVANDSGTN NGQNGAAYKKKNAEESSTELDEISNTIPGLERHEWLGR
SEQ ID NO: 29 JABFST01000 0317.1	MALKTWTLNGEERWHISVVLTEVTPLHIGSGEFCYRPELTNADQKPDINACIKGAN NLPIPGSTVKGKFNAWLTAQVDTPLEAIFGKGHNPDDDDDQSGGKVEFHDAWIST KIKDTSTWPYQVATQTFIDAATAIDRHSRTALDASLHYTECVPPGVQFTLNITGVM QEHEAALIIAALDRFDQHDDQPYFGAGDANGQGQLILVGH LAVKVMGKTEITEWLA HFNNKASDMAMSHARSLGAEDIAGLIKLGQTLTKVPPTVSLGIQLQFAGPFLVNDPY AVKKLEADPKTKIDHYPLLDNHKKPRLPSASIRGVLSQAERIIRSLGVHCCDTRDPCP SLYKHQDLSQLCLACQIFGAAGWKSVINISDFTCDANELKTQEFIAIDRFHGGGKDG AKFNAKHSERPYFQGRITLSPRMANHQLDWGKGLLALVIRDLQEGDLSFGFGANKGY GALESVLITGIDQLQTD AIEAFRRLCVTQAAPQAFITPTS AVVIGDKAPLVVTDKCLPD NSFHNPYHFIPINSPDTRHWLPTETDLAESHSHAYYRQQPELFHGGQLICRLYTETPTFI GASKKDDTLPAELDNYRLNGQLAIPATSLRGMISSLAEAAASNSAMRVLDNGLLSYRK DASLALSKIGITFINRQGGWQWLPMEKIKLKNAYSAENMRLFVEQSHSWSPDYNTVYY FSEKAGAFDVPQRTPKPGWQPGILRLLGKEGRSQELENKKHEWFIPVPENYIDKQLNA FKYQEYLKDNSSKAIDIPAPVLNRYNELAYQRTLSQKKDTEL VADGDSAPWLPHLK GQQRQPMVVGKHLVYTLPMTEYSLVYYAATNKVATEISYSSIWRGRVQDDADQAAT VNHFIPDILLPFNPKRTSLSPAELLFGFTELDPDKHSNDPSTRSFAGKVRIGAATLAAYP SNDSDLLAPEHITLALSSPKLPSPALYFRTLQGNNSNVYIPKHELNPNHHTAKGRKY YLHATRTPDQKRILKLSQGHPPQNNAVKLPWLSSHQETKNLQKVKIKPIKPKQSFYF QVDFNNLTAWELGLLCYALRPTIDFRHRIGMGKPLGLGSVKIDILALQTLDRQKRYA QDSQDSARYNQHRWVNSSVTDMLAQAGYDVIEPTANPLVPKDLKTLFSQTMAANID RALTLLGEPQHVQPVHYPQVRDTAIQVRDTAIEEESYQWVANDNLSDNSSAAKQT LHDITETSEGLPTLIRHQKKKETQP
SEQ ID NO: 30	MSDTQKQAIHENKWHFRGIKRWEISAYLKTLSPLHIGDGGTIPVTIKDTQGKNREVEV NSVITGKAALPIPGSTIKGRLRHYFSKHFSKALLNKVFGEESDATDDDQGRGGLAEF

SEQ ID NO & Protein ID/Contig	Sequence
PDPY0100000 1.1	HDAKWNPENRNLQGRYPYWNNTKTYIEVSTAINRHTGAAKDKSLHTECVPPGT VFEIKITGSMDDRCAALVVAALAIQTTGSRIFLGAEDANGNGRIGLTGKITVKQMDQ AHIIQWLQKDSTTCVASFSNVKAENETQVKQMVQRHIAPKLSNVVSAAGPSYDITLHF DGPFVVNDSDKCKAEDTPDIYPLEEKNGVPAFPVRSFRGAIRSQAERIIRTIGGCCDG SINNTCKNPKNLCAIACEMFGSTGWKTSIEMDPFLCVDRELKPFIIQEFVAIDRFHGGGK DEAKFNAAHYQAPVFKGKVRVSQRVGNDSWRKGLLALIFRDLKEGDIYFGFGTNKG YGAVKKAEPDGNASDFSESIEAFINKCREKKGLYCNPIKKPGKTKVSKNLPPAIV PLDRDTSKFYNPYHFIPVKKPNTSSWAEKTAFTGADSPHSHGFYRKQTNEQQPLYSGR LICMLTSETPFFIGAQAESDPTENENQASLRHPYQLDGEPAPSTSLRGLISTMTEAAAN CAMRVLDSIISYRKPMNPISHLSALGMVTKRGEDFWLIPLAMPALSLNDEEHNYKLD KRYRTMFPDGLAKLKVYLEKAYSNNVMKTFLLNNENTWTLAQSKIHYPPLTPIQMQN GGINSYYNNLRTPSRSNFLIGQTVAHGNGIPASGPGAGMVPGLRILGKEHRQNDLP QNKKHELFIQVPAFAVADPKTFLDTATAFLIPRNVIDAFEKIAEKQTQSQKQDKLKH EERLPFHLKGTRREQNHTLQIKTGDLVYFRPNAKGDEVEEIAFSSIWRGKTS GTTADFF PDKELLPFNRNRSRVSPAELLFGFTENNPKEMKIDRGLAFAGKIRISAGTSLDKFSDTT ESDLFEPETTLKALSSPKPPSPALYFKEKSGTQYIKKQDLNPGKHEIQGRKIYHLALR NENNQNVQRITSQGKFDNAANRTQPVWSQNEERNHLKTKCKPLKSGLNFFHFIDFNN LTQWELGLLCYALRPCETFRHKIGMGKPIGLGTVKIDIAVLQDTIDRYARYTDTTQDSE RYNQGAWISQELQNEIPNQYKKGKISNKKGMLSPEDCRKVMETMDADIQRAIELLG DPGNVTSPVHYPQLDRKNIETKNYEFWKQNEIEQQVLKPITKNTHTLTPFARWEQG
SEQ ID NO: 31 NZ_JMLA0100 0001.1	MNLPTWKLNNKRWHSIVLTTATPLHIGSGEFCEHDDVKNNDGEPVKINACIKGSKG RPIIPGSTIKGKLYEWLKTNRTEENLLEKLFKGKHNVSQDQGRGGKAEFHDAEIIIEPL TGSQPWPYWREEHQAFIAASTAIDRHKQVALQQSLHYMETVPAGIRFKFTFTGVMRD EEAALLIAALDSFDKNQNPQCFGVDRANAYGRMELHGHLHVKVMGATEISSWLNSF SENDKKMAMESARNLEQQEINTLIKQGNALFKASCDEVKLGLTLKFKGPFLVNDPYA VKILSSNENAKTDHYPLLDKNRNPYLPVSSFRGVLSQAERIIRTLGGKCCSTDDPCKP IFDKGDL SKLCLACQIFGASGWKTVINIHDFKAINKSKKTKQDFVAIDRFHGGGKDGA KFDATHFERPEFEGAISFSRMANNDLDWGKGLLALVLRDMQEGDMFTFGYGANKGY GGLESASITGIEQITSIDQAFRDKCVASPQTWLCDEAVK PANQQDKIPPAGIQVANSGF HNPYQFIPSKDPDTGHWLPVLGLNADSHSHAFYRDQTDNGEKL YHGRICLCLNTET PIFIGADKKKDTPEAEINNYRLNGELAI PATSLRGMISLAEEAASNSAMRVLDNGLLSY RKTADDALRKVGMVIYVDNKSFIKLNDAIKLLQTYTPGNMKDFIEKSNWSPEHNT VYYLDNNQIPQESYMNGMKPGILRILGKEGRELEENKLHELFIQVPLEYVDTENNKF DYQAYKKAFLYRAIEIPEPVLKRYSELADQRTMSQKSNKELKDDTCQSVGWLPFHL KGTKRQLDDKHKVGKLQIDEYDLIYYEASGKEVTEVAFSSIWRGRVETNSSQANKVY SFIPGELLPFNERSKKVSPAELLFGFTQINKDGSKADDKAQAFAGKVRISAGTISEYPES EANLLEQEVTLKALSTPKLPSPALYFRTINGNGSAYISKQELEPSKHLAKGRKYYLHA LRTGDNKVQKLGSQGETANGGDSKLPWVTHNPDERPQLKVKIKPIAEFISLDFNNL TEWELGLLCYALRPTDSFRHRIGMGKPLGLGSVKIDIMALQTINRQQRYAQDGLEEN RFNRHNWVNPPHQPRLDKAGYSISLSSTPLNPEILRATFTKTMNADIYRTLELLGNPQ NVKRPVHYPQVENHNIEQENYKWFVANDQSGSGKGRNKIDPAEKALKILTENSCLPT LSRLDWRDE
SEQ ID NO: 32 NZ_FOGH010 00010.1	MNNKGSNMTDTPVSGRWITGQFQLVTPMHIGTGLDEEMDKQSGESVDKKQNNSWI QAIALDLNKKPYIPGASIKGALKALARYYCASNINIFGDTIDTKDGDNRKSVTVAG QAEFLNAWYAADQEDKPFDTITRVAIDRVGTAEADRKLFNTRRVNPGVCFNYKIIQN ACETEIQYLLDLLRKAADPSFSLGAGANQSQGVRLSSCVRYFGKQEMHDFWRAI QNGKQEHWQIFAKPSNIKYADLERPDIIANSNLPLTLDFTPLVKASKKKDEAKNE ADAKPRTNHQGVILPASSLRGRLRAQAEKILRTMGQDIPQGHAAAPAYDGHARDLIS LLFGTAGWKGIVCASDLHSIPEYALQFNGVRETISDLSDTVKSCIIVDLVKTSTAIEKT EEQLHIRIVDSAGSLIVHKSSENSSWANDTFRDASVKDNFKARLKEIADPQDLSALRA DIKKRAFQLATLTRHEMVAIDRFTGGGKEGAKFNVDYIECPTLTGAIYLDLHRLKQA QLKNDLALKPALGLISLLRDLAEGDIAFGFGANKGYGQCREHAVLDNWEERLKKI GAGLTIDGALQALRDTVALEPPAEFPPEIEKTTDDNQPEAPDFNLKPASNGFHNYPYFI PLNNPKIGDWPEAKAETLKANREGHDQYHTGKFSGRIVCSLTQTPLFIGAETKPS DREPSEARPFKLNGKHAIPATSLRGMSSLFESVSNSNFRVLHPEHYSVRKSLDDYVA LSAMGRIVDDQGELKLQPLTLPFLFGNRRNNVPAKWEKIFGTPSEDDFLRIYFDDIPSKF SSNKRYFYNCKATELKDFIKSDKYFIGKRTPTVPFKSSTEKSHLESLEFIDVEFKKAV

SEQ ID NO & Protein ID/Contig	Sequence
	ENLEITPGNNPYIHGWVRNLKDEFREDIPDNVKKHHVFLPDTTKRVSPLIEIPPHVKKRFH ELADLALAGLHLKQGETIASPYKILPYTPIGRNKLENHIHRVPNDLTCYMTRLKKGDL VFFDVDNDGQITEISFSSIWRAGIGTKNKLQTTADLLSQRDPNLVQLGMGVRTKNTDR FKLSAERLFGVVEHRDDNTTVENVNQPNDKAQAFAGKVRIGFGLPDKKTTVN GVSPVTLKELSSPKPPSPAFYLKRKNNDDFVSKKVAAECSETMTLRGRKCYLHAWRE QNGNVMKLDAGVNSGGSTCKPPWKTHKPAANDQKEFEEDKNKFITSRQVKIAPISE NTPFYFEIDFNNLDATELAQLCATLQAPKFEHRLGMGKPLGLGSVKIEPVGLFLINRH QRYTTDSTNCDRYHYAWLKGEHAAWDWPEYFRQNVVTADCTQTFNDTFDKLVQN GLAGTDADIKHALQLLGDPOYIGVPVHYPIAGNSTLENKHFEWFGNNDKASVLRQKA QANSKNHHYQPKQQATPEEPQYLHTITKDSKQISLLKKNKIEDIENRDQQKHRYSNHR R
SEQ ID NO: 632 MVRP0100010 4.1	MFPKGRQMRRQRLLGDAEYYGGTGREQPASIVISTSDSPDHK VYEWIITGQLKAETG FFFGTKAGAGGHTDLSILLGKDGHYRVPRSVFRGALRRDLRVAFGAGCRVEVGRERP CECPVCKVMRQITVMDTISSYREAPEIRQIRLNPYTGTVDK GALFDM EVGPEGIEFPF VLRFRGSKSFPSELA AAVIGSWTKGTAWLGGAATGKGRFSLGLSIHKWNLSTAEGR KSYLAAYGLRDAADKTVKRLSIDKGGKGDVGLPAGLERDALPSSVREPLWKKLVCT VDFSSPLLLADPIAALLGVEGDERIGFDNIAYEKRRYNGETNTTESIPAVKGETFRGIVR TALGKRHGNLTRDHEDCRCLCAVFGKEQEAGKIRFEDLMPVGAWTRKHLDHVAID RFHGGAEENMKFDTYALAASTPNPLRMKGLIWVRSDLFETGHDGPTPPYVKDIIDAL ADVKRGLYPVGGKTGSGYGWIKDVTIDGLPQGLSLPPAEERVDGVNEVPYNY SAPP DLPSAAEGEYFFPHVFIKPYDKVDRVSRLTGHDRFRQGRITGRITCTLKTLTPLIIPDSE GIQTDATGHKMCKFFSVAGKPMIPGSEIRGMISSVYEALTN SCFRVFDEEKYLTRRVQ PKKGAKSSELVPGIIVWGQNGGLAVQQVKNA YRVPLYDDPAV TSAIPTEAQKNKER WESVPSVNLQGALDWNLT TANIARDNR TFLNSRPEEKDAILSGTKPISFELEG TNPND MLVRLVPDGV DGAHSGYLKFTGLNMVLKANKKTSRKLAPSEEDVRTLA IHLNDFDS RRDWRPPNSQRYFPRSVLRFSLERSTYTIPKRCERVFEGTCGEPYSVPSDVERQYNSII DDISKNYGRISETYLT KTANRKLTVGDLVYFIADLDKNMATHILPVFISRISDEKPLGE LLPFSGKLIPCEGPPTILKKMAPSLLTEAWRTL ISTHLEGFCPACRLFGTTSYKGRIRF GFAEHTGTPKWLREELDWARPFLTLP IQERPRPTWSVPDDKSEVPGRKFYLHHHGGN RIVESNLRNRPEVNQTKNNSSVEPISAGNTFTFDVCFENLEAWELGLLLYCLELSPKLA HKLGRAKAFGFGSVKIHVERIEERTTDGAYQDVTAVKKNGWITTGHDKLREWFHRD DWEVDHIRNLR T VLRFPDADQEHDVRYPELKANNVSGYVELRDKMTASERQESL RTPWYRWFPQNGTGGSGRHEQAATSQEQDTAKDES VLSATQRRQAVIDVSDPDERL SGTVESFDRQKGDGYIGCGVRQFYVRLEDIRSRTALCEGQVVTFRARKEWEGHEAYD VEIDQ

[0286] Guide Molecules

[0287] The system may comprise a guide molecule. The guide molecule may comprise a guide sequence. In certain cases, the guide sequence may be linked to a direct repeat

- 5 sequence. In some cases, the system may comprise a nucleotide sequence encoding the guide molecule. The guide molecule may form a complex with the dead Cas7-11 protein and directs the complex to bind the target RNA sequence at one or more codons encoding an amino acid that is post-translationally modified. The guide sequence may be capable of hybridizing with a target RNA sequence comprising an Adenine or Cytidine encoding said amino acid to form
- 10 an RNA duplex, wherein said guide sequence comprises a non-pairing nucleotide at a position corresponding to said Adenine or Cytidine resulting in a mismatch in the RNA duplex formed. The guide sequence may comprise one or more mismatch corresponding to

different adenosine sites in the target sequence. In certain cases, guide sequence may comprise multiple mismatches corresponding to different adenosine sites in the target sequence. In cases where two guide molecules are used, the guide sequence of each of the guide molecules may comprise a mismatch corresponding to a different adenosine sites in the target sequence.

[0288] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence. In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target DNA sequence and a guide sequence promotes the formation of a CRISPR complex.

[0289] In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site); that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cas7-11 protein used, but PAMs are typically 2-8 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different Cas7-11 orthologues are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cas7-11 protein. In certain embodiments, the Cas7-11 protein has been modified to recognize a non-natural PAM, such as recognizing a PAM having a sequence or comprising a sequence YCN, YCV, AYV, TYV, RYN, RCN, TGYV, NTTN, TTN, TRTN, TYTV, TYCT, TYCN, TRTN, NTTN, TACT, TYCC, TRTC, TATV, NTTV, TTV, TSTG, TVTS, TYYS, TCYS, TBYS, TCYS, TNYS, TYYS, TNTN, TSTG, TTCC, TCCC, TATC, TGTG, TCTG, TYCV, or TCTC.

[0290] The terms "guide molecule" and "guide RNA" are used interchangeably herein to refer to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprises a guide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of the complex to the target nucleic acid sequence. The guide molecule or guide RNA specifically encompasses RNA-based molecules having one or more chemically modifications (*e.g.*, by chemical linking two ribonucleotides or by replacement of one or more ribonucleotides with one or more deoxyribonucleotides), as described herein.

[0291] As used herein, the term "guide sequence" in the context of a CRISPR-Cas system, comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In the context of the present invention the target nucleic acid sequence or target sequence is the sequence comprising the target adenosine to be deaminated also referred to herein as the "target adenosine". In some embodiments, except for the intended dA-C mismatch, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more.

Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (*e.g.*, the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (*e.g.*, cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence.

[0292] In some embodiments, the guide molecule comprises a guide sequence that is designed to have at least one mismatch with the target sequence, such that an RNA duplex

formed between the guide sequence and the target sequence comprises a non-pairing C in the guide sequence opposite to the target A for deamination on the target sequence. In some embodiments, aside from this A-C mismatch, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In some cases, the distance between the non-pairing C and the 5' end of the guide sequence is from about 10 to about 50, *e.g.*, from about 10 to about 20, from about 15 to about 25, from about 20 to about 30, from about 25 to about 35, from about 30 to about 40, from about 35 to about 45, or from about 40 to about 50 nucleotides (nt) in length. In certain example. In some cases, the distance between the non-pairing C and the 3' end of the guide sequence is from about 10 to about 50, *e.g.*, from about 10 to about 20, from about 15 to about 25, from about 20 to about 30, from about 25 to about 35, from about 30 to about 40, from about 35 to about 45, or from about 40 to about 50 nucleotides (nt) in length. In one example, the distance between the non-pairing C and the 5' end of said guide sequence is from about 20 to about 30 nucleotides.

[0293] In certain embodiments, the guide sequence or spacer length of the guide molecules is from 15 to 50 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, *e.g.*, 15, 16, or 17 nt, from 17 to 20 nt, *e.g.*, 17, 18, 19, or 20 nt, from 20 to 24 nt, *e.g.*, 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, *e.g.*, 23, 24, or 25 nt, from 24 to 27 nt, *e.g.*, 24, 25, 26, or 27 nt, from 27-30 nt, *e.g.*, 27, 28, 29, or 30 nt, from 30-35 nt, *e.g.*, 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer. In certain example embodiment, the guide sequence is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 nt.

[0294] In some embodiments, the guide sequence has a length from about 10 to about 100, *e.g.*, from about 20 to about 60, from about 20 to about 55, from about 20 to about 53, from about 25 to about 53, from about 29 to about 53, from about 20 to about 30, from about 25 to about 35, from about 30 to about 40, from about 35 to about 45, from about 40 to about 50, from about 45 to about 55, from about 50 to about 60, from about 55 to about 65, from about 60 to about 70, from about 70 to about 80, from about 80 to about 90, or from about 90 to about 100 nucleotides (nt) long that is capable of forming an RNA duplex with a target sequence. In certain example, the guide sequence has a length from about 20 to about 53 nt

capable of forming said RNA duplex with said target sequence. In certain example, the guide sequence has a length from about 25 to about 53 nt capable of forming said RNA duplex with said target sequence. In certain example, the guide sequence has a length from about 29 to about 53 nt capable of forming said RNA duplex with said target sequence. In certain example, the guide sequence has a length from about 40 to about 50 nt capable of forming said RNA duplex with said target sequence. In some examples, the guide sequence comprises a non-pairing Cytosine at a position corresponding to said Adenine resulting in an A-C mismatch in the RNA duplex formed. The guide sequence is selected so as to ensure that it hybridizes to the target sequence comprising the adenosine to be deaminated.

[0295] In some embodiments, the guide sequence is about 10 nt to about 100 nt long and hybridizes to the target DNA strand to form an almost perfectly matched duplex, except for having a dA-C mismatch at the target adenosine site. Particularly, in some embodiments, the dA-C mismatch is located close to the center of the target sequence (and thus the center of the duplex upon hybridization of the guide sequence to the target sequence), thereby restricting the nucleotide deaminase to a narrow editing window (*e.g.*, about 4 bp wide). In some embodiments, the target sequence may comprise more than one target adenosine to be deaminated. In further embodiments, the target sequence may further comprise one or more dA-C mismatch 3' to the target adenosine site. In some embodiments, to avoid off-target editing at an unintended Adenine site in the target sequence, the guide sequence can be designed to comprise a non-pairing Guanine at a position corresponding to said unintended Adenine to introduce a dA-G mismatch, which is catalytically unfavorable for certain nucleotide deaminases such as ADAR1 and ADAR2. See Wong *et al.*, RNA 7:846-858 (2001), which is incorporated herein by reference in its entirety.

[0296] In some embodiments, the sequence of the guide molecule (direct repeat and/or spacer) is selected to reduce the degree secondary structure within the guide molecule. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%), 1%), or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see *e.g.*, A.R.

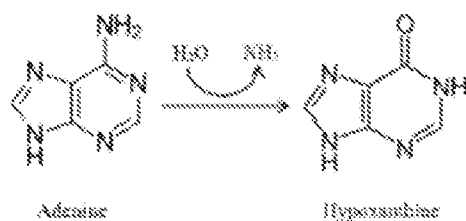
Gruber *et al.*, 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0297] In some embodiments, it is of interest to reduce the susceptibility of the guide molecule to RNA cleavage, such as to cleavage by Cas7-11. Accordingly, in particular
5 embodiments, the guide molecule is adjusted to avoid cleavage by Cas7-11 or other RNA-cleaving enzymes.

[0298] In some embodiments, the guide molecule is modified, *e.g.*, by one or more aptamer(s) designed to improve guide molecule delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can
10 include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide molecule deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends a guide molecule that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O₂ concentration, temperature, protein concentration, enzymatic concentration,
15 lipid structure, light exposure, mechanical disruption (*e.g.* ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0299] Adenosine Deaminase

[0300] The system may further comprise an adenosine deaminase or catalytic domain thereof. The adenosine deaminase protein or catalytic domain thereof deaminates an Adenine
20 or Cytidine at the one or more codons thereby changing the codon to encode for an amino acid that is not post-translationally modified. The term "adenosine deaminase" or "adenosine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts an adenine (or an adenine moiety of a molecule) to a hypoxanthine (or
25 a hypoxanthine moiety of a molecule), as shown below. In some embodiments, the adenine-containing molecule is an adenosine (A), and the hypoxanthine-containing molecule is an inosine (I). The adenine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



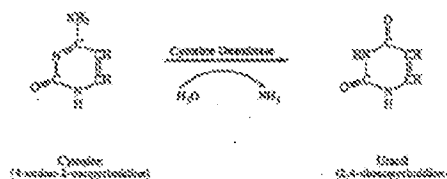
[1]

[0301] According to the present disclosure, adenosine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as adenosine deaminases that act on RNA (ADARs), members of the enzyme family known as adenosine deaminases that act on tRNA (ADATs), and other adenosine deaminase domain-containing (AD AD) family members. According to the present disclosure, the adenosine deaminase is capable of targeting adenine in a RNA/DNA and RNA duplexes. Indeed, Zheng *et al.* (Nucleic Acids Res. 2017, 45(6): 3369-3377) demonstrate that ADARs can carry out adenosine to inosine editing reactions on RNA/DNA and RNA/RNA duplexes. The adenosine deaminase can be modified to increase its ability to edit DNA in a RNA/DNA_n RNA duplex.

[0302] In some embodiments, the adenosine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the adenosine deaminase is a human, cephalopod (*e.g.*, squid) or *Drosophila* adenosine deaminase. In certain examples, the adenosine deaminase is a human adenosine deaminase. In certain examples, the adenosine deaminase is a cephalopod adenosine deaminase. In certain examples, the adenosine deaminase is a *Drosophila* adenosine deaminase.

[0303] Cytidine Deaminase

[0304] The term "cytidine deaminase" or "cytidine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts a cytosine (or a cytosine moiety of a molecule) to an uracil (or an uracil moiety of a molecule), as shown below. In some embodiments, the cytosine-containing molecule is a cytidine (C), and the uracil-containing molecule is an uridine (U). The cytosine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



[2]

[0305] According to the present disclosure, cytidine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1). In particular

embodiments, the deaminase in an APOBEC 1 deaminase, an APOBEC2 deaminase, an APOBEC3A deaminase, an APOBEC3B deaminase, an APOBEC3C deaminase, and APOBEC3D deaminase, an APOBEC3E deaminase, an APOBEC3F deaminase an APOBEC3G deaminase, an APOBEC3H deaminase, or an APOBEC4 deaminase. The cytidine deaminase can be modified to increase its ability to edit DNA in a RNA/DNA RNA duplex.

[0306] In some embodiments, the cytidine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the cytidine deaminase is a human, primate, cow, dog, rat or mouse cytidine deaminase.

[0307] CD (cytidine deaminase)-functionalized CRISPR system for RNA editing can be used for C to U conversions. In some embodiments, the cytidine deaminase protein or catalytic domain thereof is a human, rat or lamprey cytidine deaminase protein or catalytic domain thereof. In some embodiments, the cytidine deaminase protein or catalytic domain thereof is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1). In some embodiments, the cytidine deaminase protein or catalytic domain thereof is an APOBEC1 deaminase comprising one or more mutations corresponding to W90A, W90Y, R118A, H121R, H122R, R126A, R126E, or R132E in rat APOBEC1, or an APOBEC3G deaminase comprising one or more mutations corresponding to W285A, W285Y, R313A, D316R, D317R, R320A, R320E, or R326E in human APOBEC3G. In some embodiments, the cytidine deaminase protein or catalytic domain thereof is delivered together with an uracil glycosylase inhibitor (UGI), where said UGI is covalently linked to said cytidine deaminase protein or catalytic domain thereof and/or said catalytically inactive Cas7-11 protein.

[0308] Cas7-11-APOBEC fusions can perform C-to-U editing of RNA. APOBEC substrates are ssRNA and the Cas7-11-APOBEC can therefore target regions of the RNA around the guide/target duplex. Cas7-11-APOBEC fusions can perform C to U knockdown via stop codon introduction. In addition to correcting pathogenic U to C mutations that arise during the cellular life cycle, Cas7-11-APOBEC fusions can lead to the introduction of stop codons by converting a CAA, CGA, or CAG to TAA, TGA, or TAG, respectively. APOBEC orthologs in fusion with Cas7-11 can increase the efficiency of C-to-U editing, or can allow for additional types of base conversions. Mutating the APOBEC from the Cas7-11-APOBEC can lead to fusions with specific dsRNA activity, base flip activity and increased activity.

[0309] In certain example embodiments, the deaminase is selected from Table 2.

Table 2.

SEQ ID NO & Deaminase Name	Protein Sequence
SEQ ID NO: 33 Homo sapiens_ADAR2_E_Q_Mutant	QLHLPQVLADAVSRLVLGKFGDLTDFSSPHARRKVLAGVVMTTGTDVKDAKV ISVSTGTKCINGEYMSDRGLALNDCHAEIISRRSLLRFLYTQLELYLNNKDDQKRS IFQKSERGGFRLKENVQFHLIYSTPCGDARIFSPHEPILEEPADRHHPNRKARGQLR TKIESGQGTIPVRSNASIQTDGVLQGERLLTMSCSDKIARWNVVGIQGSLLSIFV EPIYFSSIILGSLYHGDHLSRAMYQRISNIEDLPPLYTLNKPILLSGISNAEARQPGKA PNFSVNWTVGDSAIEVINATTGKDELGRASRLCKHALYCRWMRVHGKVPShLL RSKITKPNVYHESKLAAKEYQAAKARLFTAFIKAGLGAWVEKPTEQDQFSLT*
SEQ ID NO: 34 Homo sapiens_ADAR1_E_Q_Mutant	SLGTGNRCVKGDSLKGETVNDCHAEIISRRGFIRFLYSELKYNSTAKDSIFE PAKGGKELQIKKTVSFHLYISTAPCGDGALFDKSCSDRAMESTESRHYPVFENPK QGKLRTKVENGGGTIPVESSDIVPTWDGIRLGERLRTMSCSDKILRWNVGLGLQGA LLTHFLQPIYLKSVTLGYLFSQGHLLTRAICCRVTRDGSFEDGLRHPFIVNHPKVG RVSIDSKRQSGKTKETSVNWCADGYDLEILDGTRGTVDGPRNELSRVSKKNIF LLFKKLCSFRYRRDLLRLSYGEAKKAARDYETAKNYFKKGLKDMGYGNWISKP QEEKNF
SEQ ID NO: 35 Octopus vulgaris_ADAR1_E_Q_Mutant	SVGTGNRCLTGDHLSLEGNSVNDSHAEMITRRGFLRYLYRHLLLEYDAEVPNDLF EKGERSICRIKTNITFHLIYSTAPCGDGALFSPRDTDSSNAKMEENKHIHNPTFSS SVQGLLRRTKVEGGGTIPIDADFTEQTWDGIQGERLRTMSCSDKICRWNVVGL QGALLSHFIEPIYLDLSTLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNH PWLGRVTACDPPRETQKTKSLINWCYDDEKSEVLDGTAGICYTAIEKNLFSRLT KHNLYEEFKRVCRCFDRNDLLTAPSYNKAAMMATPFQTAKNVMLKKLKENNC GTWVSKPIEEEMF
SEQ ID NO: 36 Sepia_ADAR1_E_Q_Mutant	SVGTGNRCLTGDRLSLEGNSVNDSHAEMVTRRGFLRYLYKHLLEYDPEKPHDLF EKGERSLCRIKTNITFHLIYSTAPCGDGALFSPRDTDSSNVKVEENKHVHNPTFS SSVQGLLRRTKVEGGGTIPIDADFTEQTWDGIQGERLRTMSCSDKICRWNVVGL QGALLSHFVEPIYLESLLGYLYDHGHLARAVCCRIERGEASVNQLLPEGYRLNH PWLGRVTACDPPRETQKTKSLINWCYDDEKSEVLDGTAGICYTAIEKNLFSRLT KHSLYEEFKKVCQKFEREDLLNVTSYNKAAMMAIPFQTAKNVMLKKLKENNC TWVSKPIEEEMF
SEQ ID NO: 37 Octopus vulgaris_ADAR2_E_Q_Mutant	GIGTGTKCINGEHMSDRGFGVNDCHAEIARRCFLRYIYDQLEHLSDNSDVRNSS IFELRDGGYQLKENIQFHLIYSTAPCGDARIFSPHGQDVETGDRHPNRKARGQL RTKIESGQGTIPVRTSGVIQTDGVLGERLLTMSCSDKIARWNVVGIQGSLLSHF MNPIYLESIILGSLYHSDHLSRAMYSRISIENLPEPFHLNRPFLSGISSPESRQPGKA PNFNGINWRKEDETFEVINAMTGRVEGGSVSRICKQALFGRFMSLYGKLSSLTGQS VTTRPTHYSDAKAAVMEYQLAKQCQVFQAFQKAGLGNWVQKPIEQDQF
SEQ ID NO: 38 Sepia_ADAR2_E_Q_Mutant	GIGTGTKCINGEYMNDRGFVNDCHAEIARRCFLRFIYDQLEMHLSEDPEVRGQ SVFELRDGGGYKLKPNIHFLIYSTAPCGDARIFSPHGQDVETGDRHPNRKARGQ LRTKIESGQGTIPVRSSGFIQTDGVLGERLLTMSCSDKIARWNVVGIQGALLCH FMHPIYLESIILGSLYHSDHLSRAVYCRISAIENLPDLFQLNRPFLSGISSPESRQPG KAPNFGINWRRNDDTFEVINAMTGRVEGGNMSRICKQALFDRFMNLYGKLSSLT GQSVTTRPTLYSEAKAAVMEYQLAKQCQVFQAFQKAGLGNWVQKPIEQDQF
SEQ ID NO: 39 Doryteusthis opalescens_ADAR2_E_Q_Mutant	GIGTGTKCINGEYMNDRGFVNDCHAEIARRCFLRFIYDQLEHLSDNAEVRGQ SIFELRDAGGYKLKPNIQFHLIYSTAPCGDARIFSPHGQDVETGDRHPNRKARGQL RTKIESGQGTIPVRSSGFIQTDGVLGERLLTMSCSDKIARWNVVGVQGALLCH FMHPIYLESIILGSLYHSDHLSRAVYCRISAIENLPDLFQLNRPFLSGISSPESRQPG KAPNFGINWRRNDDTFEVINAMTGRVEGGNMSRICKQALFDRFMNLYGKLSSLT GQSVTTRPALYSEAKATVMEYQLAKQCQVFQAFQKAGLGNWVQKPIEQDQF
SEQ ID NO: 40 APOBEC1	MTSEKGPSTGDPFLRRRIEPWEFDFVYDPRERLKEACLLYEIKWGMRSRKiWRSSG KNTTNHVEVNFIFKFTSERDFHPSMSCSITWFLSWSPCWECQAIREFLSRHPGVT LVIIYVARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDE AHWPQYPPPLWMMMLYALELHCHLSLPPCLKISRRWQNHLLTFRLLHQNCHYQTP

	PHILLATGLIHPSVAWR
SEQ ID NO: 41 APOBEC2	MAQKEEA AVATEAASQNGEDLENLDDPEKLLKELIELPPFEIVTGERLPANFFKFQ FRNVEYSSGRNKTFLCYVVEAQGGQVQASRGYLEDEHAAAHAEAEFFNTILP AFDPALRYNVTWYVSSSPCAACADRIIKTLSTKTNLRLLLVGRLFMWEEPEIQA ALKKLKEAGCKLRIMKPQDFEYVWQNFVEQEEGESKAFQPWEDIQENFLYYEEK LADILK
SEQ ID NO: 42 APOBEC3A	MEASPASGPRHLMDPHIFTSNFNNGIGRHKTYLCYEVERLDNGTSVKMDQHRGF LHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPAQIYRVTFWISWSPCFSWGCAG EVRAFLQENTHVRIRFAARIYDYDPLYKEALQMLRDAGAQVSIMTYDEFKHCW DTFVDHQGCPFPQWDGLDEHSQALSGRLRAILQNOGN
SEQ ID NO: 43 APOBEC3B	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGV FRGQVYFEPQYHAEMCFLSWFCGNQLPAYKCFQITWVFSWTPCPDCVAKLAEFL SEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDYEEFAYCWENFV YNEGQQFMPWYKFDENY AFLHRTLKEILRLRIFSV AFTAMRSCASWTWFLCS WTRPRSTGSLGSSPGAPSPGAVPGKCVRSFRRTHT
SEQ ID NO: 44 APOBEC3C	MNPQIRNPMKAMYPTFYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKTG VFRNQVDSETHCHAERCFLSWFCDDILSPNTKYQVTWYTSWSPCPDCAGEVAEF LARHSNVNLTIFTARLYYFQYPCYQEGRLSLSQEGVA VEIMDYEDFKYCWENFV YNDNEPFPKWKGLKTNFRLLKRRLRESLQ
SEQ ID NO: 45 APOBEC3D	MNPQIRNPMERMYRDTFYDNFENEPILYGRSYTWLCYEVKIKRGRSNLLWDTGV FRGPVLPKRQSNHRQEVYFRFENHAEMCFLSWFCGNRLPANRRFQITWVFSWNP CLPCVVKVTKFLAEHPNVTLTISAARLYYYRDRDWRVLLRLHKAGARVKIMD YEGERCGRGQGSMTGRNSLRD GWICNAMAGGVP GPAGVGLALATDSQETRPG RAGPGSGESLSASHLFISDFAYCWENFVCNEGQPFMPWYKFDDNYASLHRTLKEI LRNPMEAMYPHIFYHFKNLLKACGRNESWLCFTMEVTKHHS AVFRKRGVFRN QVDPETHCHAERCFLSWFCDDILSPNTNYEVTWYTSWSPCECAGEVAEFLARH SNVNLTIPTARLCYFWDTDYQEGCLSLSQEGASVKIMGYKDFVSCWKNFVYSDD EPFKPWKGLQTNFRLLKRRLREILQ
SEQ ID NO: 46 APOBEC3F	MKPHFRNTVERMYRDTFSYNFYNRPIILSRRTVWLCYEVKTKGPSRPRLD AKIF RGQVYSQPEHHAEMCFLSWFCGNQLPAYKCFQITWVFSWTPCPDCVAKLAEFL AEHPNVTLTISAARLYYYWERDYRRALCRLSQAGARVKIMDDEEFAYCWENFV YSEGQPFMPWYKFDDNY AFLHRTLKEILRNPMMEAMYPHIFYHFKNLRKAYGRN ESWLCFTMEVVKHHSVPVSWKRGVFRNQVDPETHCHAERCFLSWFCDDILSPNTN YEVTWYTSWSPCECAGEVAEFLARHSNVNLTIFTARLYYFWDTDYQEGRLSLS QEGASVEIMGYKDFKYCWENFVYNDDEPFPKWKGLKYNFLFLDSKLQEILE
SEQ ID NO: 47 APOBEC3G	MELKYHPEMRFFHWFSKWRKLHRDQEYEV TWYISWSPCTKCTRDMATFLAEDP KVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPATMKIMNYDEFQHCWSKFV YSQRELFEPPWNNLPKYIYLLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYL CYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLGRHAELCFLDVIPFWKLDL DQDYRVTCFTSWSPCFSCAQEMAKFISKKNKHVSLCIFTARIYDDQGRCCQEGRLTL AEAGAKISIMTYSEFKHCWDTFVDHQGCPFPQWDGLDEHSQDL SGRLRAILQNO EN
SEQ ID NO: 48 APOBEC3H	MALLTAETFRLOFNKRRRLRRPYYPKALLCYQLTPQNGSTPTRGYFENKKKCH AEIFINEIKSMGLDETQCYQVTCYLTWSPCSSCAWELVDFIKAHDLNLRIFASR LYYHWCKPQQDGLRLLCGSQVPVEVMGFPSRGTCAGSLHGYIV
SEQ ID NO: 49 APOBEC4	MEPIYEEYLANHGTIVKPYWLSFSLDCSNCPYHIRTGEEARVSLTEFCQIFGFPY GTTFPQTKHLTFYELKTSSGSLVQKGHASSCTGNYIHPEMMLFEMNGYLD SAIYN NDSIRHIILYSNNSPCNEANHCCISKMYNFLTITYPGITLSIYFSQLYHTEMDFPASA WNREALRSLASLWPRVVLSPISGGIWHSVLHSFISGVSGSHVFQPILTGRALADRH NAYEINAITGVKPYFTDVLQTKRNPNTKAQEALESYPLNNAFPQQFQMPSGQL QPNLPPDLRAPVVFVLPRLDLPMMHMGQNPKNPRNIVRHLNMPQMSFQETKDL GRLPTGRSVEIVEITEQFASSKEADEKKKKKKGKK
SEQ ID NO: 50	MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNK CHVELLFLRYISDWDLDPGRCYRVTFWTSWSPCYDCARHVADFLRGPNLSLRI

Activation-induced cytidine deaminase (AID)	FTARLYFCEDRKAEPGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAW EGLHENSVRLSRQLRRILL
---	--

EXAMPLES

[0310] While several experimental Examples are contemplated, these Examples are intended non-limiting.

5

Materials

[0311] Table 3 below shows examples of bacterial plasmids.

Table 3.

Plasmid number	Name	Benchling link
pDF0038	DsiCas7-11a full locus with CRISPR array	https://benchling.com/s/seq-72T3bYt8Ff9QNiG7aE2X
pDF0039	DsiCas7-11a only expression with CRISPR array	https://benchling.com/s/seq-uiS8cQyZLi2NrqlfxWF2
pDF0044	CjcCas7-11b full locus with synthetic array	https://benchling.com/s/seq-GXIXQfFqwbN2vu7g6NUV
pDF0118	DsiCas7-11a bacterial expression with Twinstrep-SUMO-Codon	https://benchling.com/s/seq-PSpbXxlK66yLBw0TbxxnY
pDF0163	GwCas7-11c bacterial expression with Twinstrep-SUMO-Codon	https://benchling.com/s/seq-CsMaJRrLeJ15XtHvMFd1
pDF0191	CjcCas7-11b bacterial expression with Twinstrep-SUMO-Codon	https://benchling.com/s/seq-ILOOF2YI8T6v2jUuQ9Q3
pDF0229	DsiCas7-11a expression vector with golden gate site for guides	https://benchling.com/s/seq-mPWEEpoNVZVKxJwllhVj

[0312] Table 4 below shows examples of bacterial spacers.

10

Table 4.

SEQ ID NO	Spacer sequence	Target
SEQ ID NO: 51	CTTTGTTGTCTTCGACATGGGTAATCCTCAT	MS2 targeting 1
SEQ ID NO: 52	TAAAGAGTTGAACCTCTTTGTTGTCTTCGAC	MS2 targeting 3
SEQ ID NO: 53	GTACCTTCGTACGGACGACCTTCACCTTCAC	RFP targeting 1
SEQ ID NO: 54	CAGGATGTCCCAAGCGAACGGCAGCGGACC A	RFP targeting 2
SEQ ID NO: 55	CAGGGAGGAGTCCTGGGTAACGGTAACAAC A	RFP targeting 3
SEQ ID NO: 56	CATGAACTCTTTGATAACGTCTTCGCTACTC	RFP targeting 4
SEQ ID NO: 57	TCGGAGGTAGCCCGACTCGCTCCGGTTCGTA	non-targeting 1
SEQ ID NO: 58	TGAAAGTTGGCGCCTTTTTTCATACGTCGAG	non-targeting 2

[0313] Table 5 below shows examples of targets and pre-crRNA.

Table 5.

Target	Sequence
SEQ ID NO: 59 DsiCas7-11 array (4DR-3S)	GGGCGAAAUCGGUAUUUCCAAUGAAAUCACUCAGGUUGGAAAGCCGGUUUUC UUUGAUGUCACGGAACCGCCAAUAUAACAAUACAGGUUUUAUAGUUGGGUUG GAAAGCCGGUUUUUCUUUGAUGUCACGGAACAUUGUAAACUGCAUAUCUGUGU UUGCCGUCUGGUUGGAAAGCCGGUUUUUCUUUGAUGUCACGGAACUCACGUGA GUCAGUCUUAGUUCUAUACUUCUUUAGGGUUGGAAAGCCGGUUUUUCUUUGAU GUCACGGAACAGGAACUUGAACAAUACGUAACUAACGAG
SEQ ID NO: 60 ssRNA 1 target	GGGGGCCAGUGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAAAUAUGGAU UACUUGGUAGAACAGCAAUCUACUCGACCUGCAGGCAUGCAAGCUUGGCGUA AUCUUGGUCAUAGCUGUUUCCUGUGUUUAUCCGCUCACAAUCCACACAACA UACGAGCCGGAAGCAUAAAG
SEQ ID NO: 61 MS2 target for body labeling	GGGGCAAACUCCGGCAUCUACUAAUAGACGCCGGCCAUUCAAACAUGAGGAU UACCCAUGUCGAAGACAACAAAGAAGUUAACUCUUUAUGUAUUGAUCUUCU UCGCGAUCUUUCUCUCGAAAUUUACCAAUCAAUUGCUUCUGUCGCUACU
SEQ ID NO: 62 ssRNA 2 target	GGGGGCCAGUGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAAAUAUGGAU UACUUGGUAGAACAGCAAUCUACUCGACCUGCAGGCAUGCAAGCUUGGCGUA AUCUUGGUCAUAGCUGUUUCCUGUGUUUAUCCGCUCACAAUCCACACA
SEQ ID NO: 63 fixed crRNA target 1	GGCCAGUGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAAAUAUGGAUUAC UUGGGUCGAAGACAACAAAGAAGUUAACUCUUUAUAGAACAGCAAUCUACU CGACCUGCAGGCAUGCAAGCUUGGCGUAAUCAUGGUCAUAGCUGUU
SEQ ID NO: 64 fixed crRNA target 2	AUGCCCGCCAUGAAGAUCGAGUGCCGCAUACCCGGCACCCUGAACGGCGUGGA GUUGUCGAAGACAACAAAGAAGUUAACUCUUUACGAGCUGGUGGGCGGCGG AGAGGGCACCCCGAGCAGGGCCGCAUGACCAACAAGAUGAAGAG
SEQ ID NO: 65 fixed crRNA target 3	AUUCUAUGUCUGGUUUUCGCUCAAAAACUCCCCGGAUAGACAACAGCACGG CAACGUCGAAGACAACAAAGAAGUUAACUCUUUAGCUGUGCCUUGGGCACC AUGCAGUACCAAACGGAACGAUAGUGAAAAACAAUCACGAAUGACCA
SEQ ID NO: 66 fixed crRNA target 4	AUUCUAUGUCUGGUUUUCGCUCAAAAACUCCCCGGAUAGACAACAGCACGG CAACGUCGAAGACAACAAAGAAGUUAACUCUUUAGCUGUGCCUUGGGCACC AUGCAGUACCAAACGGAACGAUAGUGAAAAACAAUCACGAAUGACCA
SEQ ID NO: 67 fixed crRNA target 5	GCUGUGCCUUGGGCACCAUGCAGUACCAAACGGAACGAUAGUGAAAAACAAUC ACGAAUGACCAGUCGAAGACAACAAAGAAGUUAACUCUUUAAUUCUAUGUC UGGUUUUCGCUCAAAAACUCCCCGGAUAGACAACAGCACGGCAAC
SEQ ID NO: 68 GwCas7-11 array (3DR-2S)	GGGGUUUGGAAGCCGGGUCGAAGUCAGGCCCGUUAAGACCGAGUAGAUUGCU GUUCUACCAAGUAAUCCAGUUUGGAAGCCGGGUCGAAGUCAGGCCCGUUAAG ACU AUGCUUCCGGCUCGUAUGUUGUGUGAAUUGUUUGGAAGCCGGGUCGAA GUCAGGCCCGUUAAGAC
SEQ ID NO: 69 CjeCas7-11 array (3DR-2S)	GGGGCUAGCUCAGUCCUAGGUAAUAGUCUAGCGCUGAUUUAGGCAAAAACGG CUUGAAGACUAAAGGAAGGAAUUAUUGUCACGGUACGACUAAAUGGCACCAC UAAUUGUUGCCUUCUAAAGUCCUUCUUGAAGACUAAAGGAAGGAAUUAUGU CACGGUACGUUUGUUUAACCAUAGAGGCACACCGCCUCUGAGAGCUUGAAG ACUAAAGGAAGGAAUUAUUGUCACGGUACAAUCUCCGCCCGUUCGUAAGC
SEQ ID NO: 70 EGFP ssRNA 400 nt	AUGCCCGCCAUGAAGAUCGAGUGCCGCAUACCCGGCACCCUGAACGGCGUGGA GUUCGAGCUGGUGGGCGGCGGAGAGGGCACCCCGAGCAGGGCCGCAUGACC AACAAGAUGAAGAGCACCAAAGGCGCCUGACCUUCAGCCCUUACCUGCUGAG CCACGUGAUGGGCUACGGCUUCUACCACUUCGGCACCUACCCAGCGGCUACG AGAACCCCUUCCUGCACGCCAUCAACAACGGCGGCUACACCAACACCCGCAUC GAGAAGUACGAGGACGGCGGCGUGCUGCACGUGAGCUUACGCUACCGCUACG AGGCCGGCCGCGUGAUCGGCGACUUAAGGUGGUGGGCACCGGCUUCCCCGAG GACAGCGUGAUCUUCACCGACAAGAUAUCC
SEQ ID NO: 71	GGGGUUGGAAAGCCGGUUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUUCUUUGAUGUCACGGAAC

MS2 crRNA 1 (mini array) WT	
SEQ ID NO: 72	GGCGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 1	
SEQ ID NO: 73	GGGGAUGGAAAGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGAUGGAAAGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 3	
SEQ ID NO: 74	GGGGUUCGAAAGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUCGAAAGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 5	
SEQ ID NO: 75	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 7	
SEQ ID NO: 76	GGGGUUGGAAUGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAUGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 9	
SEQ ID NO: 77	GGGGUUGGAAAGGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGGCCGGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 11	
SEQ ID NO: 78	GGGGUUGGAAAGCCCGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCCGUUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 13	
SEQ ID NO: 79	GGGGUUGGAAAGCCGGAUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGAUUUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 15	
SEQ ID NO: 80	GGGGUUGGAAAGCCGGUUAUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUAUCUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 17	
SEQ ID NO: 81	GGGGUUGGAAAGCCGGUUUUGUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUGUUUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 19	
SEQ ID NO: 82	GGGGUUGGAAAGCCGGUUUUCUAUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUAUGAUGUCACGGAAC
MS2 crRNA 1 (mini array) mutant pos 21	
SEQ ID NO: 83	GGGGUUGGAAAGCCGGUUUUCUUUCAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUCAUGUCACGGAAC

MS2 crRNA 1 (mini array) mutant pos 23	
SEQ ID NO: 84 MS2 crRNA 1 (mini array) mutant pos 25	GGGGUUGGAAAGCCGGUUUUCUUUGAAGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAAGUCACGGAAC
SEQ ID NO: 85 MS2 crRNA 1 (mini array) mutant pos 27	GGGGUUGGAAAGCCGGUUUUCUUUGAUGACACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGACACGGAAC
SEQ ID NO: 86 MS2 crRNA 1 (mini array) mutant pos 29	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCUCGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCUCGGAAC
SEQ ID NO: 87 MS2 crRNA 1 (mini array) mutant pos 31	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACCGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACCGAAC
SEQ ID NO: 88 MS2 crRNA 1 (mini array) mutant pos 33	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGUACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGUAC
SEQ ID NO: 89 MS2 crRNA 1 (mini array) mutant pos 35	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGUAGCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGUAG
SEQ ID NO: 90 MS2 crRNA 1 (mini array) WT	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAAC
SEQ ID NO: 91 MS2 crRNA 1 (mini array) double mutant pos 11	GGGGUUGGAAAGGGGGUUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGGGGGUUUUCUUUGAUGUCACGGAAC
SEQ ID NO: 92 MS2 crRNA 1 (mini array) double mutant pos 14	GGGGUUGGAAAGCCGCAUUUCUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGCAUUUCUUUGAUGUCACGGAAC
SEQ ID NO: 93 MS2 crRNA 1 (mini array) double mutant pos 17	GGGGUUGGAAAGCCGGUUAACUUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUAACUUUGAUGUCACGGAAC
SEQ ID NO: 94 MS2 crRNA 1 (mini array)	GGGGUUGGAAAGCCGGUUUUGAUUGAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUGAUUGAUGUCACGGAAC

double mutant pos 19	
SEQ ID NO: 95 MS2 crRNA 1 (mini array) double mutant pos 22	GGGGUUGGAAAGCCGGUUUUCUUACAUGUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUACAUGUCACGGAAC
SEQ ID NO: 96 MS2 crRNA 1 (mini array) double mutant pos 25	GGGGUUGGAAAGCCGGUUUUCUUUGAACUCACGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAACUCACGGAAC
SEQ ID NO: 97 MS2 crRNA 1 (mini array) double mutant pos 28	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUGUCGGAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUGUCGGAAC
SEQ ID NO: 98 MS2 crRNA 1 (mini array) double mutant pos 31	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACCCAACCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACCCAAC
SEQ ID NO: 99 MS2 crRNA 1 (mini array) double mutant pos 34	GGGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAUGCUUUGUUGUCUUCGA CAUGGGUAAUCCUCAUGGUUGGAAAGCCGGUUUUCUUUGAUGUCACGGAUG
SEQ ID NO: 100 SARS-CoV2 ssRNA	GGGAAUUCAGCGAAAUUGCACCCCCGAUUACGUUUGGUGGACCCUCAGAUUCA ACUGGCAGUAACCAGAAUGGAGAACGCAGUGGGGCGCGAUCAAAACAACGUC GGCCCCAAGGUUUACCCAAUAAUACUGCGUCUUGGUUACCCGUCUCAC
SEQ ID NO: 101 Influenza A segment 4 HA gene	GGGAUUCUAUGUCUGGUUUUCGCUCAAAAACUUCCTGGAAAUGACAACAGCA CGGCAACGCUGUGCCUUGGGCACCAUGCAGUACCAAACGGAACGAUAGUGAA AACAAUCACGAUGACCAAUUGAAGUCACUAAUGCUACUGAACUGGUU
SEQ ID NO: 102 34-nt MS2 ssRNA target	GGGATGAGGATTACCCATGTCTGAAGACAACAAAG
SEQ ID NO: 103 EGFP ssRNA 100 nt	AUGCCCGCCAUGAAGAUCGAGUGCCGCAUACCCGGCACCCUGAACGGCGUGGA GUUCGAGCUGGUGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCA
SEQ ID NO: 104 EGFP ssRNA 200 nt	AUGCCCGCCAUGAAGAUCGAGUGCCGCAUACCCGGCACCCUGAACGGCGUGGA GUUCGAGCUGGUGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCAUGACC AACAAGAUAGAAGAGACCAAAGGCGCCUGACCUUCAGCCCCUACCUGCUGAG CCACGUGAUGGGCUACGGCUUCUACCACUUCGGCACCUCACCC
SEQ ID NO: 105 EGFP ssRNA 600 nt	AUGCCCGCCAUGAAGAUCGAGUGCCGCAUACCCGGCACCCUGAACGGCGUGGA GUUCGAGCUGGUGGGCGGCGGAGAGGGCACCCCCGAGCAGGGCCGCAUGACC AACAAGAUAGAAGAGACCAAAGGCGCCUGACCUUCAGCCCCUACCUGCUGAG CCACGUGAUGGGCUACGGCUUCUACCACUUCGGCACCUCACCCAGCGGCUACG AGAACCCCUUCCUGCACGCCAUACAACGGCGGCUACACCAACACCCGCAUC GAGAAGUACGAGGACGGCGGCGUGCUGCACGUGAGCUUCAGCUACCCGCUACG

	AGGCCGGCCGCGUGAUCGGCGACUUC AAGGUGGUGGGCACC GGCUUCCCCGAG GACAGCGUGAUCUUCACCGACAAGAUC AUCCGCAGCAACGCCACCGUGGAGCA CCUGCACC CAUGGGCGAU AACGUGCUGGUGGGCAGCUUCGCCCCGACCUUCA GCCUGCGCGACGGCGGCUACUACAGCUUCGUGGUGGACAGCCACAUGCACUUC AAGAGCGCCAUCCACCCAGCAUCCUGCAGAACGGGGGCCCCAUGUUCGCCUU CCGCCGCGUGGAGGAGCUG
SEQ ID NO: 106 60 nt MS2 ssRNA	CCGGCCAUUC AAACAUGAGGAUUACCCAUGUCGAAGACAACAAAGAAGUUCA ACUCUUUA
SEQ ID NO: 107 ssRNA 1 with right PFS	GGCCAGUGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAAAUAUGGAUUAC UUGgUAGAACAGCAAUCUACUCGNNNNNACCUGCAGGCAUGCAAGCUUGGC GUAAUCAUGGUCAUAGCUGUUUCCUGUGUUUAUCCGCUCACAAUCCACACA ACAUACGAG CCGGAAGCAU
SEQ ID NO: 108 ssRNA 1 with left PFS	GGCCAGUGAAUUCGAGCUCGGUACCCGGGGAUCCUCUAGAAAUANNNNNNUG GAUUACUUGgUAGAACAGCAAUCUACUCGACCUGCAGGCAUGCAAGCUUGGCG UAAUCAUGGUCAUAGCUGUUUCCUGUGUUUAUCCGCUCACAAUCCACACAA CAUACGAGCCGGAAGCAU
SEQ ID NO: 109 31-nt MS2 ssRNA target	GGGAUGAGGAUUACCCAUGUCGAAGACAACAAAG
SEQ ID NO: 110 MS2 ssDNA and dsDNA target	CCGGCATCTACTAATAGACGCCGCCATTCAAACATGAGGATTACCCATGTCGA AGACAACAAAGAAGTTCAACTCTTTATGTATTGATCTTCCTCGCGA

[0314] Table 6 below shows examples of crRNA for *in vitro* assays.

Table 6.

Name	SEQ ID NO	DR sequence	SEQ ID NO	Spacer sequence
ssRNA crRNA 5	111	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	112	CGAGTAGATTGCTGTTCTacC AAGTAATCCA
ssRNA crRNA 1	113	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	114	TATGCTTCCGGCTCGTATGT TGTGTGGAATT
ssRNA crRNA 2	115	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	116	AACACAGGAAACAGCTATG ACCATGATTACG
ssRNA crRNA 3	117	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	118	AATTGTGAGCGGATAAACA CAGGAAACAGCT
ssRNA crRNA 4	119	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	120	TTGCTGTTCTacCAAGTAATC CATATTTCTA
ssRNA crRNA 6	121	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	122	TTTCTAGAGGATCCCCGGGT ACCGAGCTCGA
non targeting crRNA	123	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	124	CGCCAATATAACAATACAG GTATTATAGTTG
MS2 guide 1	125	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	126	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 guide 2	127	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	128	CTTCTTTGTTGTCTTCGACA TGGGTAATCCT
MS2 guide 3	129	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	130	TAAAGAGTTGAACTTCTTTG TTGTCTTCGAC
MS2 guide 4	131	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	132	TCTTTGTTGTCTTCGACATG GGTAATCCTCA
MS2 additional spacer 1	133	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	134	TAATCCTCATGTTTGAATGG CCGGCGTCTAT

MS2 additional spacer 2	135	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	136	TTCGACATGGGTAATCCTCA TGTTTGAATGG
MS2 additional spacer 3	137	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	138	CGAGGAAGATCAATACATA AAGAGTTGAACT
MS2 additional spacer 4	139	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	140	TGGTAAATTTGAGAGAAA GATCGCGAGGAA
MS2 additional spacer 5	141	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	142	TTGAATGGCCGGCGTCTATT AGTAGATGCCG
fixed position ssRNA 1 150nt guide 1	143	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	144	AATCCATATTTCTAGAGGAT CCCCGGGTACC
fixed position ssRNA 1 150nt guide 2	145	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	146	TGTTCTACCAAGTAATCCAT ATTTCTAGAGG
fixed position ssRNA 1 150nt guide 3	147	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	148	TCGAGTAGATTGCTGTTCTA CCAAGTAATCC
fixed position ssRNA 1 150nt guide 4	149	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	150	GCATGCCTGCAGGTCGAGT AGATTGCTGTTT
fixed position ssRNA 1 150nt guide 5	151	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	152	TTACGCCAAGCTTGCATGCC TGCAGGTCGAG
fixed position ssRNA 1 150nt guide 6	153	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	154	GCTATGACCATGATTACGCC AAGCTTGCATG
fixed position MS2 guide 1	155	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	156	AATCCTCATGTTTGAATGGC CGGCGTCTATT
fixed position MS2 guide 2	157	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	158	CTTCGACATGGGTAATCCTC ATGTTTGAATG
fixed position MS2 guide 3	159	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	160	ACTTCTTTGTTGTCTTCGAC ATGGGTAATCC
fixed position MS2 guide 4	161	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	162	CATAAAGAGTTGAACTTCTT TGTTGTCTTCG
fixed position MS2 guide 5	163	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	164	AGGAAGATCAATACATAAA GAGTTGAACTTC

fixed position MS2 guide 6	165	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	166	GAGAAAGATCGCGAGGAAG ATCAATACATAA
EGFP guide 1	167	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	168	AGGGTGCCGGTGATGCGGC ACTCGATCTTCA
EGFP guide 2	169	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	170	CCAGCTCGAACTCCACGCC GTTCAAGGTGCC
EGFP guide 3	171	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	172	CCCTCTCCGCCGCCACCAG CTCGAACTCCA
EGFP guide 4	173	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	174	CGGCCCTGCTCGGGGTGC CCTCTCCGCCGC
EGFP guide 5	175	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	176	ACGTGGCTCAGCAGGTAGG GGCTGAAGGTCA
EGFP guide 6	177	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	178	AGGTGCCGAAGTGGTAGAA GCCGTAGCCCAT
EGFP guide 7	179	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	180	TGGCGTGCAGGAAGGGGT CTCGTAGCCGCT
EGFP guide 8	181	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	182	GTACTTCTCGATGCGGGTGT TGGTGTAGCCG
EGFP guide 9	183	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	184	GCGGCCGGCCTCGTAGCGG TAGCTGAAGCTC
EGFP guide 10	185	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	186	AAGCCGGTGCCCACCACCTT GAAGTCGCCGA
EGFP guide 11	187	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	188	AAGATCACGCTGTCTCGG GGAAGCCGGTGC
GwCas7-11c MS2 crRNA	189	GTTTGGAAAGCCGGTTCGAAG TCAGGCCCGTTAAGAC	190	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
CjcCas7-11b MS2 crRNA	191	cttgaagactaaaggaaggaattaatgtcacg tac	192	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
GwCas7-11c non-targeting crRNA	193	GTTTGGAAAGCCGGTTCGAAG TCAGGCCCGTTAAGAC	194	CGCCAATATAACAATACAG GTATTATAGTTG
CjcCas7-11b non-targeting crRNA	195	cttgaagactaaaggaaggaattaatgtcacg tac	196	CGCCAATATAACAATACAG GTATTATAGTTG
MS2 crRNA DR length	197	GTTGGAAAGCCGGTTTTCTTT GATGTCACGGAAC	198	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	199	GGAAAGCCGGTTTTCTTTGAT GTCACGGAAC	200	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	201	AAGCCGGTTTTCTTTGATGTC ACGGAAC	202	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	203	CCGGTTTTCTTTGATGTCACG GAAC	204	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	205	GTTTTCTTTGATGTCACGGAA C	206	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	207	TCTTTGATGTCACGGAAC	208	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	209	TTGATGTCACGGAAC	210	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	211	ATGTCACGGAAC	212	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	213	TCACGGAAC	214	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA DR length	215	CGGAAC	216	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA spacer length	217	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	218	CTTTGTTGTCTTCGACATGG GTAATCCTCA

MS2 crRNA spacer length	219	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	220	CTTTGTTGTCTTCGACATGG GTAATCCT
MS2 crRNA spacer length	221	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	222	CTTTGTTGTCTTCGACATGG GTAAT
MS2 crRNA spacer length	223	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	224	CTTTGTTGTCTTCGACATGG GT
MS2 crRNA spacer length	225	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	226	CTTTGTTGTCTTCGACATG
MS2 crRNA spacer length	227	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	228	CTTTGTTGTCTTCGAC
MS2 crRNA spacer length	229	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	230	CTTTGTTGTCTTC
MS2 crRNA spacer length	231	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	232	CTTTGTTGTC
MS2 crRNA 1 DR WT	233	TTGATGTCACGGAAC	234	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 1	235	ATGATGTCACGGAAC	236	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 2	237	TAGATGTCACGGAAC	238	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 3	239	TTCATGTCACGGAAC	240	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 4	241	TTGTTGTCACGGAAC	242	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 5	243	TTGAAGTCACGGAAC	244	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 6	245	TTGATCTCACGGAAC	246	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 7	247	TTGATGACACGGAAC	248	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 8	249	TTGATGTGACGGAAC	250	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 9	251	TTGATGTCTCGGAAC	252	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 10	253	TTGATGTCAGGGAAC	254	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 11	255	TTGATGTCACCGAAC	256	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 12	257	TTGATGTCACGCAAC	258	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 13	259	TTGATGTCACGGTAC	260	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR mutant 14	261	TTGATGTCACGGATC	262	CTTTGTTGTCTTCGACATGG GTAATCCTCAT

MS2 crRNA 1 DR mutant 15	263	TTGATGTCACGGAAG	264	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR WT	265	TTGATGTCACGGAAC	266	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 1	267	AAGATGTCACGGAAC	268	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 2	269	TACATGTCACGGAAC	270	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 3	271	TTCTTGTCACGGAAC	272	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 4	273	TTGTAGTCACGGAAC	274	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 5	275	TTGAACTCACGGAAC	276	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 6	277	TTGATCACACGGAAC	278	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 7	279	TTGATGAGACGGAAC	280	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 8	281	TTGATGTGTCGGAAC	282	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 9	283	TTGATGTCTGGAAC	284	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 10	285	TTGATGTCAGCGAAC	286	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 11	287	TTGATGTCACCCAAC	288	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 12	289	TTGATGTCACGCTAC	290	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 13	291	TTGATGTCACGGTTC	292	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
MS2 crRNA 1 DR double mutant 14	293	TTGATGTCACGGATG	294	CTTTGTTGTCTTCGACATGG GTAATCCTCAT
COVID-19 N gene crRNA 1	295	GGTTGGAAAGCCGTTTTCTT TGATGTCACGGAAC	296	ACCAAACGTAATGCGGGGT GCATTTTCGCTGA
COVID-19 N gene crRNA 2	297	GGTTGGAAAGCCGTTTTCTT TGATGTCACGGAAC	298	GTTGAATCTGAGGGTCCACC AAACGTAATGC
COVID-19 N gene crRNA 3	299	GGTTGGAAAGCCGTTTTCTT TGATGTCACGGAAC	300	CATTCTGGTTACTGCCAGTT GAATCTGAGGG
COVID-19 N gene crRNA 4	301	GGTTGGAAAGCCGTTTTCTT TGATGTCACGGAAC	302	GCCCCACTGCGTTCTCCATT CTGGTTACTGC

COVID-19 N gene crRNA 5	303	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	304	ACGTTGTTTTGATCGCGCCC CACTGCGTTCT
COVID-19 N gene crRNA 6	305	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	306	GGTAAACCTTGGGGCCGAC GTTGTTTTGATC
COVID-19 N gene crRNA 7	307	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	308	AGACGCAGTATTATTGGGT AAACCTTGGGGC
COVID-19 N gene crRNA 8	309	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	310	TGAGAGCGGTGAACCAAGA CGCAGTATTATT
Influenza A segment 4 HA gene crRNA 1	311	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	312	CCGGGAAGTTTTTGAGCGA AAACCAGACATA
Influenza A segment 4 HA gene crRNA 2	313	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	314	CGTGCTGTTGTCATTTCCGG GAAGTTTTTGA
Influenza A segment 4 HA gene crRNA 3	315	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	316	CAAGGCACAGCGTTGCCGT GCTGTTGTCATT
Influenza A segment 4 HA gene crRNA 4	317	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	318	GGTACTGCATGGTGCCCAA GGCACAGCGTTG
Influenza A segment 4 HA gene crRNA 5	319	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	320	TCACTATCGTTCCGTTTGGT ACTGCATGGTG
Influenza A segment 4 HA gene crRNA 6	321	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	322	GTCATTTCGTGATTGTTTTCA CTATCGTTCCG
Influenza A segment 4 HA gene crRNA 7	323	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	324	AGTGACTTCAATTTGGTCAT TCGTGATTGTT
Influenza A segment 4 HA gene crRNA 8	325	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	326	ACCAGTTCAGTAGCATTAGT GACTTCAATTT
EGFP guide 12	327	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	328	TGGGGTGACGGTGCTCCAC GGTGGCGTTGCT
EGFP guide 13	329	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	330	AAGGTGCGGGCGAAGCTGC CCACCAGCACGT
EGFP guide 14	331	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	332	GTCCACCACGAAGCTGTAG TAGCCGCCGTCG
EGFP guide 15	333	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	334	TGCTGGGGTGGATGGCGCT CTTGAAGTGCAT
EGFP guide 16	335	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	336	GGAAGGCGAACATGGGGCC CCCCTTCTGCAG
MS2 tiling guide 1	337	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	338	AATCCTCATGTTTGAATGGC CGGCGTCTATT
MS2 tiling guide 2	339	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	340	GGTAATCCTCATGTTTGAAT GGCCGGCGTCT

MS2 tiling guide 3	341	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	342	ATGGGTAATCCTCATGTTTG AATGGCCGGCG
MS2 tiling guide 4	343	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	344	GACATGGGTAATCCTCATGT TTGAATGGCCG
MS2 tiling guide 5	345	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	346	TTCGACATGGGTAATCCTCA TGTTTGAATGG
MS2 tiling guide 6	347	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	348	GTCTTCGACATGGGTAATCC TCATGTTTGAA
MS2 tiling guide 7	349	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	350	GTTGTCTTCGACATGGGTAA TCCTCATGTTT
MS2 tiling guide 8	351	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	352	TTTGTTGTCTTCGACATGGG TAATCCTCATG
MS2 tiling guide 9	353	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	354	TTCTTTGTTGTCTTCGACAT GGGTAATCCTC
MS2 tiling guide 10	355	GGTTGGAAAGCCGGTTTTCTT TGATGTCACGGAAC	356	AACTTCTTTGTTGTCTTCGA CATGGGTAATC
Cjc MS2 60bp ssRNA crRNA 1	357	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	358	AATCCTCATGTTTGAATGGC CGGCGTCTATT
Cjc MS2 60bp ssRNA crRNA 2	359	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	360	GGTAATCCTCATGTTTGAAT GGCCGGCGTCT
Cjc MS2 60bp ssRNA crRNA 3	361	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	362	ATGGGTAATCCTCATGTTTG AATGGCCGGCG
Cjc MS2 60bp ssRNA crRNA 4	363	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	364	GACATGGGTAATCCTCATGT TTGAATGGCCG
Cjc MS2 60bp ssRNA crRNA 5	365	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	366	TTCGACATGGGTAATCCTCA TGTTTGAATGG
Cjc MS2 60bp ssRNA crRNA 6	367	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	368	GTCTTCGACATGGGTAATCC TCATGTTTGAA
Cjc MS2 60bp ssRNA crRNA 7	369	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	370	GTTGTCTTCGACATGGGTAA TCCTCATGTTT
Cjc MS2 60bp ssRNA crRNA 8	371	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	372	TTTGTTGTCTTCGACATGGG TAATCCTCATG
Cjc MS2 60bp ssRNA crRNA 9	373	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	374	TTCTTTGTTGTCTTCGACAT GGGTAATCCTC
Cjc MS2 60bp ssRNA crRNA 10	375	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	376	AACTTCTTTGTTGTCTTCGA CATGGGTAATC
Cjc MS2 60bp ssRNA crRNA 11	377	CTTGAAGACTAAAGGAAGGA ATTAATGTCACGGTAC	378	TTGAACTTCTTTGTTGTCTTC GACATGGGTA
Gw MS2 60bp ssRNA crRNA 1	379	GTTTGGAAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	380	AAACGTAATGCGGGGTGCA TTTCGCTGATTT
Gw MS2 60bp ssRNA crRNA 2	381	GTTTGGAAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	382	ACCAAACGTAATGCGGGGT GCATTTTCGCTGA
Gw MS2 60bp ssRNA crRNA 3	383	GTTTGGAAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	384	TCCACCAAACGTAATGCGG GGTGCATTTCCG

Gw MS2 60bp ssRNA crRNA 4	385	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	386	GGGTCCACCAAACGTAATG CGGGGTGCATTT
Gw MS2 60bp ssRNA crRNA 5	387	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	388	TGAGGGTCCACCAAACGTA ATGCGGGGTGCA
Gw MS2 60bp ssRNA crRNA 6	389	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	390	ATCTGAGGGTCCACCAAAC GTAATGCGGGGT
Gw MS2 60bp ssRNA crRNA 7	391	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	392	TGAATCTGAGGGTCCACCA AACGTAATGCGG
Gw MS2 60bp ssRNA crRNA 8	393	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	394	AGTTGAATCTGAGGGTCCA CCAAACGTAATG
Gw MS2 60bp ssRNA crRNA 9	395	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	396	GCCAGTTGAATCTGAGGGT CCACCAAACGTA
Gw MS2 60bp ssRNA crRNA 10	397	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	398	ACTGCCAGTTGAATCTGAG GGTCCACCAAAC
Gw MS2 60bp ssRNA crRNA 11	399	GTTTGGGAAGCCGGGTCGAAG TCAGGCCCGTTAAGAC	400	GTTACTGCCAGTTGAATCTG AGGGTCCACCA

[0315] Table 7 below shows examples of mammalian plasmids.

Table 7.

Plasmid number	Name	Benchling link
pDF0158	pCMV - huDsiGRAMP N term msfGFP mammalian expression	https://benchling.com/s/seq-16V8YUC36yaVMw4GYawz
pDF0159	pCMV - huDsiGRAMP mammalian expression	https://benchling.com/s/seq-V4NVwioQSHLp6fkKY5I
pDF0160	pCMV - huDsiGRAMP C term msfGFP mammalian expression	https://benchling.com/s/seq-6KYYZvPGjqOWNIPwLN2u
pDF0113	pU6-Eco31i-Eco31i-DsiCas7-11a full DR guide scaffold with golden gate site	https://benchling.com/s/seq-fl07Vm8NmTi2ydJnbZGS
pDF0114	pU6-Eco31i-Eco31i-DsiCas7-11a mature DR guide scaffold with golden gate site	https://benchling.com/s/seq-22EWyWJAfvfpX2EsdJxs
pDF0111	CMV-Cluciferase(STOP85)-poly A EF1a-G-luciferase-poly A (pAB0040)	https://benchling.com/s/seq-9fEQxFBZrIEQo8Ta407A
pDF0112	CMV-Cluciferase-poly A EF1a-G-luciferase-poly A	https://benchling.com/s/seq-QDT25JIo4luxluesN196
pDF0234	pCMV - huDsiGRAMP (5' UTR from BE constructs) dead mutant	https://benchling.com/s/seq-cNh2lgwflmDV66VanySf
pDF0237	pCMV - dead huDsiGRAMP - NES - huADAR2	https://benchling.com/s/seq-OKxBeGoAOWuk6GbgfAa2
pDF0228	pCMV - huDsiGRAMP - NES - huADAR2	https://benchling.com/s/seq-BnKsJ4pko0RcknikJMIr

5 [0316] Table 8 below shows examples of Cas7-11 guides for mammalian.

Table 8.

SEQ ID NO	Name	Spacer sequence	Target	Application (Knockdown/editing)
401	DsiGRAMP mammalian Gluc guide 1 F	GCAGATCAGGGCAAACAG AACTTTGACTCCC	Gluc	Knockdown
402	DsiGRAMP mammalian Gluc guide 2 F	CGATGTTGAAGTCTTCGTT GTTCTCGGTGGG	Gluc	Knockdown
403	DsiGRAMP mammalian Gluc guide 3 F	GCAACTTCCCGCGGTCAGC ATCGAGATCCGT	Gluc	Knockdown
404	DsiGRAMP mammalian Gluc guide 4 F	TGCAGCCAGCTTTCCGGGC ATTGGCTTCCAT	Gluc	Knockdown
405	DsiGRAMP mammalian Gluc guide 5 F	TGAACTTCTTCATCTTGGGC GTGCACTTGAT	Gluc	Knockdown
406	DsiGRAMP mammalian Gluc guide 6 F	GAATGTCGACGATCGCCTC GCCTATGCCGCC	Gluc	Knockdown
407	DsiGRAMP mammalian Gluc guide 7 F	ATGAACTGCTCCATGGGCT CCAAGTCCTGA	Gluc	Knockdown
408	DsiGRAMP mammalian Gluc guide 8 F	GCAGCCAGTTGTGCAGTCC ACACACAGATCG	Gluc	Knockdown
409	DsiGRAMP mammalian Gluc guide 9 F	GCGTTGCGGCAGCCACTTC TTGAGCAGGTCA	Gluc	Knockdown
410	DsiGRAMP mammalian Gluc guide 10 F	TGTCCACCTGGCCCTGGAT CTTGCTGGCAA	Gluc	Knockdown
411	DsiGRAMP mammalian Non-targeting 1 F	GGTAATGCCTGGCTTGTCTG ACGCATAGTCTG	Gluc	Knockdown
412	DsiGRAMP mammalian Non-targeting 2 F	GGGAACCTTGCCGTTATA AAGTCTGACCAG	Gluc	Knockdown
413	KRAS guide 1	TATAATGGTGAATATCTTC AAATGATTTAGT	KRAS	Knockdown
414	KRAS guide 2	ATGTATAGAAGGCATCATC AACACCCTGTCT	KRAS	Knockdown
415	KRAS guide 3	GGTTAAAAATTTACAGATT GTGCTGAGCTTG	KRAS	Knockdown
416	PPIB guide 1	GTAGATGCTCTTTCCTCCTG TGCCATCTCCC	PPIB	Knockdown
417	PPIB guide 2	CAGTTTGAAGTTCTCATCG GGGAAGCGCTCA	PPIB	Knockdown
418	PPIB guide 3	CAGTGTGGTAGGAGTTTG TTACAAAAGTGA	PPIB	Knockdown

419	MALAT1 guide 1	CTTGGCCAAGTCTGTTATGT TCACCTGAAAA	MALA T1	Knockdown
420	MALAT1 guide 2	CAAAATGTACTCAGCTTCA ATCACAAATACG	MALA T1	Knockdown
421	MALAT1 guide 3	GGTTATAGCTTGACAAGCA ATTAAC TTAAA	MALA T1	Knockdown
422	CXCR4 guide 1	ATGATAATGCAATAGCAGG ACAGGATGACAA	CXCR4	Knockdown
423	Non-targeting 1	GGTAATGCCTGGCTTGTCG ACGCATAGTCTG	-	Knockdown
424	Non-targeting 2	GGGAACCTTGGCCGTTATA AAGTCTGACCAG	-	Knockdown
425	Cluc W85X guide MM1	CATCCTGCGGCCTCTACTCT GCATTCAATTACATACTGA CACATTCGGCA	Cluc	Editing
426	Cluc W85X guide MM3	ACCATCCTGCGGCCTCTAC TCTGCATTCAATTACATACT GACACATTCGG	Cluc	Editing
427	Cluc W85X guide MM5	AAACCATCCTGCGGCCTCT ACTCTGCATTCAATTACAT ACTGACACATTC	Cluc	Editing
428	Cluc W85X guide MM7	CTAAACCATCCTGCGGCCT CTACTCTGCATTCAATTACA TACTGACACAT	Cluc	Editing
429	Cluc W85X guide MM9	TTCTAAACCATCCTGCGGC CTCTACTCTGCATTCAATTA CATACTGACAC	Cluc	Editing
430	Cluc W85X guide MM11	TGTTCTAAACCATCCTGCG GCCTCTACTCTGCATTCAAT TACATACTGAC	Cluc	Editing
431	Cluc W85X guide MM13	AATGTTCTAAACCATCCTG CGGCCTCTACTCTGCATTCA ATTACATACTG	Cluc	Editing
432	Cluc W85X guide MM15	AGAATGTTCTAAACCATCC TGCGGCCTCTACTCTGCATT CAATTACATAC	Cluc	Editing
433	Cluc W85X guide MM17	ATAGAATGTTCTAAACCAT CCTGCGGCCTCTACTCTGC ATTCAATTACAT	Cluc	Editing

434	Cluc W85X guide MM19	CCATAGAATGTTCTAAACC ATCCTGCGGCCTCTACTCTG CATTCAATTAC	Cluc	Editing
435	Cluc W85X guide MM21	TTCCATAGAATGTTCTAAA CCATCCTGCGGCCTCTACTC TGCATTCAATT	Cluc	Editing
436	Cluc W85X guide MM23	CTTTCCATAGAATGTTCTAA ACCATCCTGCGGCCTCTAC TCTGCATTCAA	Cluc	Editing
437	Cluc W85X guide MM25	CTCTTTCCATAGAATGTTCT AAACCATCCTGCGGCCTCT ACTCTGCATTC	Cluc	Editing
438	Cluc W85X guide MM27	ATCTCTTTCCATAGAATGTT CTAAACCATCCTGCGGCCT CTACTCTGCAT	Cluc	Editing
439	Cluc W85X guide MM29	GAATCTCTTTCCATAGAAT GTTCTAAACCATCCTGCGG CCTCTACTCTGC	Cluc	Editing
440	Cluc W85X guide MM31	TGGAATCTCTTTCCATAGA ATGTTCTAAACCATCCTGC GGCCTCTACTCT	Cluc	Editing
441	Cluc W85X guide MM33	ACTGGAATCTCTTTCCATA GAATGTTCTAAACCATCCT GCGGCCTCTACT	Cluc	Editing
442	Cluc W85X guide MM35	GAAGTGAATCTCTTTCCA TAGAATGTTCTAAACCATC CTGCGGCCTCTA	Cluc	Editing
443	Cluc W85X guide MM37	TGGAAGTGAATCTCTTTC CATAGAATGTTCTAAACCA TCCTGCGGCCTC	Cluc	Editing
444	Cluc W85X guide MM39	CCTGGAAGTGAATCTCTT TCCATAGAATGTTCTAAAC CATCCTGCGGCC	Cluc	Editing
445	Cluc W85X guide MM41	TTCCTGGAAGTGAATCTC TTTCCATAGAATGTTCTAA ACCATCCTGCGG	Cluc	Editing
446	Cluc W85X guide MM43	GGTTCCTGGAAGTGAATC TCTTTCCATAGAATGTTCTA AACCATCCTGC	Cluc	Editing

447	LwaCas13a guide for comparison	ACCCAGGAATCTCAGGAAT GTCGACGAT	Gluc	Knockdown
448	PspCas13b guide for comparison	GGGCATTGGCTTCCATCTCT TTGAGCACCT	Gluc	Knockdown
449	RfxCas13d guide for comparison	ACCCAGGAATCTCAGGAAT GTCGACGAT	Gluc	Knockdown
450	shRNA guide for comparison	AAAGTTCTGTTTGCCCTGAT CCTCGAGGATCAGGGCAAA CAGAACTTT	Gluc	Knockdown

[0317] Table 9 shows the representative Cas7-11 orthologs of Figure 1F.

Table 9.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 451 MBD3181016.1	RDMATSAHTTIDKETKRTKENHLINLEVVRIEDN-RYLKGSVEMIN-
SEQ ID NO: 452 MBC8521845.1	DQVRTSAHNKINRATGRVEKDLLFSYELG--ANT-VPLEGTIERVNP
SEQ ID NO: 453 RCV63412.1	DMVRTSTHNKISRATGRVEKDHLFSYELG--ANT-TPFEGMIEPVNH
SEQ ID NO: 454 MBC8461765.1	GTTRVSLHNRVDPTTGRVPEETFFSFEVG--SAE-K-FTGRIERSGP
SEQ ID NO: 455 HHR29383.1	EPIRISTHNRIDPATGRVPKDLLFTFEMG--SPA-V-FRGYIERIGI
SEQ ID NO: 456 WP_129889723.1	LQTFLYTQARLDRPRTAVPGALYSSEFG--LRE-ITFKGSVTGWLE
SEQ ID NO: 457 WP_174676620.1	NQKDVTYQVRLDRPRTAVEGALYTSEFG--TRG-LTFHGDITGSLT
SEQ ID NO: 458 TFH44231.1	LFLNTRTGTKIDRRTGTVQEDHLFTTESG--IKG-ITLTSKIYGRGN
SEQ ID NO: 459 MAT45258.1	YGQSIRSNNRIDPKTRTAANDHLFFTECV--DPD-TYFIGEIRSSSN
SEQ ID NO: 460 WP_075063448.1	YGRLLRTGIGIDRKLRVVVPDRLFTNECV--GTK-IPFTGRIIGEFP
SEQ ID NO: 461 HreCas7-11	EIAVSRVVNRVDPNSGKA-KDFFRVWEID--HKLCPNFLGKMSISLS
SEQ ID NO: 462 CmaCas7-11	DIVQKRILNRVDQTCGKA-KDFFKVCEVD--HIACPTLNIGIRINDE
SEQ ID NO: 463	ETAIERIVNRVDPSSGKA-KDYMRIWEID--PLVCSQFNIGIITIN--

Table 9.

SEQ ID NO/ NAME	SEQUENCE
HvmCas7-11	
SEQ ID NO: 464	EIGSRRTINRVDDTGKA-HDFFSIWEVD--A--VREFQGEIVLAAD
SmCas7-11	
SEQ ID NO: 465	EIGRLRTLNRIDRLTTKA-QDFFRIYEVD--Q--VRDFFGTITLAGD
OmCas7-11	
SEQ ID NO: 466	AIGSQRVLNRVDFKSGKA-HDFFKAYEVD--HTRFPRFEGEITIDNK
DisCas7-11	
SEQ ID NO: 467	EIAVRRVVNRVDYASGKA-HDFFRIFEVD--HIAFPCFHGEIAFGEN
DsbaCas7-11	
SEQ ID NO: 468	DIGTQRTLNRVDYATGKA-HDFFKVWEID--HSLLSVFQGKISIAN
SstCas7-11	
SEQ ID NO: 469	DIASGRILNRVDFDTGKA-KDYFRTWEAD--YETYGTYTGRITLRNE
CsbCas7-11	
SEQ ID NO: 470	DIASERILNRVDYTTGKA-HDYFKVWEVD--DDQWWQFTGTITMHDD
CjcCas7-11	
SEQ ID NO: 471	DVCKERVNLNRVDFRTGKA-QDYFRVFEID--HEDWGVYTGEITITEP
CbfCas7-11	
SEQ ID NO: 472	ELAPKRALNRIEQYTGKA-QDYFTVYEAL--NKEFWTFKGRIRIKED
HvsCas7-11	
SEQ ID NO: 473	EIGIRILNRVDFTTGKA-QDYFYVWEVE--HSMCPKFQGTVKINED
HsmCas7-11	
SEQ ID NO: 474	DIAQKRTSNRVDFFSKKA-HDHYGVWEVT--A--VKNLLGYIYISDA
FmCas7-11	
SEQ ID NO: 475	QKIRPRILNRMDPGTSKA-RDYFRVFEIE--NQLCSQFRGWIWLSGD
DpbaCas7-11	

[0318] Table 10 shows the representative Cas7-11 orthologs of Figure 1G.

Table 10.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 476	QRI-----TPDAPDADP-YWQADIALDADGRPYLPG
WmCas7x3	
SEQ ID NO: 477	EEMDKQSG-----ESVDKKQNN-SWIQAIALDLNKKPYIPG
NisCas7x3	
SEQ ID NO: 478	E-----KERDENKET-RWLEAVALDHKGQPYIPG

Table 10.

SEQ ID NO/ NAME	SEQUENCE
GwCas7x3	
SEQ ID NO: 479	VTRP-----ELTVADRDELVDINAVVTDYTGKPYLPG
HymCas7x3	
SEQ ID NO: 480	TDRPGLEI-----EQKDGSKVK-AEINAFIKDSNGKPYLPG
DsbCas7x3	
SEQ ID NO: 481	IPVT---I-----KDTQGKNRE-VEVNSVITGKAALPIIPG
GamCas7x3	
SEQ ID NO: 482	VEHD-----LIKNDGTP-VQINALITGAGGLPIIPG
DesCas7x3	
SEQ ID NO: 483	AEIE---I-----NDSQGDRRQ-VQANAIIRGKDDKPIIPG
GabCas7x3	
SEQ ID NO: 484	CEHD-----DVKNNDGEP-VKINACIKGSKGRPIIPG
MotCas7x3	
SEQ ID NO: 485	CYRP-----ELTNADQKP-VDINACIKGANNLPIIPG
MobCas7x3	
SEQ ID NO: 486	-----EGQKQTDDQAE-SLHLRTLRLDGHGRFRIPF
GwCas7-11	
SEQ ID NO: 487	-----SEGQTS----QAILCPDGSYRLPR
HreCas7-11	
SEQ ID NO: 488	-----KTSHTR----STIFLNMNGQFCIPR
CmaCas7-11	
SEQ ID NO: 489	-----QEGQTS----ATILLNRDGYFRLPR
HvmCas7-11	
SEQ ID NO: 490	-----PIDDDQTS----SRTLVD RDGRYRLPR
SmCas7-11	
SEQ ID NO: 491	-----GSHDHTS----RKILLTRDFYRLPR
OmCas7-11	
SEQ ID NO: 492	-----EDAKQTD----LQVLLTPDNKYRLPR
DisCas7-11	
SEQ ID NO: 493	-----KEKKQTD----LMLLLDGQNHYRIPR
DsbaCas7-11	

Table 10.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 494 SstCas7-11	-----EDAKHTN----LKVLLDRQNRRLPR
SEQ ID NO: 495 SybCas7-11	-----GAGGHTD----LSILLGKDGHYRVPR
SEQ ID NO: 496 CsbCas7-11	SDSIPGKEK-----KSEDSL VINEHTS----FNILLDKENRYRIPR
SEQ ID NO: 497 CjcCas7-11	FDSTQDDL DLVPDIVNTDEKLEANEQTS----FRILMDKKGRYRIPR
SEQ ID NO: 498 CbfCas7-11	-----EQTEEVEHTS----LRLVMDKKGRFRIPR
SEQ ID NO: 499 HvsCas7-11	-----DEEKAE----GAILLTPDNRFRIPR
SEQ ID NO: 500 HsmCas7-11	-----KTEEQID----MQILLTKDGRYRLPR
SEQ ID NO: 501 FmCas7-11	-----QAGKQTS----MPTLQDSNDHFRLPR
SEQ ID NO: 502 DpbaCas7-11	-----ATETHTD----LPILLTSDRHFRIIPR

[0319] **Table 11** shows the representative Cas7-11 orthologs of Figure 1H.

Table 11.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 503 WmCas7x3	HRKPQDNAPDGTPRQRGDRA-----LLPGASLRGRLRS
SEQ ID NO: 504 NisCas7x3	KKDEAKNEADAKPRTNHQGQV-----ILPASSLRGRLRA
SEQ ID NO: 505 GwCas7x3	KADDADAV----PRRTHDDKI-----VLPASSLRGRLRT
SEQ ID NO: 506 HemCas7x3	KKKRRSNTPNLRPLRDAIGRP-----CLPESSVRGALRA
SEQ ID NO: 507 DsbCas7x3	KKRKED--IDHQPLRDSAGNA-----RLPAKSIRGAMRS

Table 11.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 508 <i>GanCas7x3</i>	DKCKAEDTPDIYPLEEKNGVP-----AFPVRSFRGAIRS
SEQ ID NO: 509 <i>DesCas7x3</i>	NTPDITQAPDMVPLVDEEDGNP-----MLPASSFRGALRA
SEQ ID NO: 510 <i>GabCas7x3</i>	SKGDGDQPAVHQPLTDRSNNP-----ILPARSFRGAIRA
SEQ ID NO: 511 <i>MotCas7x3</i>	LSSNENAKTDHYPLLDKNRNP-----YLPVSSFRGVLS
SEQ ID NO: 512 <i>MobCas7x3</i>	LEADPKTKIDHYPLLDNHKKP-----RLPSASIRGVLS
SEQ ID NO: 513 <i>GwCas7-11</i>	ESDAPDNVAYKKPVVQYDETGRLRTTDPGPVEMLTCLKGEGVRGVVAY
SEQ ID NO: 514 <i>HreCas7-11</i>	DHDNKDAVMVQKTVLFVDESG-----NYSQMPHHFLKSGSIRGACRF
SEQ ID NO: 515 <i>CmaCas7-11</i>	DNRNPDAVMVKKTLVYEQDSSTHKNVPKEVPKY-FIKSETIRGLLS
SEQ ID NO: 516 <i>HvmCas7-11</i>	DPRNTDAIMVRKTVFCDPNA---KNRPAPATVY-MIKGESIRGILRS
SEQ ID NO: 517 <i>SmCas7-11</i>	NSDTADIISFRRTVVDNGEVL-----REP-----VLRGEGLRGLLRT
SEQ ID NO: 518 <i>OmCas7-11</i>	-RHEEDSVYFQKRIFTS DGRV-----VLVP-----ALRGEGLRGLLRT
SEQ ID NO: 519 <i>DisCas7-11</i>	DKRGTDVVTFFVKYKAEGEEAK-----PVC-----AYKAESFRGVIRS
SEQ ID NO: 520 <i>DsbaCas7-11</i>	TEDVADIVSFKKYTQGGEKIY-----AYKSESFRGVVRT
SEQ ID NO: 521 <i>SstCas7-11</i>	EKDGSDIVSFRKYADDSGKEVY-----AYKAESFRGVVRA
SEQ ID NO: 522 <i>SybCas7-11</i>	ERIGFDNIAYEKRRYNGETNT-----TESIP---AVKGETFRGIVRT
SEQ ID NO: 523 <i>CsbCas7-11</i>	EPGNRDAIAYKKRVYNDGNNA-----IEPEPRF-AVKSETHRGIFRT
SEQ ID NO: 524 <i>CjcCas7-11</i>	NKDNIDCIAYEKRKWENGKIK-----FVP-----TIKGETIRGIVRM
SEQ ID NO: 525 <i>CbfCas7-11</i>	DPNNVDAIVFEKMKLDGDQVK-----YLP-----AIKGETIRGIVRT
SEQ ID NO: 526	EGKAPDAVFFKKYVFENGKIE-----EKP-----CFKAESIRGIFRT

Table 11.

SEQ ID NO/ NAME	SEQUENCE
HvsCas7-11	
SEQ ID NO: 527	DSSNTDLVTFKKYKLEESKEVF-----AIKGESIRGVFRT
HsmCas7-11	
SEQ ID NO: 528	-HGNTDSVFYKKPILKSKEKE-----PSYQW---AIKSDTVRGLIRS
FmCas7-11	
SEQ ID NO: 529	-RNESDAVFYQKSVAGEKGPVY-----ALKGEGLRGIVSS
DpbaCas7-11	

[0320] Table 12 shows the representative Cas7-11 orthologs of Figure 1I.

Table 12.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 530	PATLVDHDM LAIDRFTGGGKDGAKFK----LRYAECPTLEGQLSLDL
WmCas7x3	
SEQ ID NO: 531	LATLTRHEMVAIDRFTGGGKEGAKFN----VDYIECP-TLTGAIYLDL
NisCas7x3	
SEQ ID NO: 442532	SIKTRRHEMLAIDRFTGGGKDGAKFN----VDYVECP-TLAGKLSLDL
GwCas7x3	
SEQ ID NO: 533	EDCDNIQEFVAIDRFTGGAKDKAKFN----AEYIGSP-RFTGTIALDK
HvmCas7x3	
SEQ ID NO: 534	TRQETVQDFVAIDRFHGGGKETAKFD----ASFWRP-QYSILMHIPS
DsbCas7x3	
SEQ ID NO: 535	LKPFIIQEFVAIDRFHGGGKDEAKFN----AAHYQAP-VFKGKVRVSQ
GamCas7x3	
SEQ ID NO: 536	FRREQEQTFVAIDRFHGGCKEGALYT----IRHAESP-RFEGHLVIDP
DesCas7x3	
SEQ ID NO: 537	YRPAKTQQFVAIDRFHGGGKDGALFS----IKYFERP-VLKGGISLKL
GabCas7x3	
SEQ ID NO: 538	KSKKTKQDFVAIDRFHGGGKDGAKFD----ATHFERP-EFEGAISFSP
MetCas7x3	
SEQ ID NO: 539	ANELKTQEFIAIDRFHGGGKDGAKFN----AKHSERP-YFQGRITLSP
MebCas7x3	
SEQ ID NO: 540	TPHAMRSDRV ALDVF-GGAMPEAKFDDRPLAASPGKPLNFKSTIWYRE
GwCas7-11	
SEQ ID NO: 541	EVEAIKCDHVAIDRFHGGTVHRMKYDDYPLPGSPNRPLRIKGNIWVKR
HreCas7-11	
SEQ ID NO: 542	NVSDCCIDHVAIDRFTGGGVEKMKFNDYPLSASPKNCLNLKGSIWITS

Table 12.

SEQ ID NO/ NAME	SEQUENCE
CmaCas7-11	
SEQ ID NO: 543	SVSDKKMDHVAIDRFTGGGVQDMKFDDYPLPGCPAQPLILEGKFWVKD
HvmCas7-11	
SEQ ID NO: 544	TVADKRLDHVAIDRFDQSVVE--KYDDRPLVGSPKQPLVFKGCFWVQT
SmCas7-11	
SEQ ID NO: 545	TWNDKKIDHVSCSRFDASVVN--KFDDRSLVGSPDSPLHFEGTFWLHR
OmCas7-11	
SEQ ID NO: 4546	DPEPVTFDHVAIDRFTGGAADKKKFDDSPGSPARPLMLKGSFWIRR
DisCas7-11	
SEQ ID NO: 547	EPEPRRFDHVAIDRFTGGAVNQKKFDDRSLVPGKEGFMTLIGCFWMRK
DsbaCas7-11	
SEQ ID NO: 548	QPEPMIFDHVAIDRFTGGAVDKKKFDDCSLPGTPGHPLTLKGCFWIRK
SstCas7-11	
SEQ ID NO: 549	AWTRKHLDHVAIDRFHGGAEENMKFDYALAASPTNPLRMKGLIWVRS
SybCas7-11	
SEQ ID NO: 550	EKLEKHIDHVAIDRFTGGALDKAKFDYPLAGSPKKPLKLKGRFWIKK
CsbCas7-11	
SEQ ID NO: 551	SYKKKLIDHVAIDRFHGGAEEDKMKFNTLPLAGSFEKPIILKGRFWIKK
CjcCas7-11	
SEQ ID NO: 552	KDIAKKIDHVAIDRFTGGARDQMKFDTLPLIGSPERPLRLKGLFWMR
CbfCas7-11	
SEQ ID NO: 553	GQEEKFFDHVAIDRFLGGAKEKYKFDDKPIIGAPDTPIVLEGKIWVKK
HvsCas7-11	
SEQ ID NO: 554	DSAPKRLDHVAIDRFTGGAKEQAKFDDSPGSPDSPLEFTGIVWVRD
HsmCas7-11	
SEQ ID NO: 555	EIKTYRMDHVAIDRISGGAVDQCKYDDEPLVGTSKHPLVFKGMFWINR
FmCas7-11	
SEQ ID NO: 556	EVREKLFHDHVSIDRFTGGAANKLKFDKPLVG---NPLVFQGVFWVHQ
DpbaCas7-11	

[0321] Table 13 shows the representative Cas7-11 orthologs of Figure 1J.

Table 13.

SEQ ID NO/ NAME	SEQUENCE
SEQ ID NO: 557	PVAGFALQNQPAIPATSLRGLLSSLFESISGSNLRVLHPTPYRSIRKTT
WmCas7x3	
SEQ ID NO: 558	EARPFKLNGKHAIPATSLRGLLSSLFESVSNSNFRVLHPEHYSVRKSL
NisCas7x3	

Table 13.

SEQ ID NO\ NAME	SEQUENCE
SEQ ID NO: 559 GwCas7x3	SLHPFKLNNGLAIPATSLRGMISSLFESVSNSNFRVLDEKTYSMRKT
SEQ ID NO: 560 HvmCas7x3	EILPFTLGGKPAIPATSLRGMLSSIAEAASNSLRVLEDKTL SYRKSM
SEQ ID NO: 561 DsbCas7x3	QVDHYTENGEIAIPATTLRGLLSSLSEAASNSMRVLDDGMMSYRQPV
SEQ ID NO: 562 GamCas7x3	LRHPYQLDGEPAPSTSLRGLISTMTEAAANCAMRVL DSEIISYRKPM
SEQ ID NO: 563 DesCas7x3	RIENYRLGNRIAIPAASLRGMLSSLAEASNSAMRVLHQGILSYRKKA
SEQ ID NO: 564 GabCas7x3	LKEHYQLNNKLAIPATSLRGLISSLAEEASNSALRVLDNGVLSYRKPA
SEQ ID NO: 565 MetCas7x3	EINNYRLNGELAIPATSLRGMISSLAEEASNSAMRVLDNGLLSYRKTA
SEQ ID NO: 566 MebCas7x3	ELDNYRLNGQLAIPATSLRGMISSLAEEASNSAMRVLDNGLLSYRKDA
SEQ ID NO: 567 GwCas7-11	TQRFHQINDEIGLPGASLRGMVLSNYQILTNSCYRNKATEEITRRMP
SEQ ID NO: 568 HreCas7-11	NYRFFRINDELAIPGSELRGMVSSVYEALTNSCFRIMEEGRYLSRRMG
SEQ ID NO: 569 CmaCas7-11	NYAFFRINDHIAIPGASIRGMISSVFETLTHSCFRVMDDKKYLTRRVI
SEQ ID NO: 570 HvmCas7-11	NYGFFRINGNVAIPGSSIRGMISSVFEALTNSCFRVFDQERYLSRSEK
SEQ ID NO: 571 SmCas7-11	NFPFFKINDEIMLPAGAPLWAAVSQVYEALTNSCFRVMKQKRFLSWRME
SEQ ID NO: 572 OmCas7-11	EYPSFRLNNTPMIPGAGLRAAVSQVYEVLTNSCIRIMDQGQTL SWRMS
SEQ ID NO: 573 DisCas7-11	SYAFFRLHKQIMIPGSELRGMVSSVYETVTNSCFRIFDETKRLSWRMD
SEQ ID NO: 574 DsbCas7-11	SYRFFTLNRVPMIPGSEIRGMISSVFEALSNSCFRIFDEKYRLSWRMD
SEQ ID NO: 575 SstCas7-11	SYAFFSVNGDIMLPGSEIRGMLSSVYEALTNSCFRVFDEGYRLSWRME
SEQ ID NO: 576 SybCas7-11	MCKFFSVAGKPMIPGSEIRGMISSVYEALTNSCFRVFDEEKYLTRRVQ
SEQ ID NO: 577	NYKFFNINGELMIPGSELRGMLRTHFEALTKSCFAIFGEDSTLSWRMN

Table 13.

SEQ ID NO\ NAME	SEQUENCE
CsbCas7-11	
SEQ ID NO: 578	NYQFFHINDEIMVPGSEIRGMISVYEAITNSCFRVYDETKYITRRLS
CjcCas7-11	
SEQ ID NO: 579	TYDFFQMNNAIMIPGSEIRGMISAVYEAMTNSCFRIFHEKQYLTRRIS
CbfCas7-11	
SEQ ID NO: 580	YYHFLKIDNKPIIPGAIEIRGAVSSIYEALTNSCFRVFGEKKVLSWRME
HvsCas7-11	
SEQ ID NO: 581	VYKFFRLGGHLCIPAAEIRGMISVYEALTNSCFRVFDEKRLISWRMT
HsmCas7-11	
SEQ ID NO: 582	SFRFFRIDDEVLPISGSEIRGMVSTVFEALTGSCFRVINQKAHLWRIN
FmCas7-11	
SEQ ID NO: 583	RKFMRMGSAQAIPGSAIRSMTSSVFEALTNSCFRVLDQKSHLSWRME
DpbaCas7-11	

[0322] Table 14 shows the representative Cas7-11 orthologs of Figure 1K.

Table 14.

SEQ ID NB\ NAME	SEQUENCE
SEQ ID NO: 584	E-----ALSA--IGRI-VE---RNGELKLYPL
WmCas7x3	
SEQ ID NO: 585	DYV-----ALSA--MGRI-VD---DQGELKLQPL
NisCas7x3	
SEQ ID NO: 586	Q-----SLSA--MGRI-VR---HDQKLYLLPL
GwCas7x3	
SEQ ID NO: 587	ANRN-----EDKPLSA--LGMI-KKIETGDKVEYRLLPL
HvmCas7x3	
SEQ ID NO: 588	SG-----SLSA--IGMV-VI---RDGKKFIYPL
DsbCas7x3	
SEQ ID NO: 589	PSH-----ILSA--LGMV-TK---RGEDFWLIPL
GamCas7x3	
SEQ ID NO: 590	N-----ALRE--IGMI-VL---RDGKRFILPL
DesCas7x3	
SEQ ID NO: 591	R-----ALRK--IGIL-FK---REEQWRLVQM
GabCas7x3	
SEQ ID NO: 592	D-----ALRK--VGMV-IY---VDNKSFIKL
MetCas7x3	
SEQ ID NO: 593	L-----ALSK--IGITFIN---RQGQWQLIPM

Table 14.

SEQ ID NB\ NAME	SEQUENCE
MebCas7x3	
SEQ ID NO: 594	DEA-----KYRK--AGRV-TVSGDGAQKKYSIQEM
GwCas7-11	
SEQ ID NO: 595	D-----EFKDFHPGIV-----VDGAKIREM
HreCas7-11	
SEQ ID NO: 596	ESETTQ-KRKSGRYQVEESDPDLF---PGRV-QK---KGNKYKIEKM
CmaCas7-11	
SEQ ID NO: 597	DPT-----ELTKYYPGKV-KR---DGNKFFILKM
HvmCas7-11	
SEQ ID NO: 598	E-----DYKDFYPGRV-----LDGGKQIKKM
SmCas7-11	
SEQ ID NO: 599	SEHK-----DYQ---PGKI-----TDNGRKIQPM
OmCas7-11	
SEQ ID NO: 600	DHQN-----VLQDFLPGRV-----TADGKHIQKF
DisCas7-11	
SEQ ID NO: 601	DVK-----ELEQFKPGRV-----ADDGKRIEEM
DsbaCas7-11	
SEQ ID NO: 602	DRN-----VLMQFKPGRV-----TDNGLRIEEM
SstCas7-11	
SEQ ID NO: 603	KKGA-----KSSELVPGII-VWG---QNGGLAVQQV
SybCas7-11	
SEQ ID NO: 604	DEKDYKIDNSIRKMEsqRNPKYRI--PDEL-QKELRNsgnGLFNRLY
CsbCas7-11	
SEQ ID NO: 605	EKKD---ESNDKNKSQDDASQKIR---KGLV-KK---TDEGFSIIEV
CjcCas7-11	
SEQ ID NO: 606	EDK-----ELREFIPGIV-RI---INGDVYIEKA
CbfCas7-11	
SEQ ID NO: 607	KDAK-----EFM---PGRV-SK---KKGKLYMVKM
HvsCas7-11	
SEQ ID NO: 608	EEAK---RPDPKKSEEQNRMRFR---PGRI-IK---KDKKFYAQEM
HsmCas7-11	
SEQ ID NO: 609	DMAK-----HYR---PGRI-IQ---NNEKMFIQPY
FmCas7-11	
SEQ ID NO: 610	DDAG-----DYK---PGRF-EK---KDDKAVIRKF
DpbaCas7-11	

Table 15 shows sequence alignments of cas nucleases.

DOMAINS	
Mutated residues; catalytic residues	
Domain 1 - Csm3-like (group 7 RAMP) with abnormal “G-rich” loop Domain 2 – putative small subunit (no sequence similarity but there is a conserved “W” and four large alpha helices predicted by Jpred in this region) Domain 3 - Csm3-like (group 7 RAMP), could be catalytic Domain 4 – Csm3-like (group 7 RAMP), * - catalytic aspartates based on Zhu X, Ye K. Nucleic Acids Res. 2015 Jan;43(2):1257-67; https://doi.org/10.1093/nar/gku1355) Domain 5 – Csm3-like with large insertion or subdomain (group 7 RAMP)	
WmCas7x3, GwCas7x3, SER16298.1, PID64649.1, WP_031436019.1, HEB50754.1 - III-D systems	
SECTION 1	
SEQ ID NO: 611	
WmCas7x3	-----
SEQ ID NO: 612	
SER16298.1	-----
SEQ ID NO: 613	
GwCas7x3	-----
SEQ ID NO: 614	
PID64649.1	-----
SEQ ID NO: 615	
WP_031436019.1	-----
SEQ ID NO: 616	
HEB50754.1	-----
SEQ ID NO: 617	
HrcCas7-11	--MSVEEFYVRLTFL-EPFRVVPWVRNGDERKGDRIYQRGGTYARWHKI----
NDSHGQPYITGTMLRSALVREIENTLT LHNTY---GCCPGGT--RTTEGKLEKPL-YLRRRD--	
GFEFENH-AEKPC-----SEEDPCPLCLIQGRFDKLRDEKKQFVRQGNISFCNVFNSNLNISS--	
GIKSFSWE---EI	
SEQ ID NO: 618	
CmaCas7-11	----MLKLKVKITYF-QPFRVIPWIKEDDRNSDRN-YLRGGTFARWHKDKK---
DDIHGKPYITGTLLRSALFTEIEKIKIHHSDFIH---CCNAID---RTEGKHQPS-FLRKRK----	
VYTENKNIQAC-----NKCPCLIMGRGDDRGEDLKKKKHYNGKHYQNWTVHFSNFD-----	
TQATFYWK---DI	
SEQ ID NO: 619	
HvmCas7-11	----MTQITIQTFF-HPFRVVPWNHRDHRKTDRK-YLRGGTFAKWHCTAS---
EGKSGRPYITGTLLRSALFAEIEKLI AFHDPF---KCCRGKD---KTENGNAKPL-FLRRRP-----	
RADC-----DPCGTCPLCLLMGRSDTVRRDAKKQKK-----DWSVHFCNLR-EA---TERSFNWK---ET	
SEQ ID NO: 620	
SmCas7x3	----MRLKINIHFL-EPFRLIEWHEQDRRNKGNRSRWQRGQSFARWHRRKD---NDQ-
GRPYITGTLLRSVIRAVEEELARPDATWQ---SCGGLF---ITPDGQTKPQ-HLRHRA--TVRARQT-	
AKDKC---A-DRQSACPFCLLLGRFDQVGKDGDKKGEGL-----RFDVRFNSNLD-LPKDFSPRDFDGPQ--	
EI	
SEQ ID NO: 621	
oral_meta	MIPDLRSLVVHISFL-TPYRQAPWFPPEKRRNNNRDWLRMQSYARWHKVAP---EE--
GHPFITGTLLRSVIRAVEEELCLANGIWRGVACCPGEF---NSQAKKKPK-HLRRRT--TLQWYPE-	
GAKSCSKQD-GRENACPFCLLLDRFGGEKSEEGRKKNN-----DYDVHFSNLPFYPGSSPKVWSGPE-	
-EI	
SEQ ID NO: 622	

DisCas7-11 ---MTTMMKISIEFL-EPFRMTKWQESTRRNKNKEFVRGQAFARWHRNKK---
DNTKGRPYITGTLRSAVIRSAENLLTSLSDGKISEKTCCPGKF---DTEDKDRLL-QLRQRS--TLRWT---
DKNPC--PD-NAETYCPFCCELLGRSGNDGKKAEEKDW-----RFRIHFGNLS-LP---GKPDFDGP--AI
SEQ ID NO: 623

DsbaCas7-11 -----MKITLRL-EPFRMLDWIRPEERISGNKAFQRGLTFARWHKSKA---DDK-
GKPFITGTLRSAVIRAAEHLVLSKGVGEKACCPGKFLTETDTETNKAPTMLRKR--TLKWT---
DRKGC-----DPDFPCPLCELLGPGAVGKKEGEAGI-----NSYVNFNLS-FP---GDTGYSNAR--EI
SEQ ID NO: 624

SstCas7-11 ----MIINITVKFL-GPFRMLEWTDPDNRNRKNREFMRGQAFARWHNSNP---QKG-
SQPYITGTLVRSVIRSAENLLMLSEGKVGKEKCCPGEF--RTENRKKRDAMLHLRQRS--TLQWKT--
DKPLC----NGKSLCPICELLGRRIGKTDEVKKKG-----DFRIHFGNLT-----PLNRYDDPS--DI
SEQ ID NO: 625

CsbCas7-11 -----MNITVELTFF-EPYRLVEWFDWDARKKSHS-AMRGQAF AQWTWKGGK---
GRTAGKSFITGTLVRSVIRSAEELLNLNGKWEVPCNGSF--QTDESKGKKPS-FLRKRH--
TLQWQAN-NKNIC-----DKEEACPFCELLGRFDNAGKVHERNK-----DYDIHFSNFD-LD---
HKQEKNDLRLVDI
SEQ ID NO: 626

CjcCas7-11 ---MHTILPIHLTFL-EPYRLAEWHAKADRKKNR-YLRGMSFAQWHKDKD---GI--
GKPYITGTLRSAVLNAEELISLNQGMWAKEPCCNGKF-----ETEKDKPA-VLRKR--TIQWKTG-
RPAICDPEKQEKDACPLCMLLGRFDKAGKRHRDNKYDKH---DYDIHFDNLN-LI---TDKKFSPHD--
DI
SEQ ID NO: 627

CbfCas7-11 MSKTDDKIDIKLTFL-EPYRMVNWLENGLRMTDPR-YLRGLSFARWHRNKN---
GKA-GRPYITGTLRSAVIRAAEELSLNLGKWKQLCCPGQF--ETEREMRKNKT-FLRRRP--
TPAWSAETKKEIC---T-THGSACAFCLLGRRLHGGKEDVNEDAPGS---CRKPVGFGNLS-LP---
FQPTKRQIQ--DV
SEQ ID NO: 628

HvsCas7-11 ----MKSIPITLTL-EPYRILPWAEEKGRDKKE--YLRGANYVRLHKDKN---GK--
FKPYITGTLIRSAVLSAIEMLLDITNGEWNGKECCAKF-----HTEGEKPS-FLRKKP---IYIRAEKDEIC-
--T-SRETACPLCLILGRFDKAEKKEKDKE-----KFDVHFSNLN-LY---SSKEFSTIE--EL
SEQ ID NO: 629

HsmCas7-11 ----MTKIPISLTFL-EPFRLVDWVSESERDKSE--
FLRGLSFARWHRIKNQREDENQGRPYITGTLRSAVIRAAEELIFLNGGKWQSEECNGQF--
KGSKAKYRKVE-CPRRRHRATLKW---TDNTC---S-DYHNACPFCLLGLCKPNSKE-----
NSDIHFSNLS-LP---NKQIFKNPP--EI
SEQ ID NO: 630

FmCas7-11 ----MPRFQLSLTFFDEPFRLIEWTDKSNRNSANTQWMRGQGFARWHKITL---EK--
GFPFVTGTAVRSKIIREVEALLSRNKGTWNGIPCCSGFF-----DTKGPSPT-HLRYRP--TLEWEY--
GKTVC---T-SEADVCPCLLLGRFDQAGKSDTPCQST---DYHVHWNLS--A---GVAQYR-LE--DI
SEQ ID NO: 631

DpbaCas7-11 -----F-E-----SYARWCKSNS---GL--
WKPYIPGTLRSAVLESVEYLLAL-IGSKNKVEICPGLY---TQSENNPDTK-YLRRRP--WYELHA--
QKEIC---K-TRDTACPLCLLMRTKLDNDGDGETEK-----NVKFGNL-----YPTSPLEPL--QK

Domains

.....
.....

SECTION 2

	D177	atypical G-rich loop
WmCas7x3	-----	-----
SER16298.1	-----	-----
GwCas7x3	-----	-----
PID64649.1	-----	-----
WP_031436019.1	-----	-----
HEB50754.1	-----	-----

HreCas7-11 AVSRVVRVDPNSGKAKDFFRVWEIDHKLCPNFLGKMSISLS---
 EKLEDVKALLAVGLAQVNVLSGALCRVDIID-----
 PETQKDTVHQHLIQFVTRIQQDKEKGDAADIPAFTLPPAGLSPSSNEWNDTIKSLAEKIR--
 KIKELEQGQKLRQMADVIREL-RRKTPAYLDQ---
 CmaCas7-11
 VQKRILNRVDQTCGKAKDFFKVCEVDHIACPTLNGIIRINDEKLSQEEISKIKQLIAVGLAQIESLAGGI
 CRIDITN-----QNHDDLKISFFETKPSKILQPNL---
 KESGEERFELAKLELLAEYLT-QSFDANQKEQQLRRLADAIRDL-RKYSPDYLDK---
 HvmCas7-11 AIERIVNRVDPSSGKAKDYMRIWEIDPLVCSQFNGIITINLD---
 TDNAGKVKLLMAAGLAQINILAGSICRADIIS-----
 EDHDAIKQFMAIDVREPEVSTSFPLQDDELNNAPAGCGDDEISTDQPVGHNLVDRVRISKIAESIE-
 DVFSQEQAQQLRRMADAIRDL-RRSKPDETTLDA--
 SmCas7x3 GSRRTINRVDDDETGKAHDFFSIWEVDA--VREFQGEIVLAAD-L--
 PSRDQVESLLHHALGFVDRLCGARCVISIAD-----
 QKPAEREERTV---AAGDEKATIADYDQVKGLPY-----TRLRPLADAVRNL-RQLDLAELNKPDKG
 oral_meta GRLRTLNRIDRLTTKAQDFFRIEVDQ--VRDFFGTITLAGD-L--
 PRKVDVEFLLRRGLGFVSTLCGAQCEIKVVD-----
 LKKKQNNKEDSI---LPVSEVPFFLEPEVLAKMCQ-DVFPS---GKLRLADVILRL-REEGPDNLT----
 DisCas7-11 GSQRVLNRVDFKSGKAHDFFKAYEVDHTRFPRFEGETIDNK-V--S--
 AEARKLLCDSLKFTDRLCGALCVIRFDE-----YTPAADSGKQT---
 ENVQAEPNANLAEKTAEQII-SILDDNKKTEYTRLLADAIRSL-RRSSKLVAG----
 DsbaCas7-11 AVRRVVRVDYASGKAHDFFRIFEVDHIAFPCFHGEIAFGEN-V--S--
 SQARNLLQDSLRFTRDRLCGALCVIRYDG-----DIPKCGKTAPLPETESI--
 --QNAAEETARAIVRVFHGGRKDPEQAQIDKAEQIQLLSAAVREL-GRDKKKVSA----
 SstCas7-11 GTQRTLNRVDYATGKAHDFFKVWEIDHSLLSVFQKGKISIADN-I---
 GDGATKLLLEDRLFTDRLCGAICVISYDC-----
 IENS DGKENGKTGEAAH----IMGESDAGKTDAENIANIAIA-DMMGTAGEPEKLRLADAVRAL-
 RIGKNTVSQ-----
 CsbCas7-11 ASGRILNRVDFDTGKAKDYFRTWEADYETYGTYTGRITLRN-----
 EHAKKLLLASLGFVDKLCGALCRIEVIK-----
 KSESPLPSDTKEQSYTKDDTV----EVLSEDHNDLRLKQAEVIV-EAFKQNDKLEKIRILADAIRTL-
 RLHGEVIEKDE--
 CjcCas7-11 ASERILNRVDYTTGKAHDFYFKVWEVDDDQWWQFTGTITMHDD-----
 CSKAKGLLLASLFCFVDKLCGALCRIEVTG-----
 NNSQDENKEYAHPDTGIITSLNLKYQNNSTIHQDAVPLSGSAHDNDEPPVH--
 DNDSSLDNDTITLLSMKAKEIV-GAFRESGKIEKARTLADVIRAM-RLQKPDWEEK---
 CbfCas7-11 CKERVLNRVDFRTGKAQDYFRVFEIDHEDWGVYTGEITITE-----
 PRVQEMLEASLKFDVTLCGALCRIEIVG-----
 SADETKRTTSSKEGCPASTTTRDCSSSEND---DTSPEDPVREDLKKIAHVIA-
 NAFQNSGNREKVHALADAIRAM-RLEESSIINT---
 HvsCas7-11 APKRALNRIEQYTGKAQDYFTVYEALNKEFWTFKGRIRIKED-I----
 YDKVTDLLFSALRCVEKIAAGALCRIEIDK-----EPSQQKGFVK-----
 -----RQLSKQAKEDIE-KIFQVVKDAQKLRLSDCFRELTRMANKDELA----
 HsmCas7-11 GIRILNRVDFTTGKAQDYFYVWEVEHSMCPKFQGTVKINED-M--
 PKYNNVVKDLLISSIQFVDKLCGALCVIEI-----GKTKNYIC---
 QSFSNNIPEEEIKKLAQEIR-DILKGEDALDKMRVLADTVLQM-RTKGPEIVNE---
 FmCas7-11 AQKRTSNRVDFFSKKAHDHYGVWEVTA--VKNLLGYIYISDA-
 ITESHQKTVISLLKAALSFTDTLCGANCKLELSD-----
 EPVDSIHSNQASNFNPHSGAAPSQCSQSM---PFNMDQETKELANTLCKAFT-----
 GNMRLRLTLADAVREM-RRMSPGISS-----
 DpbaCas7-11 IRPRILNRMDPGTSKARDYFRVFEIENQLCSQFRGWIWLSGD-L--
 PNMELVKSLLAAGLSNVATLAGAVCRIRIVSTDNPSMKQDLTTQDLIDDFTNYYLKGDTPPANLAAS
 GKGDAPFRFSPGSGDHPDTTGVSHA---DMASSHEGTALAKDIAEKCK-
 DILSQISASEQLRRLADIMRDL-RQDSNREIMYRQ--
Domains

SECTION 3**D429**

WmCas7x3 -----
 -----MTTPSAPKSSL-----PA---LHWLIRAELEVLTPHLHG-----TGTDQRITPDAPDA-----
 DPY
 SER16298.1 -----
 -----MTDTV-----KS---GRWIITGQFQLVTPMHIG-----
 TGLDEEMDKQSGESVDKKQNNNS
GwCas7x3 -----
 -----MSNQ---TRWIEGTLELITPLHIG-----TGLDE-----KERDENKETR
 PID64649.1 -----
 MSDTQKQAIHENKWHF----RGI---KRWEISAYLKTLSPLHIG-----DG---
 GTIPVTIKDTQGKNREV
 WP_031436019.1 -----
 -----MNLPTWKL-----NNE---KRWHISIVLTTATPLHIG-----SG--
 EFCEHDDVKNNNDG--EPV
 HEB50754.1 -----
 -----MAAVQDRWTLMDQQGNEL---KRFRITAELETASSLHIG-----AS--
 ETVEHDLIKNDDG--TPV
HreCas7-11 -LPAGKPEG-----RESIWEK-----TPTGE--TLTLRQLLKSANVPGE-----
 SWRAFCEELGEQLYRLEKNLYS--HARPLPRLLGETEFYGPARKS---DDPPMIRASYR---AFPS----
 YVWVLGDGILRAETPFYFG-----TET-----SEGQTS-
CmaCas7-11 -LPKGKGG-----RTSIWNK-----KVADD--FTLRDCLKNQKIPNE-----
 LWRQFCEGLGREVYKISKNISN--RSDAKPRLLGETEYAGLPLRKE--DEKEYSPTYQN---QESLPK--
 TKWIISGELQAITPFYIG-----HVN-----KTSHTR-
HvmCas7-11 -LPKGKTDK-----
 DNSVWDKPLKKDILSPRMPASEDDDTPTLRKVLKDEINGQED-----
 MWRKFCEALGNSLYDL SKK AKERKRTEALPRLLGETEYGLPMREN--KEDEPLPSSLT---YK-----
 FKWLIAGELRAETPFFFG-----TEV-----QEGQTS-
SmCas7x3 FLPPGRV NK---DGRRVPHYVWDIP-----LGKGD---TLRKRLEFLAASCEGDQA--
 -KWRNICESEGOALYKSKKLKD--SPAAPGRHLGAAEQVRPPQPPV---SYSEESINSDL---PL-----
 AEWIITGTLRAETPFAIGM-----DAPI-----DDDQTS-
 oral_meta -LPMGSQGL---GGRL-PHHLWDVP-----LVSKDRETQTLRSCLEKIAAQCKSEQT--
 QFRLFCQKLGSSLFRINKGVYL--APNSKISP---EPCLDPSKTI--RTKGPVPGKQK---
 HRFSLLPPFEWIITGTLKAQTPFFIP-----DEQG-----SHDHTS-
DisCas7-11 -LPKDHDGK---D---DHYLWDIG-----KKKKDENSVTIRQILTTSADTKELKNAG--
 KWREFCEKLGEALYLKSKDMSG--GLKITRRILGDAEFHGKPDRL--KRSRSVSIGSVL-----
 KETVVCGELVAKTPFFFG-----AIDE-----DAKQTD-
DsbaCas7-11 -LPLNHEGK---E---DHYLWDK-----KAGGE---TIRTILKAAAEKEAVAN---
 QWRQFCIELSEELYKEAKKAHG--GLEPARRIMGDAEFSKSDKVPDT--VSHSIGISVE-----
 KETIIMGTLKAETPFFFG-----IESK-----EKKQTD-
SstCas7-11 -LPLDHEGK---E---NHHLWD-----IGEGK---SIRELLLEKAESLPSD-----
 QWRKFCEDEVGEILYLKSKDPTG--GLTVSQRILGDEAFWSKADRQL--NPSAVSIPVT-----
 TETLICGKLISETPFFFG-----TEIE-----DAKHTN-
CsbCas7-11 -LPDGKEER---DK---GHHLWDI-----KVQGT---ALRTKLKELWQSNKDI----
 GWRKFTEMLGSNLYLIYKKETG--GVSTRFRILGDTEYYSKAHDSE---GSDLFIPVTPP---EGIET---
 KEWIIIVGRLLKAATPFYFGVQQPSDSIPGKEK-----KSEDSL V-----INEHTS-
CjcCas7-11 -LPKGINDK-----HHLWDR-----EVNGK---KL RNILEELWRLMNKRN----
 AWRTFCEVLGNELYRCYKEKTG--GIVLRFRITLGETEYYPEPEKTEPCLISDNSIPITPL---GGV-----
 KEWIIIGRLKAETPFYFGVQSSFDSTQDDLDLVPDIVNTDEKLE-----ANEQTS-
CbfCas7-11 -LPKGKSEKTTEQIEVNKHLYWDEI-----PVNDT----
 SVRHILIEQWRRWQSKKDD-PEWWKFCDFLGECLYKEYKKLTS--
 GIQSRARVMGETEYYGALGMPD---KVIPLLKSDKT-----KEWILVGSLLKAETPFFFGLET-----
 -----EQTE-----EVEHTS-
HvsCas7-11 -LPLGPEDD-----GHYLWDKI-----KVEGK---TLRIFLRNCFSSQYKD-----
 NWLCFCDEASKKGYQKYREKRH--KLT DRELPTATPKHFAEKKDPQ---ISPIYIDKDDK---V-----
 YEWIIVGRLLIAQTPFHFG-----DEEKAE-
HsmCas7-11 -LPRGIEKK---G---GHWLWDKL-----R-----LRKKFKEIANNYKD-----
 SWQELCEKLGNELYISYKELTG--GIAVKKRII GETEYRKIPEQEI---SFLPSKAGYS-----
 YEWIILGKLISENPFFFG-----KETK-----TEEQID-

FmCas7-11	-LPRGRLNK---EGEITAHYLWDE-----RIDEK----
TIRQVLEDTIELSPARSIHYKNWISFCNQLGQKLYERAKD-ND--PILERKRPLGEAAFSKVPTSSH---	
APRHDMNSRVK----	GGFT---REWIIVGTLRALTPFYMG-----TGSQ-----AGKQTS-
DpbaCas7-11	-VAEENHEK-----ASLLYKK-----TKKGD---SIAALIAGKTEGMDA-----
ETWRRRLCEFLGQTFYGEAKEAGL--VETPVPRILGESERYSLQKKPT---VRTDLAAELVP----D-----	
IEFIKGNLIAETPFFFG-----TDIA-----TETHTD-	
Domains	
.....	
SECTION 4	
	*
G-rich loop	
WmCas7x3	WQADIALDADGRPYLPGLASLKGALKALARRRQVDAPCL-
PLFGDLNRGGAPHPDCIPPRRTRAGLAEFRDALQSHAT--QAD--GPD-----	
AQPRIADRITGSVVDKLFHTQTVPVGTFRFSVEI--ILRRADQ--NLAAQLVALLQHGPT----	
DPDFRLGAHANLGFGRVGL-YGNIDTRRFGPQQAQ-H	
SER16298.1	WIQAIALDLNKKPYIPGASIKGALKALARRYCASNL--NIFGDTI--
DTKDGDNKRKSVTVAGQAEFLNA-WYAAD--QED--KPF-----	
TITRVAIDRVGTAEADRKLFNTRRVNPGVCFNYKI--IIQNACE--TEIQYLLDLLRKAAC----	
DPSFSLGAGANQSQGVRC-LSSC-VRYFGKQEMH-D	
GwCas7x3	WLEAVALDHKGQPYIPGASLKGALRALAKRHD-----FRNLF--DNKEVDGD-----
FVRQAEFLSA-WCVPD--TDK--GRL-----IQPRVAIDRVGTGAQDKLQFTRLITPGTRFAMKI--	
VVQNAVE--NEIADLLGLLNLLPD-----DPQFSLGAYANQGQGRVQW-FGKIQTRCFGINEAK-A	
PID64649.1	EVNSVITGKAALPIPGSTIKGRLRHYFSKHFSKALLNKVFGEES--DATDDDQG----
RGGLAEFHDAKWNEK--	
NRNLQGRYPYWNTRKTYIEVSTAINRHTGAADKSLHHTECVPPGTVFEIKITGSMDDRCA--	
ALVVAALAEIQTTS-----RIFL--GAEDANGNGRIGL-TGKITVKQMDQAHII-Q	
WP_031436019.1	KINACIKGSKGRPIPGSTIKGKLYEWLKRNTNTEENLLEKLFQKGH--
NSVSQDQG-----RGGKAEFHDAEIIPL--TGS--	
QPWPYWREEHQAFIAASTAIDRHKQVALQQSLHYMETVPAGIRFKFTFTGVMRDEEA--	
ALLIAALDSFDKNQN-----QPCF--GVDRANAYGRMEL-HGHLHVKVMGATEIS-S	
HEB50754.1	QINALITGAGGLPIPGSTIKGRFLARLRERGVDSALLETLFGKGH--DRETEDQG----
RGGRAEFHDAPLCHRL--SGA--	
RHFPYWRPERQIWVKAQTAVDHRGTALRRSLRYTEMVPPGVRFRLTITGCMTDAEA--	
DVLFALLEDLGDPQ-----ACSF--GGAGADNGTMRL-FGRPEVYCLDRSGIL-G	
HreCas7-11	--QAIIICPDGSYRLPRSLRGVIRRDILRAIL-GTGCNVSLGKVRP--CSCPVE-----
IMRRITVQQG---VSS--YRE--PAE-----VRQIRSNPHTGTVEEGALFDLETGPQGMTPFRL----	
YFRTRSPYIDRALWLTINHWQE-----GKAI-FGGDIGVGMGRFRL-E-NLQIRSADLVSRD	
CmaCas7-11	--STIFLNMNGQFCIPRSTLRGALRRDLRLVF-GDSCNTPVGS-RV--CYCQVCQ-----
IMRCIKFEDA---LSD--VDS--PPE-----VRHRIRLNCHTGVVEEGALFDMETGFQGMIFPRL----	
YYESKNEIMSQHLYEVLNNWTN-----GQAF-FGGEAGTGFRFKL-L-NNEVFLWEIDGEEED	
HvmCas7-11	--ATILLNRDGYFRLPRSVIRGALRRDLRLVMGNDGCNMPIGG-QM--CECGVCR----
--VMRHIVIEDG---LSD--CKI--PPE-----VRHRIRLNCHTGTVEEGALFDMETGYQGMTPFRL----	
YCETENSOLDLSYLWEVLNNWQN-----GQSL-FGGDTGTGFRFEL-T-EPKVFLWNFSKKE-K	
SmCas7x3	--SRTLVDGRYRLPRSTLRGILRRDLRLSLASGDQGCQVRLGPERP--CTCPVCL-----
ILRQVVIADT---VSE--TTV--PAD-----IRQIRRNPTGTAAAGGLFDTERGPKGAGFPFSL---	
RYRGH-APMPKALRTVLQWWSA-----GKCF-AGSDGGVGCGRFAL-D-NLEVYRWDLGTFA-F	
oral_meta	--RKILLTRDFYRLPRSLRLGIIRDLHEATDKGGCRVELAPDVP--CTCQVCR-----
LLGRMLLADT---TST--TKV--APD-----MRHRVGVDRCGIVRDGALFDTEYGIEGVCFPLEI---	
RYRGN-KDLEGPIRQLLSWWQ-----GLLF-LGGDFGIGKGRFRL-E-NMKIHRWDLRDES-A	
DisCas7-11	--LQVLLTPDNKYRLPRSAVRGILRRDLQTYF-DSPCNAELGG-RP--CMCKTCR-----
IMRGITVMDA---RSE--YNA--PPE-----IRHRTRINPFTGTVAEGALFNMEVAPEGIVFPFQL----	
RYRGSEDGLPDALKTVLKWWAE-----GQAF-MSGAASTGKGRFRM-E-NAKYETLDLSDEN-Q	
DsbaCas7-11	--LMLLLDQGNHYRIPRSALRGILRRDIRSVL-GTGCNAEVGG-RP--CLCPVCR-----
IMKNITVMDT---RSS--TDT--LPE-----VRPRIRLNPTGTGSVQEKALFNMEMGTGIEFPFVL---	
SYRGK-KTLPKELRNVLNWWTE-----GKAF-LGGAASSTGKSIFQL-S-DIHAFFSSDLSDET-A	
SstCas7-11	--LKVLLDRQNRYLPRSAIRGVLRRDLRTAFGGKGCNVELGG-RP--CLCDVCR-----
IMRGITIMDA---RSE--YAE--PPE-----IRHRIRLNPTGTVAEGALFDMELGPQGLSDFIL---	
RYRGKGSIPKALRNVLKWWTK-----GQAF-LSGAASSTGKGIFRL-D-DLKYISFDLSDKD-K	

CsbCas7-11 --FNILLDKENRYRIPRSALRGALRRDLRTAF-GSGCNVSLGGQIL--CNCKVCI-----
 EMRRITLKDS---VSD--FSE--PPE-----IRYRIAKNPGTATVEDGSLFDIEVGPEGLTFPFVL---
 RYRGH--KFPEQLSSVIRYWEENDGKNGMAW-LGGLDSTGKGRFAL-K-DIKIFEWDLNQ---K
 CjcCas7-11 --FRILMDKKGRYRIPRSIRGLRRDLRTAFGGSGCIVELGRMIP--CDCKVCA-----
 IMRKITVMDS---RSE--NIE--LPD-----IRYRIRLNPYTATVDEGALFDMEIGPEGITFPFVF---
 RYRGE-DALPRELWSVIRYWMD----GMAW-LGGSGSTGKGRFAL-I-DIKVFEWDLNCEE-G
 CbfCas7-11 --LRLVMDKKGRFRIPRSVLRGALRRDMRIAF-DSGCDVKLGSP--CDCSVCQ-----
 VMRSITIKDS---RSE--AGK--LPQ-----IRHRIRLNPFSGTVDEGALFDIEVAPEGVIFPFVM---
 RYRGE--EFPPALLSVIRYWQD----GKAW-LGGEGATGKGRFALAK-DLKMYEWKLEDK--S
 HvsCas7-11 --GAILLTPDNRFRLPRTALRGILRRDLKLAG-ASACEVEVGRSEP--CPCDVCK-----
 IMRRVTLLDT--VSEDLRDF--LPE-----LRKRIRINPQSGTVAEGALFDTEVGPEGLSFPFVL---
 RYKCE--KLPDSLTTVLCWWQE----GLAF-LSGESATGKGRFRL-E-INGAFVWDLQK---G
 HsmCas7-11 --MQILLTKDGRYRLPRSVLRGALRRDLRLVI-GSGCDVELGSKRP--CPCPVCR-----
 IMRRVTLKDA--RSD--YCK--PPE-----VRKRIRINPLTGTVQKGALFTMEVAPEGISFPFQL---
 RFRGE-DKFHDALQNVLVWWKE----GKLF-LGGGASTGKGRFKL-E-IEHVLKWDLKN---N
 FmCas7-11 --MPTLQDSNDHFRLPRTALRGALRRDINQASDGMGCVELGPHNL--CSCPVCQ-----
 --VLRQIRLLDT--KSK--FSM--PPA-----IRQKICKNPVLSIVNEGSLFDVELGIEGETFPFVM---
 RYRGG-AKIPDTIITVLSWWKN---ERLF-IGGESGTGRGRFVL-E-CPRIFCWDVEK---G
 DpbaCas7-11 --LPILLTSDRHFRIIPRSVLRGILRRDLRLVT-GSGCSVKLGRSEP--CACDVCQ-----
 IMRSLTMRDC---VSS--CKV--PPE-----IRHRIRLNPVTETVEEGALFDMEIGPQGISFPFVL---
 RSRGVNSSFSTRLKNVLTWWSE----GKIF-MGGDKGTGKGRFTL-A-ELEAYFRLTTKRIG

Domains**SECTION 5****D654**

WmCas7x3 WF----AAQQTQADARW-TDFAEA---VT-----LTAPAPAPAQP-----
 APHRLALP-----LSLTFHTPFLVKQPEHKH--R---KPQDNAPDGTTPR---QRGD-----
 ---RA---LLPGASLRGRLRSQAERILRTLGLCK
 SER16298.1 WF----RAIQNGKQEHW-QIFAKP---SN-----IKYADLERPDI-----
 IANSLNLP-----LTLDFTHTPFLVKASKKK-----DEAKNEADAKPR---TNHQG-----
 ---QV---ILPASSLRGRLRAQAEKILRTMGQD
 GwCas7x3 WY----EEIRKDESKCW-TAFAPK---KN-----VSTP---PTPA-----
 KEAQLTLP-----LNLAFHTPFLVKQAGIK-----ADDADAVPR---RTHDD-----
 --KI---VLPASSLRGRLRTQSERILRTLGLCE
 PID64649.1 WL----Q----KDST--TCVASF---SN-----VKAENETQVKQ-----
 MVQRHIAPKLNSVVSAAAG-----PSYDITLHFDGPFVVNDSDKC-----KAEDTPDIYPL---
 EEKNG-----VP---AFPVRSFRGAIRSQAERIIRTIGGQ
 WP_031436019.1 WL----NSFSE-NDKK--MAMESA--RN-----LEQQEINTLIE-----
 QGNALFKA-----SCDE-----VKLGLTLKFKGPFLVNDPYAVKILS---SNENAKTDHYPL---
 LDKNR-----NP---YLPVSSFRGVLSQAERIIRTGLGK
 HEB50754.1 WL----ASFEGGNGG--MAMTAA--AL-----LQADTVQRRAD-----
 KVRQAWQP-----PDVG-----PRLHVELRFSGPFLVNDPSRN---T---PDITQAPDMVPL---
 VDEDG-----NP---MLPASSFRGALRAQAERIIRTGLGR
 HrcCas7-11 FS----LYLRA--RGL-KGLSR---EE-----VTRIGLNEEQWEAVMA---DD-----
 PGTHYNPF-----P--W-----EKISYTLIHSPLISNDPIAAM--L---D--
 HDNKDAVMVQKTVLFVDESIGN---YSQMPHHFLKSGSIRGACRFLGRK-----
 CmaCas7-11 YL----QYLFS--RGY-KGIE---TD-----EIKKVADPIKWKTFL---TK-----
 LEIPPEKI---P--L-----TQLNYTLTIDSPILSRDPIAAM--L---D--
 NRNPDAVMVKKITLVYEQDSSSTHKNPKEVPKYFIKSETIRGLLRISIISRT-----
 HvmCas7-11 HE----AYLLN---RGF-KGQMPV---QD-----VKTKSFKTKTWFIQH---RE---
 -LDISPKKL---P--W-----YSTDYRFNVTSPLISRDPIGAM--L---D--
 PRNTDAIMVRKTVFCPDPA---KNRPAPATVYMIKGESIRGILRSIVVR-----
 SmCas7x3 RQ----AYSSEN---NGL-RSPEEFDLAV-----IHELAEGLAKEDGQK-----
 ILKGTEPF-----TC-W-----QERSWQFSFTGPLLQGDPLAAL--N-----
 SDTADIISFRRTVVDNGEVL-----REP---VLRGEGLRGLLRTAVGRV-----
 oral_meta RA----DYVQK---CGLRRGVGDDTAINL-----EKDLSLNLPE-----
 SGY-----P--W-----KKHAWKLSFQVPLLTADPIMAQ--T-----
 RHEEDSVYFQKRIFTSRGRVV---LVP---ALRGEGLRGLLRTAVSRA-----

DisCas7-11 RN----DYLKN---WGW-RDEKGL---EE-----LKKRLNSGLP-----
 EPGNYRDP-----K--W-----HEINVSIMASPFINGDPIRAA--V---D--
 KRGTDVVTFVKYKAEGEEAK-----PVC---AYKAESFRGVIRSAVARI-----
 DsbaCas7-11 RE----SYLSN--HGW-RGIMEN---SI-----VHESPLEGGAGGCSF-----
 GLSDLPKL-----G--WHAEDLKLSDIEKYKPFHRQKISVKITLNSPFLNGDPVRAL--T-----
 EDVADIVSFKKYTQGGEK-----IIY---AYKSESFRGVVRTALGLR-----
 SstCas7-11 RK----DYLDN---YGW-RNRIEA---LS-----LEKMPLD-----
 RMNDYAEF-----L--W-----QKVSVEIEIGSPFLNGDPIRAL--I---E--
 KDGSDIVSFRKYADDSGK-----EVY---AYKAESFRGVVRAALARQ-----
 CsbCas7-11 IN----EYIKE--RGM-RGKEKELLEMG-----ESSLPDGLIPYKFFEEREC-LF--
 --PYKENLKP-----Q--W-----SEVQYTIEVGSPLLTADTISAL--T---E--
 PGNRDAIAYKKRVYNDGNA-----IEPEPRFAVKSETHRGIFRTAVGRR-----
 CjcCas7-11 LK-----AYICS---RGL-
 RGIEKEVLLENKTIAEITNLFKTEEVKFFESYSKHIKQLCHECIINQISFLWGLRSYYEYLG-----L--W-
 -----TEVKYEIKIASPLLSSDTISAL--L---N--KDNIDCIAYEKRK WENGGIK-----FVP---
 TIKGETIRGIVRMAVGKR-----
 CbfCas7-11 LH----AYIDT---YGH-RGNEHAIGTGQ-----
 GIDGFRSGSLSDLLSDISKESFRDPLASYHNYLDK-----R--W-----
 IKVGYYQITIGAPLLSADPIGAL--L---D--PNNVDAIVFEKMKLDGDQVK-----YLP---
 AIKGETIRGIVRTALGKR-----
 HvsCas7-11 LF-----NYIKN---HGF-RGEERLFLEGN-----EAELEKMGIIQINTELLQPEMIK--
 --KEKNFTDF-----P--Y-----DLIKYQLNISSPLLLNDPIRAI--ALYEGE--
 GKAPDAVFFKKYVFENGKIE-----EKP---CFKAESIRGIFRTAVGRI-----
 HsmCas7-11 FH----SYLQY--KGL-RDKGDFNSIKE-----IEGLKVETEEFKVK-----
 --KPF-----P--W-----SCVEYTIFIESPFVSGDPVEAV--L---D--SSNTDLVTFKKYKLEESK--
 --EVF---AIKGESIRGVFRTAVGKN-----
 FmCas7-11 QN----DYIQY---HGF-RNKEDEL--LS-----VYSTVSGLAEKNDV-----
 NLNNARDF-----S--F-----DKICWEVQFDGPVLTGDPLAAL--F-----
 HGNTDSVFYKKPILKSGEKE-----PSY--QWAIKSDTVRGLIRSAFGKR-----
 DpbaCas7-11 KNVWVIGNYLKS---QGW-RG-----AE-----LETHFDSLKEWKS-----SF---
 -SDSDVKVF-----T--W-----HKITWKVSFEGPVL TNDPIAAD--I-----RNESDAVFYQKSV--
 AGEKG-----PVY---ALKGEGLRGIVSSSLCKK-----

Domains**SECTION 6**

*

WmCas7x3 VAQG-----HAVPPVKNNTCP-DPATL-----LFGTAGWRGLLRTDD-----
 -----CTGTA-----PAT-----
 --LVD
 SER16298.1 IPQG-----HAAPAYDGIHR-DLISL-----
 LFGTAGWKIVCASDLIHSIPEYALQFNGVRETISDLSDTVKSCIIVDLVKTSTAAEKTEEQLHIRIVDS
 AGSLIVHKSENSSWANDTFRDASVKDNFKARLKEIADPQDLSALRADIKKRAFQLAT-----
 -----LTR
 GwCas7x3 TPQG-----HTAPAYRKGPDDLA VL-----LFGAAGWRGIVQTS-----
 -----CIVED-----KSIK-----
 -TRR
 PID64649.1 CCDG-----SINNTCK-----NPKNLCIACEMFGSTGWKTSIEMDPFL-----
 ---CVDRE-----LKP-----
 FII
 WP_031436019.1 CCS-----TDDPCKPIFDKGDLSKLCLACQIFGASGWKTVINIHDFK-----
 -----AINKS-----K-----
 --KTK
 HEB50754.1 CCD-----TSSPCRPLGSSDKVGELCLACQVFGAPGWGTTLHIQGFT-----
 -----CTSV-----FRR-----
 --EQE
 HreCas7-11 -DAPNE-----NGLTYFE--ADHEE--CDCLLCSLFGSKHYQGKLRFEDAE-----
 -----LQDEV-----EA-----
 ----IK

CmaCas7-11	-EIKLE----	DGKKERIFN--LDHED--	CDCLQCRLFGNVHQQGILRFEDAE-----
-----	ITNKN-----	-----	VS-----
---DCC			
HvmCas7-11	-----	NEELYD--TDHED--	CDCILCRLFGSIHQGSLRFEDAE-----
--VQNSV-----	-----	-----	SD-----
KK			
SmCas7x3	-----	AGDDLLT--RSHQD--	CKCEICQLFGSEHRAGILRFEDLP-----
-PVSP-----	-----	-----	TV-----
ADKR			
oral_meta	-----	YGISLIN--DEHED--	CDCPLCKIFGNEHHAGMLRFDDMV-----
-PVGW-----	-----	-----	ND-----
KK			
DisCas7-11	-HME-----	DGVPLTE--LTHSD--	CECLLCQIFGSEYEAGKIRFEDLV-----
---FESDP-----	-----	-----	EP-----
VT			
DsbaCas7-11	-NQGNDDIT--GKKNVPLIA--	LTHQD--	CECMCRRFFGSEYEAGRLYFEDLT-----
-----	FESEP-----	-----	EP-----
-----RR			
SstCas7-11	-HFDKEGKPLDKEGKPLLT--	LIHQD--	CECLICRLFGSEHETGRLRFEDLL-----
-----	FDPQP-----	-----	EP-----
---MI			
CsbCas7-11	-----	TGDLG-K--EDHED--	CTCDMCIIFGNEHESSKIRFEDLE-----
LINGN-----	-----	-----	EF-----
EKLEKH			
CjcCas7-11	-----	SGDLG-M--DDHED--	CSCTLCTIFGNEHEAGKLRFEDLE-----
-VVEEK-----	-----	-----	-----
LPSEQNSDSNKIPFGPVQDGDGNREKECVTAVKSYKKKL			
CbfCas7-11	-----	NNLLA-K--NDHDD--	CTCSLCAIFGNEHETGKIRFEDLE-----
-VYDKD-----	-----	-----	IA-----
KK			
HvsCas7-11	-----	KNVLT---KNHED--	CICVLCHLFGNVHETGRLKFEDLK-----
--IVSGQ-----	-----	-----	EE-----KF
HsmCas7-11	-----	EGKLTTE--NEHED--	CTCILCRLFGNEHETGKVRFEDLE-----
---LINDS-----	-----	-----	AP-----
KR			
FmCas7-11	-----	DALLI---KSHED--	CDCLLCEAFGSKHHEGKLRFEDLT-----
PKSDE-----	-----	-----	IK-----TYR
DpbaCas7-11	-----	KNLSS---NLHED--	CECLRCKIFGSKHQEGNIRFEDMT-----
-VSQES-----	-----	-----	EV-----
REKL			
Domains			
.....			
.....			

SECTION 7

	D745	D758	G-rich loop
WmCas7x3	HDMLAIDRFTGGGKDGAKFKLRYA----	ECPTLEGQLSLDLSRLRSARLD-----	
GANAADTPWIALGLLTLVLRDLAEGDIPFGHGSAGYGRCAQGLPDRWRQALEAHFGPNADARA			
LAALRAWCRTHATAAL--DAPCSLAGSAPTAAAAPS--GQAAPAD--AFHN--PYHFI-P-F-			
SQPDIDRWLSPDA			
SER16298.1	HEMVAIDRFTGGGKEGAKFNVDYI----	ECPTLTGAIYLDLHRLKQAQLK-----	
NDEDALKP--			
ALGLISLLLRDLAEGDIAFGFGANKGYGQCREHAVLDNWEERLKKIGAGLTIDGALQALRDTVALEP			
PAEF--PPEIEKTTDDNQPEAPDF--NLKPASN---GFHN--PYHFI-P-L-NNPKIGDWPEAKA			
GwCas7x3	HEMLAIDRFTGGGKDGAKFNVDYV----	ECPTLAGKLSLDLARLKNALK-----	
GKGDALLP--			
ALGLMTLMLRDLAEGDIPFGYGISKGYGQCRASSALGDWAELLKQHLGADSADTTVQALREYLG			
PKGQELKLDPPSADATQAGVPAQQNAAKTQAQGAQE--KFHN--PYHFI-P-L-SKPDISQWPEPQK			
PID64649.1	QEFVAIDRFHGGGKDEAKFNAAHY-----	QAPVFKGKVRVS-QRVGN-DIS-----	WRK--
-----GLLALIFRDLKEGDIYFCFGTNKGYGAVKKAEPDGN-----			

ASDFSESDIEAFINKCREKKGLYNC--NPIKKPGKTKVSKNLPPAIVPLDRTDS---KFYN--PYHFI-P-V-KKPNTSSWAEKTA	
WP_031436019.1 QDFVAIDRFHGGGKDGAKFDATHF----ERPEFEGAISFS-PRMANNDLD-----WGK-----GLLALVLRDMQEGDMTFGYGANKGYGGLESASITGIEQ-----	
ITSDIQAFRDCKVASPQTWLCDEAVKPANQQD----KIPPA--GIQVANS---GFHN--PYQFI-P-S-KEPDTGHWLPVLG	
HEB50754.1 QTFVAIDRFHGGCKEGALYTIRHA----ESPRFEGHLVID-PRMPA-----WGR-----GLLALVFRDLREGDITFGLGAGKGYGVVDAAVVQDMAE-----	
LEPYVEAFREQCRQHQMADCHSAPSPQPLRDHDLAEIPPA---EEAPGE---TFLN--PYHFV-P-I-REPDTGSWLARDE	
HreCas7-11 CDHVAIDRFHGGTVHRMKYDDYPLPGSPNRPLRIKGNIVVK-----RDLN-----DTEK-----EAVKDVLTELRLDGLIPLGANGGAGYGRIQRLMI-DDGPGW-----	
LALPERKEDERPQPSFSPVS-----LGP-VHVNLSKSGSDTADVYYHPHYFL-E-PPSQTV-----CmaCas7-11 IDHVAIDRFTGGGVEKMKFNDYPLSASPKNCLNLKGSIWIT-----SALK-----DSEK-----EALSKALSELKYGYSALGGLSAIGYGRVKELTLEEN-----	
DIIQLTEITESNLNSQSRLS-----LKP-DVKKELSN--HFYY--PHYFI-K-PAPKEV-----HvmCas7-11 MDHVAIDRFTGGGVDQMKFDDYPLPGCPAQPLILEGKFWVK-----DDID-----DESK-----SALEKAFADFRDGLVSLGGLGAIGYGQIGDFELIGGSADW-----	
LNLKPPEENRTDVPCGDRSA-----QGP-EIKISLDAD--KIYH--PHFFL-K-PSDKNV-----SmCas7x3 LDHVAIDRFDQSVVE--KYDDRPLVGSPPKQPLVFKGCFWVQ-----TSGMT---HQLT-----ELLAQAWRDIAAGHYPVGGKGGIGYGWINSLVVDGEK-----	
ITCRPDGDSISLTTVTGDIP-----PRP-ALTPPAG----AIYY--PHYFLPPNP-EHKP-----oral meta IDHVSCSRFDASVVN--KFDDRSLVSPDSPLFHFGTFWLH-----RDFQ-----ND-----VEIKTALQDFADGLYSIGGKGGIGYGWLFDMEIPLSLRKLNSGFREA-----	
SSIQDALLDSAKEIPLSAPLT-----FTP-VKG-----AVYN--PYYYL-PFP-AEKP-----DisCas7-11 FDHVAIDRFTGGAADKKKFDDSLPGSPARPLMLKGSFWIR-----RDVLED---EEYC-----KALGKALADVNNGLYPLGGKSAIGYGQVKSILGKGGDKRI-----	
SRLMNPADFETDVAVPEKPK-----TDA-EVRIEAE----KVYY--PHYFV-E-P-HKKV-----DsbCas7-11 FDHVAIDRFTGGAVNQKKFDDSLVPGKEGFMTLIGCFWMR-----KDKELS---RNEI-----EELGKAFADIRDLGLYPLGAKGSMGYGQVAELSIVDDEDSDDENNPAP-----LLAESMKNASPSLGTPTSLKK-----KDA-GLSLRFDEN--ADYY--PYYFL-E-P-EKSV-----SstCas7-11 FDHVAIDRFTGGAVDKKKFDDCSLPGTGPGLTLKGCFWIR-----KELEKPEDKSER-----EALSKALADIHNGLYPLGGKGAIGYGQVMNLKIKGAGD-----VIKAALQSESSSRMSASEPEH-----KKP-DSGLKSLFDDKKAIVY--PHYFL-K-PAEEV-----CsbCas7-11 IDHVAIDRFTGGALDKAKFDITYPLAGSPKKPLKLKGRFWIK-----KGFS-----GDHK-----LLITTALSDIRDGLYPLGSKGGVGYGWVAGISIDNVPDDFKEMINK-----	
TEMPLPEEVEESNNGPINNDY-----VHP-GHQSPKQDHKNKNIYY--PHYFL-D-S-GSKV-----CjcCas7-11 IDHVAIDRFHGGAEKMKFNTLPLAGSFEKPIILKGRFWIK-----KDIV---KDYK-----KKIEDAMVDIRDGLYPIGGKTGIGYGWVTDLTILNPQSGFQ-----	
IPVKKDISPEPGTYSTYPSH-----STP-SLNKG-----HIYY--PHYFL-A-P-ANTV-----CbfCas7-11 IDHVAIDRFTGGARDQMKFDTLPLIGSPERPLRLKGLFWMR-----RDVS---PDEK-----ARILLAFLEIREGLYPIGGKTGSGYGWVSDLEFDGDAPFAFKEMNSK-----	
RKGQASFEKEISFRYPSGAPK-----HIQ-NLKAT-----SFYY--PHYFL-E-P-GSKV-----HvsCas7-11 FDHVAIDRFLGGAKEKYKFDDKPIIGAPDTPIVLEGKIWVK-----KDIN---DEAK-----ETLSQAFSDINTGIYYLGANGSIGYGWIEEVKA-LKAPSWLKI-----EKNFEKDTSLNISAIMNEF-----KKD-IQTLNLD----KTYL--PYGFL-K-L-LEKV-----HsmCas7-11 LDHVAIDRFTGGAKEQAKFDDSPGSPSPLEFTGIVWVR-----DDID---EEEK-----KALKSAFLDIKSGYYPLGGKKGVGYGWVSNLKI-ESGPEWLRLEVQE-----KSSQENVLSPVILSEVMDIEF-----NPP-KIDEN-----GVYF--PYAFL-R-P-LNEV-----FmCas7-11 MDHVAIDRISGGAVDQCKYDDEPLVGTSKHPLVFKGMFWIN-----RDSS-----VEMQ-----RALIAAFKEIRDGLYPLGSGNGGTGYGWISHLAI-TNGPDWLNLEEV-----LPQPTADIPVEECTAEPYPKF-----QKP-DLDQN-----AVYY--PHYFL-Q-P-GKPA-----DpbaCas7-11 FDHVSIDRFTGGAANKLKFDKPLVG--NPLVFQGVFWVH-----QSIGNN---EKTQ-----EALSDAFKDVRDGLYPVGAKGSIGYGWIKGIEV-VEGPDWLK-----DALSAEKTVEAGIASEEY-----KLP-DLPWISLLPKGRAIYN--PHYFL-GIP-KVTP-----	
Domains	
.....	
.....	
SECTION 8	
E959	

WmCas7x3 H--RKTGGH-SRYRG-----LSGRLVCALTTVTPLFVGAAA-RTP-ASDQHPK---
 -----P---VAGFALQNQPAIPATSLRGLLSSLFESISGSNLRVLHPT-PYSIRKTTKE-----ALSA--I-
 -----G-----
 SER16298.1 ETLKANREGH-DQYHTGK-----FSGRIVCSLTTQTPLFIGAET-KPS-TSDREPS---
 -----E---ARPFKLNGKHAIPATSLRGMLSSLFESVSNSNFRVLHPE-HYSVRKSLDDYV-----
 ALSA--M-----G-----
GwCas7x3 L--TEKGHSH-DRYAS-----LSGRIVCRLTTQTPLFIGSEQ-TTP-TNPQAPK-----
 ---S---LHPFKLNNGLAIPATSLRGMISSLFESVSNSNFRVLDEK-TYSMRKTMQQ-----SLSA--
 M-----G-----
 PID64649.1 FGTADSPHSH-GFYRKQTNE-----QQPLYSGRLICMLTSETPFFIGAQA-ESD---
 PTENE-----NQASLRHPYQLDGEPAIPSTSLRGLISTMTEAAANCAMRVLDSE-IISYRKPMNPSH---
 -----ILSA--L-----G-----
 WP_031436019.1 L-NADSHHSH-AFYRDQTDN-----GEKLYHGRILCCLNTETPIFIGADK-KKD-
 TEPAEIN-----N-----YRLNGELAIPATSLRGMISSLAEAAASNSAMRVLDNG-LLSYRKTTADD-----
 -----ALRK--V-----G-----
 HEB50754.1 L-DSSCCHSH-GFYRQQVDD-----RPLYHGRILTCLLETETPLFIGATG-DSS-V-
 PSRIE-----N-----YRLGNRIAIPAASLRGMLSSLAEAAASNSAMRVLHQG-ILSYRKKAHN-----
 --ALRE--I-----G-----
HrcCas7-11 S-RELDIISHARTRDS---G-----GEALLTGRILCRLITRGPIFIPDTN-NDNAFGLEGGI-G-
 ----HKN---YRFFRINDELAIPGSELRGMVSSVYEALTNCFRIMEEGRYLSRRMGADFEK-----
 -DFHP-----G-----
CmaCas7-11 V-RESRLISHVQGHDT--E-----
 GEFLTGTGIKRLQTLGPLFIANNNDKGDDYFELQHNNPG-----HLN---
 YAFFRINDHIAIPGASIRGMISSVFETLTHSCFRVMDDKKYLTRRVIPESETTQKRKSGRYQVEESDPD
 LFP-----G-----
HvmCas7-11 Y-RERELVSHAKKKGP--D-----
 GKSLFTGKITCRLSTEGPVFIPDTDLGEDYFEMQASHKK-----HKN---
 YGFFRINGNVAIPGSSIRGMISSVFEALTNCFRVFDQERYLSRSEKPDPT-----LTKYYP-----
 -----G-----
SmCas7x3 K-RSDKIIGH-HTFAT--D-----PDSFTGRITCKLEVVTPLIVPDTE-G-----EQPKDQ---
 --HKN---FPFFKINDEIMLPAGPLWAAVSQVYEALTNCFRVMMKQKRFLSWRMEADYK-----
 -DFYP-----G-----
oral_meta E-RCLVPPSH-ARLQS--D-----RYTGCLTCELETVSPLLLPDTC-----REKDGN-----
 YKE---YPSFRLNNTPMIPGAGLRAAVSQVYEVLTNSCIRIMDQGQTLSSWRMSTSEHK-----
 DYQP-----G-----
DisCas7-11 E-REEKPCGH-QKFHE--G-----RLTGKIRCKLITKTPLIVPDTS-
 NDDFFRPADKEARKEKDEYHKS---
 YAFFRLHKQIMIPGSELRGMVSSVYETVTNSCFRIFDETKRLSWRMDADHQN-----VLQDFLP---
 -----G-----
DsbaCas7-11 H-RDPVPPGHEEAFRG--G-----LLTGRITCRLTVRTPLIVPNT-
 TDDAFNMKEKAGK-KKDAYHKS---
 YRFFTLNRVPMIPGSEIRGMISSVFEALSNSCFRIFDEKYRLSWRMDADVKE-----LEQFKP-----
 -----G-----
SstCas7-11 N-RKPIPTGH-ETLNS--G-----LLTGKIRCRLTTRTPLIVPDTS-NDDFF--QTGVEG---
 --HES---YAFFSVNGDIMLPGSEIRGMLSSVYEALTNCFRVFDEGYRLSWRMEADRN-----
 LMQFKP-----G-----
CsbCas7-11 Y-REKDIITH-EEFTE--E-----LLSGKINCKLETLTPLIIPDTS-DENGLKLQGNKPG--
 ---HKN---YKFFNINELMIPGSELRGMLRTHFEALTCKSCFAIFGEDSTLSWRMNADEKD---
 YKIDNSIRKMEQSRNPKYRIPDELQKELRNSGNLGNLYTSERRFWSDVSNKFENSIDYKREILR
CjcCas7-11 H-REQEMIGH-EQFHK--E---QKGELLVSGKIVCTLTVTPLIIPDTE-
 NEDAFGLQNTYSG---HKN---
 YQFFHINDEIMVPGSEIRGMISSVYEAITNSCFRVYDETKYITRRLSPEKKD--
 ESNDKNKSQDDASQKIRK-----G-----
CbfCas7-11 I-REQKMIGH-EQYEE--SYPSGASGEKLLSGRIICSMITHTPLIVPDTE-----
 VIKDPENK-----HAT---YDFFQMNNAIMIPGSEIRGMISAVYEAMTNSCFRIFHEKQYLTRRISPEDKE-
 -----LREFIP-----G-----
HvsCas7-11 K-RTSSPITH-ERFYE--N-----HLTGFIKSLKVLSPILIPDTE-----TPEKEENG-----
 HKY---YHFLKIDNKPIPGAIRGAVSSIYEALTNCFRVFGEKKVLSWRMEGKDAK-----
 EFMP-----G-----

HsmCas7-11 K-RTREPIGH-NEWKK---S-----LISGYLTCRLELLTPLIIPDTS--EEVIKEKVNNGE-
 ----HPV---YKFFRLGGHLCIPAAEIRGMISVVYEALTNSCFRVFDEKRLISWRMTAEEAK---
 RPDPPKSEEQNRMRFRP-----G-----
 FmCas7-11 E-RERHPVSH-DHIDD--K-----LLTGRLVCTLTTKTPLIIPDTQ-
 TNTMLPPNDAPEG-----HKS----
 FRFFRIDDEVLPGEIRGMVSTVFEALTGSCFRVINQKAHLSWRINADMAK-----HYRP-----
 -----G-----
 DpbaCas7-11 E-REREPVGH-DRFQT--D-----LHTGRIICLTKTITPLIIPDTE-
 NDKAFEVENASAD-----HER----
 FKFMRMGSQAAPGSAIRSMTSSVFEALTNSCFRVLQKSHLSWRMEADDAG-----DYKP---
 -----G-----

Domains

.....

SECTION 9**D998**

WmCas7x3 -----RIVERNGELKLYPL-----TLPTIHQNA-DNAY-----PVPA-----
 RWRKV-FYWESPV-----

 SER16298.1 -----RIVDDQGEKLLQPL-----TLPTLFGNR---N-----NVPA-----
 KWEKIFGTPSEDD-----
 --
 GwCas7x3 -----RIVRHDQKLYLLPL-----TLPTLPQGP-HGVY-----DLGE-----
 KWSAVF-DWQP-P-----
 --
 PID64649.1 -----MVTKRGEDEFWLPL-----AMPALSLNDEEHNY-----KLDK-----
 -RYRTMF--PDGLA-----
 --
 WP_031436019.1 -----MVIYVDNKSFIKL-----
 NDAI-----
 HEB50754.1 -----MIVLRDGKRFLPL-----VPLM-----
 EVT-----
 HreCas7-11 -----IV--VDGAKIREM-----KRYRLPLYD-T--P-----DKTS-----
 RTKEMT-CPELFTRKDG-----

 CmaCas7-11 -----RVQKKGNKYKIEKM-----DEIVRLPIYD-N--F-----SLVE-----
 RIREYH-YSEECASYVP-----

 HvmCas7-11 -----KVCRDGNKFFILKM-----KDFRRLPLYD-F-----DFEG-----
 EAESLR-PNYDEDRNEEENKGNKNTQ-----

 SmCas7x3 -----RV--LDGGKQIKKMG-----DKAIRMPLOYD-D--S-----TATG-----
 SIKDDQ-LISDCCPKSDE-----

 oral_meta -----KI--TDNGRKIQPMG-----KQAIRLPLYD-E--
 VIHVVSTPGDTDDLEKLKAIVLELTRPWKELP-
 EEQKKRFEKCKNILDGRMLQKELRALENSGFAYWRDKTSLTFDSFLKDAIEQEYPRYSGDYQRI
 KALVVNITLPWKLLKKEERHKRFDKCRRLKGGQPLTKDERKALEESG
 DisCas7-11 -----RV--TADGKHQKF-----SETARVPFYD-K--TQKHFDILDEQEIAGE----
 ----KPVRMW-VKRFIKRLSLVDPKHPQKKQDNKWKRRKEGI-----

 DsbaCas7-11 -----RV--ADDGKRIEEM-----KEIRYPFYD-R--T-----YPER-----
 NAQNGY-F-RWDARISLTDNSMRKM-EKDG-VPRNVYKLNLTKNKAYKSEKSFLFDL-----
 -----KNKAGGV
 SstCas7-11 -----RV--TDNGLRIEEM-----KEYRYPFYD-R--D-----CSDK-----
 KSQEAY-FDEWERSITLTDLSLEKMAERKGDISPDLKVLKSLKGKNYKSTEGLLAAF-----
 -----KDKGGDT
 CsbCas7-11 CAGRPNYKGGIIRQRKDSLMAEEL-----KVHRLPLYD-N--F-----DIPD---
 -----SAYK-ANDHCRKSATCSTSRGCRERFT-----

CjcCas7-11	-----LVKKTDEGFSIIEVERYSMKTKGGTKLVKVVYRLPLYD-S-E-----
AVIA-----	SIQ-FEQYGEKNEKRNA-----
CbfCas7-11	-----IVRIINGDVYIEKA-----EREYRLPLYD-D-V-----HIIT-----
NYEELE-YEKYIKKNPGREQ-----	
HvsCas7-11	-----RVSKKKGKLYMVKM-----QALRLPVYD-N-P-----ALAN-----
-EIRSGS-IYE-----	
HsmCas7-11	-----RIKKDKKFYAQEM-----LELRIPVYD-N-K-----D-----
KRNEIS-QNDPTRPS-----	
--	
FmCas7-11	-----RIIQNNEKMFIQPY-----KMFRLPFYA-G-F-----
DpbaCas7-11	-----RFEKKDDKAVIRKF-----KKKARFPFYA-G-P-----
Domains	
SECTION 10	
WmCas7x3	-----PLRVY---FGS-R-----KQTYDS-R---QP---HYLP-IQELS-Y-----
---LPND---SDCI-----	APDQGDRLRFPDRDRKFLIGQCPISRYDCPIPETDLP-----KLSPQERPRY-----
TR--GWVR--SLWTS-----	NREK---ELPHTVK-H
SER16298.1	-----FLRIY---FDDIP-----SKFSSN-K---RY---FYNCKATELKDF-----
---IKSD---KYFIG---KRTPTV---	FPKSSTEKSHLESLEFID--VEKFKKAVEN-----LEITPGNNPY---IH-
-GWVR--NLKDE---FRE---	DIPDNVK-H
GwCas7x3	-----PLRIY---FDPPP-----RRTYQS-Q---QP---CYMK-LSTVK-Y-----
---SESN---PNQI---	IAGENLGALRFPRGNQNTQFLIGQSNQD--ECPITQAEYA-----QKSEDERNEY---
TS--GWVR--TLVKP-----	GR-----DLPRSVK-H
PID64649.1	-----KLKVYLEKAYSNNV-----MKTFLN-N---EN---TWTLAQSKIH-Y---
-----LPLT---PIQMQ---	NGGINSYNNLRTPSRSNFLIGQTVAH--GNGIPASGPG-----AGMVP---
-----GILR--ILGKE---	HRQN---DLPQNK-K-H
WP_031436019.1	-----KLKQT---YTPGN-----MKDFIE-K---SN---SWSPEHNTVY-Y---
-----LD-----	NNQIPQESYM-----NGMKP-----GILR--ILGKE---
GREQ---EL-ENKL-H	
HEB50754.1	-----KLRHA---YTDPA-----MKHFLD-D---KN---SWSPRCNRVY-Y---
-----LGRD-----	GNQIPAETRG-----AGMRP-----GILR--LLGRE---
-GRHD---AL-QNKK-H	
HreCas7-11	-----RPERA---KKFNEEIAKVAVQNRAYLL-SLDEKERR---
EVLLGNREVT-FDECPDDEYSDEYSELKYAQKYKDFIAVLKKNQKRGYIKFTGPNTAN-----	
-----KKN-----	EDAPDKNY---RS--DWDP--FKLNI-----LLESDEPCRVSN--IH
CmaCas7-11	-----SVKKA---IDYNRMLAQAAADSREFLY-NHPEAK---
SILQGKKEVY-Y-ILHKQESKNRGKTKEINPNARYACLTENTPGSRKGFIKFTGPDMMVT-----	
--VNK-----	ELKSKIPIYDPEWEKDIPDWERSNQESNHKYSFILHNEIEMRSSQKK---
HvmCas7-11	-----KVKNA---VEFNIKMAGFAKHNRDFLK-KYKEQEIK---
DIFMGKKKVY-F-----	TAGKHKPNEAHDNDKIALLTGGSNKAEEKGYFKFTGPGMVN-----
-VKA-----	GVEGEECDFHIDESDPDV-YWNMS-----SILPHNQIKWRPSQKK---
SmCas7x3	-----KLQKA---LATNQKIALAAKHNQEYLA-QLSPDERE---
EALQGLKKVS-F-----	WTESLANNEAPPFLIAKLGEERGKPKRAGYLKITGPNNAN-----
IAN-----	TNNPDDGGY---IP--SWKD--QFDYS---FRLGPPRCLPNTKGNR
oral_meta	FANWHGRELLFDRFLKDENSCLIA---ETTDRIASVAKNNRDYLF-
EIKQQDFARYKRIIQGLERVP-F-----	SLRSLAKSKETSFQIACLGLRRGRFLRKGYLKISGPNNAN-----
-----VEI-----	SGGSHSNSGY---SD--IWDD--PLDFS---FRLSGKSELRPNTQKTR
DisCas7-11	-----ATFIE---QKNGSYFNVVTNNGCTSF-
HLWHKPDNFDQEKLEGIQNGE-K----	LDCWVRDSRYQKAFQEIPENDP--
DGWECKEGYLHVVGPSKVE-----	FSDKKGDVINNFQGTLPVS---PN--DWKT--IRTND-----
-----FKNRKRKNE--	
DsbaCas7-11	-GRYKKLVLKHAIEVRGGEIPYYSHP---TPTDCKLLSLVGPNRQLCR-
QDTLVQYRIKHRRGAKPEED-F-----	MFVGTPSENQK---GHKEN---NDHGGGYLKISGPNKIE-----
-----KEN-----	VLTSGVPSV---PE--NMGA--VVHNCPP---RLVEVTVRCGRKQEEEC

SstCas7-11 GGNILGLIFKYAE-RIGDVPRYEHF---TDTDRMMLSLSEYNRNQ-K-
SDGKRAYKIIKPASKLGKAY-F-----MFAGTSVENKRICNPACTDKA---NKS VKGYLKISGPNKLE---
-----KYN-----ISEPELDGV---PE--DRNCQIIHNRI---YLRKIFVANAKKR---K
CsbCas7-11 -----CGIKV---RDKNRVFLNAANNRQYLN-
NIKKSNDLYLQYLKGEKKIR-F-----NSK VITGSESPIDVIAEL--NERGRQTGFIKLSGLNNSN-----
-----KSQ-----GNTGTTF---NS--GWDR--FELNI---LLDDLETRPSKSD---
CjcCas7-11 -----KIRAA---IKRNEVIAEVARKNLIFLR-SLTPEELK---KVLQGEILVK-F-
-----SLKSGKNPNDYLAELH--ENGTERGLIKFTGLNMVN-----IKN-----V--NEEDKDF---
--ND--TWDW--EKLNI---FHNAHEKRNSLKQ---
CbfCas7-11 -----KIKNA---HRFNKNIARIAESNRNYLC-SLDRAVRR---EILSGRKKVN-
F-----RLVKVNDNKNPDKEAVELC--KTGPLEGLVKFSGLNAV-----ISN-----
LRPGTAEEGF---DA--KWDM--WSLNI---ILNRMDVRNSQKK---
HvsCas7-11 -----KYKN---SKVEIIFQTVEGIRKFLRGFNFNVEWKK--
VLVTGIDPLA-I-----LPSQKIPGNDKWVKNLQ--SKISPVRGYFKFTGPNKIE-----
TKRREEEKDEKLRTKANKVSLQ-KD--KWYE--AMHNHVE---YKQDYTPPNPKTEP-L
HsmCas7-11 -----EYNHP---TEPERIFFSNAEKIRNFLK-RNS-----NYLHGSTPLL-F---
--RQWSISNRYDKIA-----LIGNKSQGHKFTGPNKIE-----VSE-----
GTKCPKYETIPGRD--EWDK--AVHNYVEPGKFVTVISRKKGQKPKA-V
FmCas7-11 -----DPRNCLSE-----KQLLGIEPVK-L-----
WVKDFVASLVKPQTDIDIE---WKEKIGFVRVTGPNKVE-----VDS-----SNTPDPSLPEC-----
ES--DWKD--IHTE-----DGSTPSKNDR---
DpbaCas7-11 -----DTREAFS-----DQIMGKEKVT-L-----
WVKDFEASLTPDE-----IGWKKKRGYLVKVTGPNKVE-----IDT-----ENISENNPSP-----PD-
-SWQD--VRIND-----DGTIPDKKNR---
Domains
.....
.....

SECTION 11

WmCas7x3 QLFI--PDPVE--TPAA--DDL-----LPIP-QGV-
LDTFHALADLALAGQH---W-GKDE-TPADD-QLLPFTPAGRQR---HDA---
DRPPRDADRQTRLQPGDLVCFD--LG-----D-----DGAVSEISFSSIWREGLRLAG-KPN-
LATTADLLAQ-----V
SER16298.1 HVFL--PDTTKRVSP-----LEIP-PHV-KKRFHELADLALAGLH---
-L-KQGE-TIASPYKILPYTPIGRKN---LENHIHRVPNDLTCYMTLRLKKGDLVFFD--VD-----N-----
DGQITEISFSSIWRAGI--GTKNK-LQTTADLLSQ-----R
GwCas7x3 HVFL--PDVFIDAPPPV--NDL-----YPIP-DSV-
IQRFHDLADQVLASMN---L-KPEE-IVDST-NLLPYTPVGRRS---DSD-----CRDTRLQAGDLIFFD--
IDPPLHPG-E-----KSQITEISFSSIWRSGI--G-KDH-LLTTPDLLTN-----F
PID64649.1 ELFI--PVPDAFVADPK--TFLD-----TATAFLIP-RNV-
IDAFEDIAEKQTQSQKQDK-L-KHDE-----ERLPFHLKGTRR---EQN-----HTLQIKTGDLVYFR--
PN---A-K-----GDEVEEIAFSSIWRGKT-----SGTTADFFPD-----K
WP_031436019.1 ELFI--PVPLEYVDTEN--NKFDYQAYKK-----A-----FLYRAIEIP-EPV-
LKRYSELADQRTMSQKSNKEL-KKDD-TCQSV-GWLPFHLKGTKR---QLD-----
DKHNVGKLQIDEYDLIYYE--AS-----GKEVTEVAFSSIWRGRV---ETNSS-QANKVYSFIP-----
--G
HEB50754.1 EYFI--EIPERYVD-QD--HCFDYRMFIR-----D-----
RARNGTLVPISPAWERYHCLAEERTLSQKNDPEL-REDK-ACASL-KWLPFHPKGRVR---ERD-----
PENDVCHLSLRHGDLVFYA--EQ-----NRVVSEISFSIAWRSRV---ETSDSYQAVTVDCFP-----
---K
HreCas7-11 CYPR--PLLVC---IK--DKAEYRIHKRCEAIFCSI---GSPSDLYDIP-QKV-SNQYRTIL--
--QDYNDNTGK--IVE-IF-----RTQ---IKH-----DQ--LTTGDLVYFK--PA-----A-----
NGQVNAVIPVSISRKTD-----ENPLAKRFGN-----D
CmaCas7-11 KYPR--PVFIC---KK--NGVEYRMQKRCERIFDFT---KEEEKDKEIVIP-QKV-
VSQYNAIL---KDNKENTET--IPG-LF-----NSK---MVN-----KE--LEDGDLVYFK--YK-----
-----EGKVTELTPVAISRKTD-----NKPMGKRFPKISINGKMKPND
HvmCas7-11 EYPR--PVLKC---VK--DGTEYVMLKRSEHVFAEA-----SSEDSYPVP-GKV-
RKQFNSIS---RDNVQNTDH--LSS-MF-----QSR---RLH-----DE--LSHGDLVYFR--HD-----E-
-----KRKVTDIAYVRVSRTVD-----DRPMGKRFGN-----E

SmCas7x3 EYPR--PGFTC---VI--DGKEYSLTKRCERIFEDI---SGGENQVVRAVT-ERV-
 REQYREIL---ASYRANAAG--IAE-GF-----RTR---MYD-----TE-ELRENDLVYFKTAKQ----
 ADG-----KERVVAISPVCISREAD-----DRPLGKRLP-----A
 oral_meta EYPR--PSFTC---TV--DGKQYTVNKRCEVFEDS-----AAPAIELP-RMV-REGYKGIL-
 ---TDYEQNAKH--IPQ-GF-----QTR---FSS-----YR-ELNDGDLVYYK--TD-----S-----
 QGRVTDLAPVCLSRAD-----DRPLGKRLP-----E
 DisCas7-11 -----PVFCC---ED-DKGNYYTMAKYCETFFFDL-----KENEEYEIP-EKA-RIKYKELL--
 --RVYNNNPQA--VPESVF-----QSR--VARE-----NVEKLKSGDLVYFK--HN-----
 EKYVEDIVPVRISRTVD-----DRMIGKRMS-----A
 DsbaCas7-11 KRKRLVPEYVC---ADPEKKVTYTMTKRCERIFLEK-----SRRIIPFT-NDA-
 VDKFEILV---KEYRRNAEQQDTPE-AF-----QTI---LPE-----NG-TVNPGLDLYFR--EE-----
 ---KGKAAEIVPVRISRKVD-----DRHIGKRID-----P
 SstCas7-11 ERDRLVGEFAC---YDPEKKVTYSMTKRCERIFIKD-----RGRTLPIIT-HEA-
 SELFEILV---QEYRENAKRQDTPE-VF-----QTL---LPD-----NG-RLNPGDLVYFR--EE-----
 ---KGKTVEIIPVRISRKID-----DSPIGKRLR-----E
 CsbCas7-11 -YPR--PRLLF---TK--DQYEYNITKRCERVFEID---KGNKTGYPVD-DQI-
 KKNYEDIL---DSYDGIKDQ-EVAE-RF-----DTF---TRG-----SK--LKVGDLVYFH--ID-----
 G-----DNKIDSLIPVRISRKCA-----SKTLGGKLD-----K
 CjcCas7-11 GYPR--PVLKF---IK--DRVEYTIPKRCERIFCIP---VKNTIEYKVS-SKV-CKQYKVDL-
 ---SDYEKNFGH--INK-IF-----TTK---IQK-----RE--LTDGDLVYFI--PN-----EGA-----
 DKTQVQAIMPVPLSRITD-----SRTLGERLPH-----K
 CbfCas7-11 EYPR--PALHF---NH--DGKEYTIPKRCERVVRAEAGKRAETEGSYKVP-RKV-
 QEYQYNIL---RDYESNIGH--IDN-TF-----RTL---IEN-----CG--LNNGSLVYFK--PD-----N-S--
 ---RKEVVAITPVKISRKTD-----RLPQGDPRFPTS-----S
 HvsCas7-11 ERPRNIPCFVC---SD--KEKIYRMTKRCERVVSL---GENAPKYEIP-ISA-
 IKRYEVIL---SAYRENWERNKTPE-LF-----RTR---LPG-----DGRTLNEDDLVYFR--AD-----
 E-----NEKVKDIPVCISRIVD-----EVPLIKRLS-----Q
 HsmCas7-11 QRRRNVPFACC---YDYNTNRCFVMNKRCEVFVVS---RDKPKYEIP-PDA-
 IRRYEHVL---RKYRENWERYDIPE-VF-----RTR---LPG-----DGETLNEGDLVYFR--LD-----
 -E-----NNRVLDIIPVSRISD-----TQYLGRRLP-----D
 FmCas7-11 -----VYRC---QL--KGVTYTVAKWCEAFWVKD---EGKKPITVN-AEA-
 INRYHLIM---KSYQDN PQS--PPI-IF-----RSLPVLNYKQ---DQKII--GSMIFYR--ES-----
 AKS-----DKIVNEIIPVKISRAD-----TELLAKHLPN-----N
 DpbaCas7-11 -----KFIC---QY--GTTTYTVDKWCEAFFCDE-----EKDPYELA-PDV-ERKYRLLM-
 ---DSYHNNPQA--PPQ-IF-----RSLPLFSETG-----PKKTLEHGDLVYFRLSEV-----
 NKQSQSKKQVRERVTDIVPVSISRAN-----NQPIGKHIA-----A

Domains**SECTION 12**

WmCas7x3 SPHLLPLG-----MPGRS-A---RLSPVEQLFGVVEYRPPQTAK--
 GTRKPTDAPAAALAGKLQVGFRP--ARPF-----ER-----EP-AVTLKELSTPKPPSPALYFRPKA-
 -GDGYVSKAKLASQ-----PQ---DYAPAGRKHLYH--ALRR
 SER16298.1 DPNLVQLG-----MGVRT-
 KNTDRFKLSPAERLFGVVEHRDDNTTVENVNQPNDKAQAFAGKVRIGFGLP--DKKT-----
 TV----NGVS-PVTLKELSSPKPPSPAFYLKRKN--NDDFVSKKVAEC-----SE---
 TMTLRGRKCYLH--AWRE
 GwCas7x3 DVNLQPHG-----MPGRT-Q---SLSPAELLFGLVGTQNDQATT-----
 -----AYAGKVRIGFGLP--EEGH-----NP---RLDA-RITLKELSSPKPPSPALYFRKKSG-
 KDEYVSKANLADK-----PE---DYILRGRKMYLH--AWRK
 PID64649.1 E--LLPFN-----RNRS-----RVSPAELLFGFTENNPKEMKI-----D---
 RGLAFAGKIRISAGTL--SDKF-----SDTTESDLFEP-ETTLKALSSPKPPSPALYFKEKKS-
 GTQYIKKQDL--N-----PG---KHEIQGRKIYH--ALRN
 WP_031436019.1 E--LLPFN-----ESRK-----KVSPAELLFGFTQINKDGSKA-----DD-
 --KAQAFAGKVRISAGTI--SEYP-----EN--EANLLEQ-
 EVTLKALSTPKLSPALYFRTINGNGSAYISKQEL--E-----PS---KHLAKGRKYLYH--ALRT
 HEB50754.1 E--LRPFN-----RDRR-----AISPAELLFGFVELDESEHST-----
 EKSRYEQMAFAGKVRLSAGLP--VEDV-----ED---SALLEPKPIVLKALSSPKPPSPFFYFVMRDG-
 SGAYIAKKDL--S-----PD---RHRIKGRKHLYH--GLRQ

HrcCas7-11	S--LRPCA-GLCVE---DCNECPARCKK-VADYFNPHP-R----GLCPACHLFG-----
TT-----	FYKGRVRFGLAWLTGEDGAPRWYKGPD---PC--DSG-KGR-PMTIPLLERPRPT----
WSIPD-----	N-----SFDIPGRKFYVHHP-YSV
CmaCas7-11	S--LRSCS-HTCTE---DCDDCPNLCS-VKDYFKPHP-D----GLCPACHLFG-----
TT-----	FYKSRLSFGGLAWL---ENNAKWYISNDFQQKDS--KKE-KGG-KLTLPLLERPRPT----
WSMPN-----	N-----NAEVPGRKFYVHHP-WSV
HvmCas7-11	S--LRPCN-HVCVE---GCDECPDRCKE-LEDYFSPHP-E----GLCPACHLFG-----
TT-----	DYKGRVSFGLGWH--ESNTPKWYMP-----ED--NSQ-KGS-HLTLPLLERPRPT----
WSMPN-----	K-----KSEIPGRKFYVHHP-WSV
SmCas7x3	G--FQPCS-HVCLE---DCNTCSAKNCP-VPLYREGWPVN----GLCPACRLFG-----
AQ-----	MYKGRVNFGLFARL-----P-----DD--KQP-ETK-TLTLPLLERPRPT----WVLPK-----
-----SV-----	KGSNTEDATIPGRKFYLRLHDGWRI
oral_meta	E--YRPCA-HVCLE---ECDPCTGKDCP-VPIYREGYPAR----GFCPACQLFG-----TQ--
-----	MYKGRVRFSGFVP--VN-----ST--RSP-QLK-YVTLPSQERPRPT----WVLPE-----
S-----	CKGKEKDVPGRKFYLRLHDGWRE
DisCas7-11	D--LRPCH-GDWVE---DGDLSALNAYP-EKRLLLRHP-K----GLCPACRLFG-----
TG-----	SYKGRVRFGLFASL--ENDPEWLIPGK--NPG--DPF-HGG-PVMLSLLERPRPT----
WSIPG-----	S-----DN---KFKVPGRKFYVHHHAWKT
DsbaCas7-11	E--LRPCH-GEWIE---DGDLSKLDAYPAEKKLLTRHP-K----GLCPACRVFG-----
TG-----	SYKSRVRFGLFAAL---KGTpkWLKED-----PA--EPS-QGK-GITLPLLERPRPT----
WAVLH-----	ND-----KE---NSEIPGRKFYVHHNGWK
SstCas7-11	D--LRPCH-GEWIE---GDDLSQLSEYP-EKKLFTRNT-E----GLCPACRLFG-----TG--
-----	AYKGRRLRFGLFAKL--ENDPKWLMKN----SD--GPS-HGG-PLTLPLLERPRPT----WSMPD----
-----	DTLNRLKKDGKQEPKKQKGGKGPQVGRKFYVHHDGWKE
CsbCas7-11	A--LHPCT-GLS-----D-----GLCPGCHLFG-----TT-----
DYKGRVKFGFAKY--	ENGPEWLI-----TR--GNN-PER-SLTLGVLESPPA---FSIPD-----D-----
-----	ESEIPGRKFYLHHNGWRI
CjcCas7-11	N--LLPCV-HEVN---EGLLSGILDSL-DKKLLSIHP-E----GLCPTCRLFG-----TT----
-----	YYKGRVRFGLFANL--MNKPKWLT-----ER--ENG-CGG-YVTLPLLERPRLT----WSVPS-----
-----D-----	KCDVPGRKFYIHHNGWQE
CbfCas7-11	D--LRPCV-RDCLDT---EGDIRMLENSP-FKRLFHIHP-E----GLCPACQLFG-----TT--
-----	NYRGRVRFGLFASL---SDGPKWFR---KD--EGN-ETC-HITLPLLERPRPT----WSMPD-----
-----D-----	TSTIPGRKFYVHHMGYET
HvsCas7-11	E--LWPCVLAECPLLG-FECKKCELEGLP-EKIWFIRINK-D----GLCPACRLFG-----
TQ-----	IYKSRVRFSGFAYA-----KNW-----KF-YDG-YITLPRLESPPRAT----WLILK-----
---E-----	KDKHYIKYKVCGRKFYLHNSTYED
HsmCas7-11	H--LRSCV-RECLYEGWGDCKPCKLSLFP-EKMWIRINP-E----GLCPACHLFG-----
TQ-----	VYKGRVRFGLFARA---GSNW-----KF-REE-QLTLPRFETPRPT----WVIPK-----
---R-----	KD---EYQIPGRKFYLHHNGWEE
FmCas7-11	D--FLPCA-ATCLN---ECDTCNAKTCKFLPLYREGYPVN----GLCPSCHLFG-----
TT-----	GYQGRVRFGLFAKM--NGNAKFCQ-----GG--ERP-EDR-AVTLPLQERPKLT----
WVMPN-----	E-----NSTIPGRKFFLHHQGWKK
DpbaCas7-11	A--FRPCA-YVCIE---ECEPCDAKTCP-IPVYREGYPIK----GLCRACHLFG-----TT-
-----	GYKGRVRFSGFAKL---NGDAVW-----AK--GAG-GKD-YFTLPLLEKPRPT----WTMPN----
-----E-----	GAKIPGRKFYVHHNEWKT

Domains

SECTION 13

WmCas7x3	QG---QVARLDNSGHVPSDGSGRPPW-----QSRFDG-----QEDSGNK----
RRVRVEPIAGETHFEIDFDNLSPTLEQLCATLLP-HPAFEHRLGMGKPIGLGSVKLAVEGL---	
LLVDRPRRYAEDEPNAPRHRGW--RAN--ADAG--WPDHLQGDSPAAPLEAT-EQP-AALAERAMA---	
SER16298.1	QNG--NVMKLDAGVNSGGSTCKPPWKTHKPAANDQKE-----
FEEDKNKFITSRQVKIAPISENTPFYFEIDFNNLDATLAQLCATLQP-	
APKFEHRLGMGKPLGLGSVKIEPVGL--FLINRHQRYTTDSTNCDRYHYAW--LKG--	
EHAAWDWPEYFRQNVVTADCTQTFNDTFDKLVQNGLA-----	
GwCas7x3	NE---QVVELSDTGH---DGGVRPPW-----VSKFDE-----SADEGNK----
RRVSIEPIAKDESIFYFEVDFHNLSTELAQLCATLYP-NEKFEHRLGMGKPLGLGSIKITPLSL---	

FLVNRSQRYATDGLDKPRYHAVW--HTG--TASEPRWPDHLQREQQGIAFEGVSTAP-TVMSLAEEA--

 PID64649.1 ENN-QNVQRITSQGKFDNAANRTQPW-----VS-----QNEERNH---
 LKTKCKPLKSGLNFFHIDFNNLTQWELGLLCYALRP-CETFRHKIGMGKPIGLGTVKIDIAVL---
 QTIDRYARYTDTTQDSERYNQGA--WIS--QELQNEIPN--QYKKGKISNKKGMLSP-EDCRKVFME-----

 WP_031436019.1 GDN--KVQKLGSQGETANGGDSKLPW-----VTH-----NPDERPQ---
 LKVKIKPIKAE--FIFSLDFNNLTWELGLLCYALRP-TDSFRHRIGMGKPLGLGSKVIDIMAL---
 QTINRQORYAQDGLEENRFNRHN--WVN--PPHQPR---EKAGYSISLSSTPLNP-EILRATFTK-----
 HEB50754.1 RGNPDRVQSLDRYGH-ATETAANPPW-----ETC-----HPEERPQ---
 IKVRVQPVRRKTKFFHLDLFSNLSRWELGLLCYVLRP-TACFRHKLGMGKPLGLGSVRIDIASL---
 QLIDRVRRYGTDDLTAGRYNMGG--HFN--ASCLDLLP---QQDSPAPDDSGAAPDP-GTLRQDFVK-----

 HreCas7-11 ---D-----GI-----DGETRTP---
 NNRTIEPLAEGNEFVFDIDFENLRDWELGLLLYSLEL-EDSLAHKLGLGKPLGFGTVQINIRGI---
 SLKN-----GSKG--WDTKTGDDKNQWIK---KGF-----AHLGIDIKE-----
 CmaCas7-11 ENI-K-----NNQGN-----QKDISLKPDSDAIKIKE---
 NNRTIEPLGKDNVFNFEISFNNLRDWELGLLLYAIEL-EDHLAHKLGMKAFGMGSVKIEIKNL---
 LIKG-----SIND-----ISKAELIK---KGF-----KKLGIDSLEKD-----
 HvmCas7-11 DKI-R-----NRQFD-----PAK-EKQPD-DVIKPNE---
 NNRTVEPLGKGNFTFEVRFNNLRWELGLLLYSLEL-EDNMAHKLGMGKALGMGSARIKAEAI---
 ELRC-----ESA--GQNAELKDKA AFVR---KGF-----EFLEIDKPGEN-----
 SmCas7x3 -VM-A-----GTNPI-----TG-----ESIEKTA---
 NNATVEAIMPGATFTFDIVCENLDQQELGLLLYSLEL-EEGMSHTLGRGKPLGFGNVRIKVEKI---
 EKRL-----SDG-----SRREMIPP---KGA-----GLFMTDKVQDALRGLTEG
 oral_meta MWG-D-----DDKPD-----SRPSSEE---
 CQDIIEGIGPGEKFHFRVAFENLDKNELGRLLYSLEL-DAGMNHHLGRGKA FGFGQVKIRVTKL---
 ERRL-----EPGQ--WRS--EKICTDLPV---TSS-----ELVISSLKK-----
 DisCas7-11 -IK-D-----GNHPT-----TG-----KAIEQSP---
 NNRTVEALAGNSFSFEIAFENLKEWELGLLIHSLLQLEKGLAHKLGMKSMGFGSVEIDVESV---
 RLRK-----DWKQ--WRN-GNSEIPNWLG---KGF-----AKLKEWFRD-----
 DsbaCas7-11 -IS-E-----GIHPI-----SG-----ENIEPDE---
 NNRTVEVLDKGNRFVFEISFENLEPRELGLLIHSLLQLEKGLAHKLGMKSMGFGSVEIDVESV---
 RVKH-----RSGE--WDYKDGGETVDGWIE---EGK-----RGVAAKGKA-----
 SstCas7-11 -IN-C-----GCHPT-----TK-----ENIVQNQ---
 NNRTVEPLDKGNTFSFEICFENLEPYELGLLLYTLEL-EKGLAHKLGMKPMGFGSIDIEVENV---
 SLRT-----DSGQ--WKD-ANEQISEWTD---KKGK-----KDAGKWFKT-----
 CsbCas7-11 -IR-Q-----KQLEI-----RETVQPE---
 RNVTTVEVMDKGNVFSFDVRFENLREWELGLLQSLDP-
 GKNIHKLGLGKPYGFGSVKIKIDSLHTFKINS-----NNDK--IKRVPQSDIREYIN---KGY-----
 ----QKLIWWSGNNISIQKGNVL
 CjcCas7-11 -VL-R-----NN-----DITPKTE---
 NNRTVEPLAADNRFTFDVYFENLREWELGLLCYCLEL-
 EPGMGHKLGMGKPMGFGSVKIAIERLQTFTVHQ-----DGIN--WKP-SENEIGVYVQ---KGR---
 -----EKLVEWFTPSAPHKNM--
 CbfCas7-11 -VK-K-----NQ-----RTL VKTE---
 NNRTVKALDKENFTFEVFFENLREWELGLLHLCLEL-
 EPEMGHKLGMGKPLGFGSVKIRIDKLQKCVVNV-----KDGCVLWEP-EEDKIQHYIA---KGL--
 -----GKLTTWFGK-----
 HvsCas7-11 IIN-N-----SKK-----EKEKKTE---NNASFEVLKEG-
 EFTFKVYFENLENWELGLLLLSLT--GLGEAIKIGHAKPLGFGSVKIEAKKI---YFRE-----
 EAGK--FHP--CEKADEYLK---KGL-----NKLTSWFGK-----
 HsmCas7-11 IYK-K-----NKK-----NEIKKEK---NNATFEVLKQG-
 TFYFKVFFENLELWELGLLIFSAELGGEFAHKLGHGKALGFGSVKISVDKI---ILRR-----
 DPGQ--FEQRGQKFKRDAVD---KGF-----CVLENRFGK-----
 FmCas7-11 IVD-E-----GKNPI-----NG-----DVIEPDA---
 NNRTVEPLAAGNDFSFEVFFENLREWELGLLRYTLEL-ESELAHKLGMKAFGFGSVKIKIKSV---
 DLRK-----QGE-----WEK---ATN-----TLVSEDKKS-----

DpbaCas7-11	-VQ-E-----GKNPI-----DQ-----KAIRPNP----
NNSSVEVLNLGNEFQFEVSFENLEEWELGLLLYCLEL-EPGLAHKLGRGKAFGFGSIEAEVSKI--	
EMRI-----KSGT--WKNE-TSGKEKFIQ---SGL-----SQVPSFFKQDEK-----	
Domains	
.....	
SECTION 14	
WmCas7x3	--RVPADVRRALQLLGNP---GAVAAPVHYPQVKDAQIE--EKH---YLWVANDD--
-----EKTAGGNRH-----LPRL---HANSPGLPTLPRLV---KREKDHSSNTGKPRRK-----	
SER16298.1	--GTDADIKHALQLLGDP---QYIGVPVHYPIAGNSTLE--NKH---FEWFGNNDK-----
---ASVLRQKAQANSKNHHYQPKQQATPEEPQYLHTI----	
TKDSKQISLLKKNKIEDIENRDQQKHRYSNHRR-----	
GwCas7x3	---KVSDDVKRALELLGNP---DEITVPVHYPQLHNGLME--SKQ---FDWVQNDK----
-----SGRDQPANNRQH-----LSSF---TKDTEKLEPLIRIM-----RR-----	
PID64649.1	---TMDADIQRAIELLGDP---GNVTSPVHYPQLDRKNIE--TKN---YEWFKQNEI-----
-----EQQV-----LKPI---TKNTHTLTPFAR-----WEQG-----	
WP_031436019.1	---TMNADIYRTLELLGNP---QNVKRPVHYPQVENHNIE--QEN---YKWFVANDQ--
-----GSGKGRNKIDPAEKA-----LKIL---TENSCLPTLSRLD-----WRDE-----	
HEB50754.1	---TMDETVFRALDLLGNP---AHVQRPVHYPQVREMDIE--DQT---FLWFDNDK----
-----QWKDA-----LQPL---TSSSTQLPPLTRN-----KR-----	
HreCas7-11	--ANERPVIKQLRELLWVP--TGDNLPHVRYPELESKTKD----VPGYTSLLKE-----
--KDL-----ADR-----VSL--KA---PWKPWKPWS----	
GTAPHPDKGTNRLRASIVERDRIQRKTDAKPEKKEETKVGKSSSSDIEKRYVGTWKWFNDKKGYG	
FI-LYGTD	
CmaCas7-11	-DLSEYLHIKQLREILWFS---DKPVGITIEYPKLENKTNS---
RIPSYTDFVQEKDHETGFKNPKYQNL-----KSR-----LHIL---QN--PWNAWWKNE----E----	
HvmCas7-11	-DPMNFDHIRQLRELLWFL--PENVSANVRYPMLEKEDDG----
TPGYTDFIKQEEPSTGKRNPSTYSS-----EKR-----RNIL---QT---PWKHLYLIP----	
PFQASAQSETVFE-----GTVKWFDDKKGFGFIKINDGG	
SmCas7x3	GDWHQRPHISGLRRLTRY-----PEIKARYPKLSQGEDK----
EPGYIELKSQKDENGVPY---YNP-----NRE-----LRVS---EN---GPLPWFLA---KK-----	
oral_meta	----VEERRKLLRLVMTP---YKGLTACYPGLERENGR-----PGYTDLKML-----
ATY-----DPY-----RELVVQIGSNQ---PLRPWYEPG----	
KSFKPSPGNDCTGRGGSVSKSLISEPKVV-----PAIAPFCEGVVKWFNSVKGFGFI-ETKEQ	
DisCas7-11	---ELDFIENLKKLLWFP---EGDQAPRVCPMLRKKDDP--NGN-SGYEELKDGE-----
---FKK-----EDR-----QKKL---TT---PWTPWA-----	
DsbaCas7-11	-----NDLRKLLYLP--GEKQNPVHYPTLKKKEKKG---DPPGYEDLKKSFR-----
EKK-----LNR-----RKML---TT---LWEPWHK-----	
SstCas7-11	-DWEAAEHKLNKLLFLP--GEEQNPRVIYPALKQKDIP--NSRLPGYEELKKN-----
---LNM-----EKR-----KEML---TT---PWAPWHPIK---K-----	
CsbCas7-11	PQWHVIPHDKLYKLLWVPFLNDSKLEPDVRYPVNLNEESKGYIEGSDYTYKKLGDKDN-----LPY-
-----KTR-----VKGL---TT---PWSPWNPFG-----	
VIAEHEEQEVNVTGSRPSVTDKIERDGKMGV-----	
CjcCas7-11	-EWNGVKHIKDLRSLLSIP---GDKPTVKYPTLNKDAEG--AISDYTYERLSDTKL----
---LPH-----DKR-----VEYL---RT---PWSPWNAFV---KEAEYSPSEKSDEKGRETIRTKPK-	
-----SLPSVKSIGKVKWFDEGKGFGIL-IMDDG	
CbfCas7-11	-EWDRLHIQGLRSLQRL-----PL-----

HvsCas7-11	--NEINEHMRNLLLFMTYY----QNLPKVKYPDF-----DGYAKWRCS-----
YVE-----QDK-----VEYF---QN---RWIVAS-----	
HsmCas7-11	-TNFKI-YLNNFLQLLYWP---NNKKVKVRYPYLRQEDDP--EKL-PGYVELKKH-----
-----QML-----KDD-----NRYSLFARPA---VWLKWTEMV-----QRDKS-----	
FmCas7-11	-SWYNIHTVNNLRTALYYV---EDDKIQVNYPKLKKDNES---DNRPGYVEMKKA--
-----FPV-----RDI-----L-----TT---PWWPWWPPT-----PPPMNQSGNQSYARSEEPARITE---	
	-----SQPEVYKTGTGVKFKYKHDKKFGFITMDGRE
DpbaCas7-11	-QWNKVEQVKNIRKLLQLSWNKGNAVEPEVRYPALREKDDE--NKR-
PGYVELKDNG-----Y-----DA-----GKKL---VS---PWAPWHPIK---K-----	
Domains	
SECTION 15	
WmCas7x3	-----
SER16298.1	-----
GwCas7x3	-----
PID64649.1	-----
WP_031436019.1	-----
HEB50754.1	-----
HrcCas7-11	EEIFVHRSGVADNSIPKEGQKVGFRIERGARGSHAVEVKAIE---
CmaCas7-11	-----
HvmCas7-11	KDVFVHHSSIVGTGFKSLNEGDSVAFKMGVGPKGPCAEKVKKIGN
SmCas7x3	-----
oral_meta	RDIFVHFSAIRGEGYKILEPGEKVRFEIGEGRKGPQAINVIRIR-
DisCas7-11	-----
DsbaCas7-11	-----
SstCas7-11	-----
CsbCas7-11	-----
CjcCas7-11	KEVSISKNSIRGNILLKKGQKVTFHIVQGLIPKAEDIEIAK----
CbfCas7-11	-----
HvsCas7-11	-----
HsmCas7-11	-----
FmCas7-11	NIHFAGNQICRPET--SLQSGDKVKFIEGENYKGPTALKVERLKG
DpbaCas7-11	-----
Domains	

EXAMPLES

[0323] While several experimental Examples are contemplated, these Examples are intended non-limiting.

5

Example 1 Cas7-11 Family

[0324] New Cas7-11 family members were mined from bacterial and metagenomic data downloaded from NCBI, JGI, and ENA repositories. CDS were extracted using MetaGeneMark (Besemer and Borodovsky 1999; Zhu *et al.* 2010), and used to build BLAST databases (Camacho *et al.* 2009). Existing Cas7-11 family members (Makarova *et al.* 2019) were used as seeds to find putative orthologs. In parallel CRISPR arrays were identified using both CRISPRCasFinder (Couvin *et al.* 2018) and PILER-CR (Edgar 2007), and putative orthologs were filtered for proximity to CRISPR arrays.

10

[0325] Cas7-11 family members were computationally identified and expressed. The Cas7-11 family is summarized in the trees presented in FIG. 2A. The type III-E Cas7-11 locus from *Desulfonema ishimotonii* (DisCas7-11) comprises multiple Cas7-like gRAMP domains, including a small subunit domain and an unknown insertion in the Csm3-like domain 5 (FIG. 2B). The CRISPR array was expressed in *Desulfonema ishimotonii* and processed into mature crRNAs containing 14-nt direct repeats (DRs) and ~31 nt spacers (FIG. 2C). The cleavage site for processing and maturation is illustrated in FIG. 2D.

[0326] By screening a database of bacterial genomic and metagenomic sequences from diverse sources containing more than ~11.6 billion protein sequences, the set of subtype III-E systems was expended to 17 loci from various bacteria and metagenomes (FIG. 2A). Apart from the identification of many additional type III-E loci, a novel subset of III-D systems (subtype III-D2) that was characterized by an the fusion of 3 Cas7 domains into a single protein, referred to as Cas7x3 was identified (FIG. 2A and FIG. 2B and Table 1). This similarity of the domain architectures, together with the phylogenetic analysis of the Cas7 domain (FIG. 2A), suggests an evolutionary path from subtype III D1 to III-D2 via the fusion of 3 Cas7 proteins, and then, from III-D2 to III-E via the fusion with an additional Cas7 protein and the Cas11 protein.

[0327] Examination of the extended set of type III-E systems confirmed that almost all III-E loci (15 of the 17) contained the TPR-CHAT protein (Csx29), and a majority encompassed uncharacterized proteins denoted Csx30 (10/17) and Csx31 (9/17) as well as the alternative sigma factor RpoE (8/17) (FIG. 1E). The frequent presence of these genes in the III-E loci suggests that the respective proteins play accessory roles in the CRISPR response. More than half of the III-E loci (9/17) also contain the adaptation gene Cas1 which, in 8 of these 9 loci, is fused to a reverse transcriptase (RT), in a common arrangement among the type III systems, as well as a CRISPR array (12/17). Searching the spacer sequences against the NCBI NT (nucleotide sequences) database did not identify any matches, in agreement with the previous observations that the majority of the CRISPR spacerome consists of “dark matter” that apparently reflects vast, still unexplored viromes. Furthermore, alignments of the Cas7 domains of the Cas7x3 and Cas7-11 proteins show conservation of acidic residues implicated in crRNA guided target cleavage (FIGS. 1F-1K). The activities of Cas7-11 and Cas7x3 proteins were experimentally characterized to assess the path of evolution that resulted in the unique Class 1 single protein effector that is predicted to function as a programmable RNase without collateral activity.

Example 2
Cas7-11 Defense

[0328] The RNA-guided RNA-targeting Cas7-11 was demonstrated to be capable of defense against ssRNA MS2 phage and RNA knockdown in bacteria.

[0329] The CRISPR array screen of all crRNAs targeting the MS2 genome was performed (FIG. 3A). The results of the MS2 interference screen is shown in FIG. 3B (for MS2 targeting spacers) and FIG 3C (for non-targeting spacers). Regarding FIG. 3B, the results of the MS2 interference screen are shown as box plots. Enrichment of DisCas7-11 spacers in the phage targeting condition denote survival of bacteria and enhanced representation of specific active spacers. Boxes denotes 25th and 75th percentiles with the median marked by the middle line. The whiskers are calculated via the Tukey method (1.5 times the inter-quartile range).

Outliers are denoted by blue plus symbols. FIG. 3C shows the results of the MS2 interference screen showing enrichment of DisCas7-11a non-targeting spacers across varying phage dilution amounts as box plots. Boxes denotes 25th and 75th percentiles with the median marked by the middle line. The whiskers are calculated via the Tukey method (1.5 times the inter-quartile range). Outliers are denoted by blue plus symbols.

[0330] The enrichment of DisCas7-11 crRNAs in the phage targeting condition denotes the survival of bacteria and enhanced representation of specific active crRNAs. The results of the MS plaque dilution assay are shown in FIG. 3D. The assay shows significant survival of *E. coli* containing the top crRNA identified in the MS2 screen. In the non-targeting spacer, dilutions of phage up to $1e9$ are capable of lysing bacteria. For the top targeting spacer, only dilutions up to $1e4$ are capable of lysis because of DisCas7-11 interference.

[0331] The number of DisCas7-11a spacers that display survival enrichment over a threshold of 1.7 across different phage dilution conditions are shown in FIG. 3E. The quantification of resistance conferred by top MS2-targeting DisCas7-11 spacers compared against a panel of 4 non-targeting spacers is shown in FIG. 3F. The resistance is quantified as the highest surviving titer of MS2 phage that generates plaques in the dilution assay.

[0332] The quantification of resistance conferred by two MS2-targeting DisCas7-11 spacers compared against a panel of 6 non-targeting spacers is shown in FIG. 3G. Resistance is quantified as the final dilution of MS2 phage that generates plaques in the dilution assay.

[0333] The RFP knockdown assay containing a DisCas7-11 and spacers expression vector co-transformed with an RFP-expressing vector is illustrated in FIG. 3H, and the results from the knockdown assay are shown in FIG. 3I. The RFP knockdown assay by multiple targeting

spacers is normalized to a non-targeting condition as determined by flow cytometry.

DisCas7-11-mediate RFP knockdown was found to exceed 50%.

[0334] The DisCas7-11a full locus with TPR-CHAT nearby and DisCas7-11a effector expressed alone with TPR-CHAT are illustrated in FIG. 3H, and the MS2 interference activity of DisCas7-11a in the full locus and as a single effector with and without TPR-CHAT expression is shown in FIG. 3I.

Example 3

Cas7-11 Cleavage Activity

[0335] The DisCas7-11 programmable RNA cleavage activity was biochemically characterized. The CRISPR array processing assay for DisCas7-11 is illustrated in FIG. 4A and shows robust processing activity by DisCas7-11. The DisCas7-11 was assayed at concentrations ranging between 0 nM and 233 nM. The *in vitro* cleavage of ssRNA 1 target is illustrated in FIG. 4B and demonstrates that programmable DisCas7-11 RNA targeting is Mg²⁺ and crRNA dependent. Note that the targets are either 5' or 3' fluorescently labeled.

The results from the cleavage of EGFP ssRNA at multiple sites with tiled crRNAs are shown in FIG. 4C. A varying cleavage pattern based on target position is observed. The protospacer flanking site (PFS) sequence screen is illustrated in FIG. 4D and shows a varying cleavage pattern based on target position and a lack of sequence preference flanking the target site for DisCas7-11 cleavage. The kinetics of fluorescence due to RnaseAlert collateral cleavage by DisCas7-11 or LwaCas13a targeting and non-targeting crRNA against the MS2 RNA target are illustrated in FIG. 4E. The processing activity of a synthetic DisCas7-11 CRISPR array by DisCas7-11 protein with predicted catalytic processing mutants in the protein insert region with a characteristic KxYxH catalytic triad motif is illustrated in FIG. 4F. The measurement of bacterial growth (OD₆₀₀) during target interference by DisCas7-11 and LwaCas13 is shown in FIG. 4G.

Example 4

Cas7-11 Catalytic Residues

[0336] The catalytic residues in DisCas7-11 were identified.

[0337] The protein domains of DisCas7-11 with putative catalytic residues highlighted in Cas7-like domains 3 and 4 are illustrated in FIG. 5A. The results from the cleavage of synthetic ssRNA target with two crRNAs and wild type, D429A, D654A, and D429A/D654A DisCas7-11 proteins are shown in FIG. 5B. The cleavage is affected by the single mutants and completely inactivated by the double mutant D429A/D654A. The processing of a synthetic DisCas7-11 CRISPR array by catalytic DisCas7-11 mutants is not affected by

D429A and D654A catalytic mutations (FIG. 5C). The MS2 phage interference in bacteria is affected by D429A/D654A mutations, although there is remaining residual interference above the lack of DisCas7-11 altogether (FIG. 5D). The RFP knockdown in *E. coli* is prevented by the catalytic mutation D429A/D654A in DisCas7-11 (FIG. 5E).

- 5 **[0338]** The DisCas7-11:crRNA complex binding to a complementary MS2 ssRNA target is determined by electrophoretic mobility shift assay (EMSA) and illustrated in FIG. 5F. An EMSA is performed for both the WT DisCas7-11 and dead DisCas7-11 complexes and is compared to WT DisCas7-11 in complex with a non-targeting crRNA. The K_d value for the WT and dead DisCas7-11 complexes are shown below corresponding gels.

10

Example 5 Cas7-11 Cleavage Sites

- [0339]** DisCas7-11 cleavage sites were characterized. The DisCas7-11 guides targeting multiple sites on a synthetic 100nt EGFP ssRNA target are illustrated in FIG. 6A and the cleavage patterns of each EGFP-targeting guide incubated with DisCas7-11 on a 3' labeled EGFP ssRNA are shown in FIG. 6B. The DisCas7-11 guides targeting of multiple sites on
15 either MS2 ssRNA or ssRNA target 2, two targets of equivalent length (153 nt), are illustrated in FIG. 6C. The guides are designed to target sites at the same position on both targets in order to better understand whether sequence specific cleavage preferences exist and whether cleavage cut sites are position dependent. The cleavage patterns of each guide
20 incubated with MS2 ssRNA or ssRNA target 2 are shown in FIG. 6D.

Example 6 Cas7-11 Orthologs Activity

- [0340]** The CRISPR processing and RNA cleavage activity were demonstrated for different DisCas7-11 orthologs.
- 25 **[0341]** The schematics of three Cas7-11 orthologs (GwCas7-11, CjcCas7-11, and DisCas7-11) chosen for additional characterization are shown in FIG. 7A, wherein a multiple sequence alignment of the three orthologs compared against the domain organization of a typical Cas7-11 protein is shown. The multiple sequence alignments of the DR sequences for GwCas7-11, CjcCas7-11, and DisCas7-11 are shown in FIG. 7B. Mismatches against the consensus are
30 highlighted in red and the grey region shows the sections of the DRs that are cleaved from the pre-crRNAs to generate mature crRNAs.
- [0342]** The activity of the GwCas7-11 orthologs at concentrations between 2 uM to 125 nM was assessed. The activity of Cas7-11 from *Candidatus Jettenia caeni* (CjcCas7-11) shows robust processing of CRISPR arrays specific to the CjcCas7-11 locus (FIG. 7C). The activity

of Cas7-11 from ground water metagenomes (GwCas7-11) shows robust processing of CRISPR arrays specific to the GwCas7-11 locus (FIG. 7D). MS2 ssRNA target cleavage by GwCas7-11 and CjcCas7-11 incubated with their cognate crRNAs using guides complementary to the MS2 ssRNA target is illustrated in FIG. 7E. The incubation of GwCas7x3 and CjcCas7-11 proteins with their cognate crRNAs complementary to the MS2 ssRNA target at varying positions along the target is illustrated in FIG. 7F.

Example 7

Cas7-11 RNA Knockdown in Mammalian Cells

[0343] The knockdown of RNA in mammalian cells by DisCas7-11 was evaluated. The schematic of a luciferase reporter assay for RNA knockdown using a vector expressing human codon-optimized DisCas7-11 and guide RNA is shown in FIG. 8A. The knockdown activity of DisCas7-11 against the *Gaussia* luciferase (Gluc) transcript normalized to two non-targeting conditions is illustrated in FIG 8B. Note that the dotted line in this figure represents 85% knockdown. The knockdown activity of DisCas7-11 against the multiple endogenous transcripts normalized to two non-targeting conditions is illustrated in FIG. 8C. Note that the dotted lines in this figure represent background expression and 65% knockdown by DisCas7-11.

[0344] RNA A-to-I editing of *Cypridina* luciferase (Cluc) mRNA W85X mutation in mammalian cells by dead DisCas7-11a-NES-ADAR2 was investigated. Guides were designed with mismatch distances between 2-50nt. Editing was measured by two methods: 1) restoration of Cluc luciferase activity that is normalized to the non-targeting guide condition (FIG. 8E) and 2) amplicon sequencing of the targeted adenosine to measure percent conversion of the adenosine to inosine (FIG. 8F).

[0345] The RFP knockdown by DisCas7-11 or LwaCas13a in *E. coli* bacteria was assessed (FIG. 8G). The knockdown of Gluc mRNA in mammalian cells by DisCas7-11 with a panel of guides containing the mature DR sequence was assessed (FIG. 8H). The guides were designed to be tiled across the Gluc transcript. The knockdown of Gluc mRNA in mammalian cells by DisCas7-11 with a panel of guides of different lengths was assessed (FIG. 8I). Comparison of the number of significant off-targets when targeting the Gluc transcript in HEK293FT and U87 cells by DisCas7-11, shRNA, LwaCas13a, PspCas13b, and RfxCas13d using transcriptome-wide RNA-sequencing data is shown in FIG. 8G. Off-targets were determined by significance testing of differentially expressed transcripts between targeting and non-targeting guide conditions. Comparison of the number of significant off-targets when targeting the Gluc transcript in HEK293FT and U87 cells by DisCas7-11,

shRNA, LwaCas13a, PspCas13b, and RfxCas13d using transcriptome-wide RNA1217 sequencing data is shown in FIG. 8K. Off-targets were determined by significance testing of differentially expressed transcripts between the targeting guide condition and cells only expressing EGFP. The data were mean ($n=3$) $\pm$ s.e.m.

- 5 [0346] Given that DisCas7-11 lacked collateral activity *in vitro* or *in vivo*, the efficiency and effects of targeting with DisCas7-11, multiple Cas13 family members, and RNA interference through small hairpin RNA (shRNA) was compared. The activity in HEK293FT cells was assessed, for which no detectable growth impact from Cas13a targeting has been reported. Comparing Gluc knockdown between DisCas7-11, LwaCas13a, PspCas13b, RfxCas13d, and
- 10 shRNA, comparable knockdown efficiencies was found for the different systems, with the exception of LwaCas13a, which could only reach ~35% knockdown (FIG. 8L). By assessing the specificities of the different systems by whole transcriptome profiling in HEK293FT cells and comparing between targeting and non-targeting guide conditions, it was found that there were more differentially expressed genes for shRNA and RfxCas13d than for DisCas7-11
- 15 (Fig. 8M). LwaCas13a and PspCas13b had minimal off-target effects on the transcriptome when comparing targeting versus non-targeting guide conditions, and none of these knockdown tools had any effect on HEK293FT cell viability (FIGS. 8M-8N). In addition, when comparing targeting conditions of the different tools to a no-protein, GFP condition, only RfxCas13d expression was found to result in detectable differential gene expression.
- 20 Because cell toxicity of Cas13 has been reported in certain cell types, such as the U87 glioblastoma cell line⁷, the degree of toxicity due to the collateral activity of Cas13 may be cell-type specific. Thus, the comparison of shRNA, Cas13a/b/d, and Cas7-11 knockdown activities was repeated in U87 cells, finding efficient knockdown by DisCas7-11, PspCas13b, and RfxCas13d, less efficient knockdown with LwaCas13a, and no activity from shRNA
- 25 (FIG. 8O). Whereas DisCas7-11 and the Cas13 variants showed minimal differential gene expression off-targets when comparing between targeting and non-targeting conditions, evaluation of targeting conditions versus the no protein, GFP-only control showed increased numbers of differentially expressed genes in RfxCas13d, LwaCas13a, and PspCas13b compared to DisCas7-11 (FIG. 8P). Moreover, Cas13 enzymes and shRNA caused cell
- 30 toxicity resulting in ~30-50% cell death, whereas DisCas7-11 expression had minimal impact on cell viability (FIG. 8Q), consistent with the specific knockdown demonstrated by transcriptome sequencing. The observation that certain cell lines can be more susceptible to non-specific RNA degradation than others is consistent with other reports on substantial

RNAi toxicity in diverse cell lines. Moreover, the ineffectiveness of the RNAi constructs in U87 cells herein, in contrast to their functionality in HEK293FT cells, is consistent with the low Dicer and Argonaute 2 expression in U87 cells, causing poor on-target knockdown, with off-target induced toxicity. The highly specific and low toxicity knockdown activity of DisCas7-11 compared to other RNA knockdown tools, RNAi and Cas13, could make Cas7-11 proteins more generally applicable and appropriate tools for RNA knockdown across diverse cells and tissues.

Example 8

Heterologous Expressions of Type-III-E CRISPR Cas System

[0347] The heterologous expression of the *Desulfonema ishimotonii* Type III-E CRISPR-Cas system and associated CRISPR array were performed. The *Desulfonema ishimotonii* Type III-E full locus matures crRNAs (FIG. 9A), single effector protein DisCas7-11 matures crRNAs in *E.coli*. (FIG. 9B) and DisCas7-11 CRISPR array without any effector or accessory proteins (FIG. 9C) were expressed. Additional type III-E family members and loci architecture are illustrated in FIG. 9D.

Example 9

Cas7-11 Dilution Assays

[0348] Dilution assays with DisCas7-11 and different spacers were performed. The DisCas7-11 assays for MS2 targeting spacers, non-targeting spacers and RFP targeting spacers were performed with dilution of phage up to 1e-1 (FIG. 10). The assay shows that bacteria with DisCas7-11a and MS2 targeting guides can survive at higher viral loads

Example 10

Effects Genes on MS2 Interference

[0349] The effects of genes in the *Desulfonema ishimotonii* type III-E CRISPR-Cas locus on MS2 interference were analyzed. The Type III-E locus of the *Desulfonema ishimotonii* is illustrated in FIG. 11A. Results of the phage plaque assay of the incomplete Type III-E DisCas7-11 CRISPR-Cas locus are shown in FIG. 11B. The knock-out of TPR-CHAT shows the highest positive effect on the interference by DisCas7-11. Results from the phage plaque assay of the individual gene knock-in together with of the Type III-E DisCas7-11 CRISPR-Cas system are shown in FIG. 11C. Knock-in data shows that RpoE has an inhibitory effect on DisCas7-11.

Example 11

Cas7-11 Pre-crRNA Processing to DR Mutations

[0350] The robustness of DisCas7-11 pre-crRNA processing to DR mutations was evaluated. The *in vitro* transcription of pre-crRNA and processing by DisCas7-11 is illustrated in FIG. 12A. The sequence of a single spacer CRISPR array showing locations of mutated bases is shown in FIG. 12B. The transcribed pre-crRNA was processed and the results showed the effect of consecutive double mutations on processing (FIG. 12C). Transcribed pre-crRNA was processed and the results showed the effect of consecutive double mutations on processing (FIG. 12D).

Example 12
Effect of Ion on Cas7-11 Activity

[0351] The effects of ions on pre-crRNA processing and ions, complex concentrations, and incubation time on target cleavage was analyzed. The *in vitro* transcription and 5' end labeling of pre-crRNA and processing by DisCas7-11 were assessed. Pre-crRNA was processed by DisCas7-11 in the presence of different ions or chelating agents (FIG. 13A). CRISPR array processing by DisCas7-11 is shown to be ion independent. A schematic of *in vitro* cleavage of ssRNA target with DisCas7-11 and crRNA 5 is shown in FIG. 13B. The ssRNA target was incubated with DisCas7-11 and targeting or non-targeting crRNA in the presence of different ions or chelating agents (FIG. 13C). The results show the dependence on magnesium, manganese, and calcium. The cleavage of ssRNA target with increasing amounts of DisCas7-11-crRNA complex from 0 nM to 233 nM was performed (FIG. 13D). The cleavage of ssRNA target at increasing incubation periods from 0 minutes to 180 minutes was also performed (FIG. 13E).

Example 13
Effect of DR Length, Spacer Length and DR Mutations on Cas7-11 Activity

[0352] The effects of DR length, spacer length, and DR mutations on *in vitro* target cleavage by DisCas7-11 were evaluated. A schematic of the sequence of DisCas7-11 crRNA 5 targeting the ssRNA target is shown in FIG. 14A. The ssRNA target was cleaved with crRNA of varying DR and spacer lengths (FIG. 14B). A schematic of the sequence of DisCas7-11 crRNA 1 targeting the MS2 target is shown in FIG. 14C. The MS2 ssRNA target was cleaved (FIG. 14D). The results show the effect of single and double mutations at various positions in the DR, and several single and double mutations cause inhibition of DisCas7-11 target cleavage.

Example 14
DisCas7-11 In Vivo Collateral Activity

[0353] The *in vivo* collateral activity of DisCas7-11 was evaluated. The results from the three-hour end point fluorescence of RnaseAlert reporter incubated with either DisCas7-11 or LwaCas13a targeting and non-targeting guides against MS2 ssRNA target are shown in FIG. 15. The LwaCas13a demonstrates robust collateral activity whereas DisCas7-11 has no collateral cleavage of the RNase alert reporter.

Example 15

Activity of DisCas7-11 Catalytic Mutants

[0354] The cleavage activity of DisCas7-11 catalytic mutants on MS2 RNA target was evaluated. A schematic showing the positions of DisCas7-11 crRNA 1 and crRNA 2 targeting the MS2 target is shown in FIG. 16A. The cleavage of a synthetic MS2 ssRNA target with two cRNAs and wild type, D429A, D654A, and D429A/D654A DisCas7-11 proteins was performed (FIG. 16B). The D429A/D654A mutant demonstrates reduced cleavage activity on the target.

[0355] The DisCas7-11 catalytic mutants that were assessed are presented in Table 16 below.

Table 16.

Mutation	<i>In vitro</i> status	Domain
E70	Purified, no effect	1
D77	Purified, no effect	1
E93	Purified, no effect	1
D111	Purified, no effect	1
D143	Purified, no effect	1
D160	Purified, no effect	1
D162	Purified, no effect	1
D177	Purified, cut inactivated, bottom	1
D185	Purified, no effect	1
D193	Purified, no effect	1
D226	Purified, no effect	1
D238	Purified, no effect	1
E280	Purified, no effect	2
D287	Purified, no effect	2
W315	Purified, no effect	2
D316	Purified, no effect	2
D379	Purified, no effect	2
E381	Purified, no effect	2
E390	Purified, no effect	2
D422	Purified, no effect	3
D429	Purified, cut inactivated, middle (top shifts up)	3
D437	Purified, no effect	3

D454	Purified, no effect	3
D487	Purified, no effect	3
E497	Purified, no effect	3
E520	Purified, no effect	3
D543	Purified, no effect	3
D581	Purified, no effect	3
D642	Purified, no effect	4
D654	Purified, cut inactivated, bottom	4
D727	Purified, no effect	4
D733	Purified, no effect	4
D740	Purified, no effect	4
D745	Purified, cut inactivated, bottom	4
D753	Purified, no effect	4
D758	Purified, cut inactivated, bottom	4
D781	Purified, no effect	4
D798	Purified, no effect	4
D825	Purified, no effect	4
D909	Purified, no effect	5
D914	Purified, no effect	5
E959	Purified, cut inactivated, bottom	5
D970	Purified, no effect	5
D982	Purified, no effect	5
D998	Purified, cut inactivated, bottom	5
D1015	Purified, no effect	5
D1025	Purified, no effect	5
D1118	Purified, no effect	5
D1179	Purified, no effect	5
D1273	Purified, no effect	5
D1297	Purified, no effect	5
D1308	Purified, no effect	5
D1367	Purified, no effect	5
D1404	Purified, no effect	5
D1539	Purified, no effect	5
D1569	Purified, no effect	5
D1588	Purified, no effect	5

Example 16
Cas7-11 EMSA Binding Assays

5 [0356] Electrophoretic mobility shift assays (EMSA) for crRNA-target interactions and quantification of EMSA intensities were performed (FIGS. 17A-F). The EMSA assays for targeting crRNA in the absence of DisCas7-11 show minimal binding of the crRNA and shift of the labeled target). Target binding was found to be crRNA-dependent, as expected, and the dead mutant to have a similar affinity to ssRNA as the active DisCas7-11 protein.

Furthermore, agreeing with the lack of ssDNA cleavage, DisCas7-11 showed weak ssDNA binding, and efficient binding required the full protein complexed with a.

Example 17

DisCas7-11 Cleavage Sites on ssRNA Targets

- 5 [0357] The characterization of DisCas7-11 cleavage sites on ssRNA targets was performed. [0358] DisCas7-11 was incubated with a panel of crRNAs targeting either MS2 ssRNA or ssRNA target 2, two targets of equivalent length (153 nt) (FIG. 18A). The crRNAs are designed to target sites at the same position on both targets in order to better understand whether sequence- specific cleavage preferences exist and whether cleavage cut sites are
- 10 position dependent. The cleavage of ssRNA 1 and MS2 ssRNA at multiple sites with tiled crRNAs were also performed (FIG. 18B). A varying cleavage pattern based on target position was observed.
- [0359] A schematic of the position of the tested crRNAs along the SARS-CoV-2 ssRNA is shown in FIG. 18C. The results from the cleavage of synthetic SARS-CoV-2 ssRNA at
- 15 multiple sites with tiled crRNAs showing a varying cleavage pattern based on target position are also shown in the same figure. A schematic of the position of the tested crRNAs along the Influenza A ssRNA target is shown in FIG. 18D. The results from the cleavage of synthetic influenza A segment 4 HA gene ssRNA at multiple sites with tiled crRNAs showing a varying cleavage pattern based on target position are also shown in the same figure.
- 20 [0360] DisCas7-11a was incubated with a crRNA targeting a crRNA site flanked by different sequences (FIG. 18E). While the spacer:target complementarity region is fixed, the sequences around the site are completely variable, allowing determination of any sequence preferences for cleavage. DisCas7-11a was also incubated with a crRNA targeting a double-stranded (dsDNA) target with 5' labeling of the bottom strand (FIG. 18F) or top strand (FIG. 18G).

Example 18

Cleavage Activity on Various Length EGFP ssRNA Substrates

- [0361] Biochemical characterization of DisCas7-11 programmable RNA cleavage activity on various length EGFP ssRNA substrates was performed. A schematic of the position of the tested crRNAs along the 100 nt EGFP ssRNA target is shown in FIG. 19A. The biochemical
- 30 characterization of programmable DisCas7-11 RNA cleavage of 100 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs is also shown in this figure. A schematic of the position of the tested crRNAs along the 200 nt EGFP ssRNA target is shown in FIG. 19B. The biochemical characterization of the programmable DisCas7-11 RNA cleavage of 200 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs

is also shown in this figure. A schematic of the position of the tested crRNAs along the 600 nt EGFP ssRNA target is shown in FIG. 19C. The biochemical characterization of the programmable DisCas7-11 RNA cleavage of 600 nt long and 3' fluorescently labeled EGFP ssRNA target with tiling crRNAs is also shown in this figure. The DNA sequencing analysis of selected EGFP targeting crRNA conditions on an unlabeled 400 nt EGFP ssRNA target was performed (FIG. 19D). The DisCas7-11a cleavage was found to occur at both sides directly flanking the crRNA binding site, with cleavage occurring approximately 20 bases downstream of the 5' position of the spacer. Other biochemical characterizations of DisCas7-11 programmable RNA cleavage activity are shown in FIGS. 19E-19G

Example 19

Candidatus *Jettenia caeni* Type III-E CRISPR-Cas system

[0362] The heterologous expression of the *Candidatus Jettenia caeni* Type III-E CRISPR-Cas system and associated CRISPR array in *E. coli* was performed. The schematic of *Candidatus Jettenia caeni* locus (minus Cas1) in *E. coli* is shown in FIG. 20. The processing of crRNAs occurs when the CjcCas7-11 locus is heterologously expressed in *E. coli*, and mature 37 nt crRNAs are generated containing a 15-nt DR.

Example 20

Processing of Cas7-11 orthologs

[0363] The processing of Cas7-11 orthologs was found to be specific to the cognate pre-crRNAs and to do not occur on other synthetic RNAs. The DisCas7-11 pre-crRNA processing is specific to the DisCas7-11 array and does not occur on the MS2 ssRNA target as illustrated in FIG. 21A. The GwCas7-11 pre-crRNA processing is specific to the GwCas7-11 array and does not occur on the MS2 ssRNA target as illustrated in FIG. 21B. The CjcCas7-11 pre-crRNA processing is specific to the CjcCas7-11 array and does not occur on the MS2 ssRNA target as illustrated in FIG. 21C. FIG. 21D shows the comparison of target cleavage between active CjcCas7-11 and dead CjcCas7-11.

Example 21

Cas7-11 Correction of Cluc W85X

[0364] The correction of Cluc W85X mRNA by dDisCas7-11 was measured. A schematic of the DisCas7-11 guide design for RNA editing of Cluc W85A target mRNA is shown in FIG. 22A. A table disclosing the different guides used for the correction of the mRNA is shown in FIG. 22B. The dDisCas7-11 correction of Cluc W85X RNA by the different guides is shown in FIG. 22C, and the percentage editing for the correction is shown in FIG. 22D. The luciferase correction that correlates to the correction percent is shown in FIG. 22E.

Example 22

Harvesting and Next-generation Sequencing of crRNA and pre-crRNA From *Desulfonema ishimotonii* and *E. coli*

[0365] To extract total RNA, freeze-dried *Desulfonema ishimotonii* cultures (DSMZ 9680) were resuspended in TRIzol (R2051, Zymo), homogenized by bead beating with zirconia/silica beads, and processed with the Direct-Zol RNA miniprep protocol (R2051, Zymo). Purified RNA samples were treated with T4 Polynucleotide Kinase (M0201S, New England Biolabs), both with and without ATP to allow for the enrichment of 3'-P and 5'-OH ends. Samples has ribosomal RNA depleted using the Ribo Minus Kit (K155004, Thermo Fisher Scientific) before being processed for next-generation sequencing with the NEBNext Small RNA Library Prep Set for Illumina sequencing (E7330S, New England Biolabs). Notably, the PCR extension step was increased to 1 minute to allow for longer templates to be included in the library. Libraries were sequenced on an MiSeq (Illumina) to sufficient depth and analyzed using the alignment tool BWA (Li and Durbin 2009). Paired-end alignments were used to extract entire transcript sequences using Galaxy tools (<https://usegalaxy.org/>), and these sequences were analyzed using Geneious 8.1.5 (Biomatters, Auckland, New Zealand) and custom scripts (<https://github.com/abugoot-lab>).

Example 23

Design and Cloning of Bacterial Constructs

[0366] To clone DisCas7-11 constructs, genomic DNA from *D. ishimotonii* cultures (DSMZ 9680) was extracted using the Blood & Cell Culture DNA Mini Kit (13323, Qiagen). The Type III-E *D. ishimotonii* CRISPR locus was PCR amplified and cloned into a pACYC184 backbone with chloramphenicol resistance using Gibson cloning. For the Cas7-11 only construct, DisCas7-11a was cloned with a J23119 promoter and B0015 terminator. Removal or addition of other genes in the CRISPR locus was cloned using Gibson cloning; when adding genes to the Cas7-11-only construct, expression of these genes was driven by a pLac promoter. In order to facilitate simple cloning of spacers, a minimal spacer construct with two flanking direct repeats (DR) surrounding a Golden Gate acceptor site was used as a backbone for Golden Gate cloning (Engler and Marillonnet 2014). To clone the CjcCas7-11b locus, DNA corresponding to the locus was synthesized by GeneArt (Thermo Scientific). Unless otherwise noted, all clonings were transformed into Stbl3 cells (C737303, Thermo Fisher Scientific) made competent with the Mix and Go kit (T3001, Zymo Research), all colonies were picked into Terrific Broth (TB) (24g/L Yeast Extract, 12g/L Tryptone, 9.4g/L K₂HPO₄, 2.2g/L KH₂PO₄, pH 7.2) supplemented with the appropriate antibiotic (chloramphenicol,

25µg/mL; ampicillin, 100µg/mL) and all plasmids were isolated using the QiaPrep Spin Miniprep kit (27104, Qiagen) protocol and verified by next generation sequencing on a MiSeq (Illumina). All bacterial plasmids used are listed in Table 3.

Example 24

Cloning of Screening Libraries for MS2 Interference Activity Screen

[0367] To design the MS2 library for screening, all possible spacers targeting the MS2 genome were computationally extracted, and synthesized as a library by Twist Biosciences. This library was Golden Gate cloned into the minimal CRISPR array containing the Golden Gate acceptor site on the DisCas7-11a-only backbone and transformed into Endura Duo electro-competent cells (60240-1, Lucigen) and plated on chloramphenicol-supplemented Bioassay agar plates. 16 hours post transformation, libraries were harvested by scraping and purified using a NucleoBond Xtra MaxiPrep EF (740424.10, Takara Bio).

Example 25

Bacterial Phage Interference PFS Screen Assay

[0368] To transform the MS2 spacer library, 400µL NovaBlue(DE3) Gigasingle competent cells (71227-3, EMD Millipore) were transformed with a total of 2µg of the spacer library. Cells were recovered in SOC while shaking at 37°C for 1 hour. After recovery, cells were split into two biological replicates and used to inoculate 8 mL of Luria Broth (LB) (10g/L tryptone, 5g/L yeast extract, 5g/L NaCl) supplemented with chloramphenicol. Cells were further incubated in LB-chlor while shaking at 37°C for 4 hours, after which the volume was brought to a total of 12mL for each biological replicate. Replicates were split into 6 2mL tubes, and phages of different dilutions were added to each tube. Cells were incubated with phage while shaking at 37°C for 3 hours, after which they were plated on Bioassay plates with LB-agar supplemented chloramphenicol. After 16 hours of growth, all colonies were harvested by scraping and scraped and plasmid DNA was extracted with the NucleoBond Xtra MaxiPrep EF (740424.10, Takara Bio). Spacer sequences were PCR amplified from extracted plasmids and sequenced using a MiSeq (Illumina) with a paired-end 150 cycle kit.

Example 26

Cloning of Libraries and Screening for Beta-lactamase and Transcribed/non-transcribed PFS Screens

[0369] Plasmid libraries for PFS screens were cloned from synthesized oligonucleotides (IDT) consisting of 7 randomized nucleotides either upstream or downstream of the spacer 1

target. To generate dsDNA fragments for cloning, the ssDNA oligonucleotides were annealed to a short primer for second strand synthesis by large Klenow fragment (M0210S, New England Biolabs). dsDNA fragments were Gibson cloned into digested pUC19, either at the 5'-end of the beta -lactamase (Ampicillin resistance) transcript (RNA targeting library), or in a non-transcribed region of pUC19 (DNA targeting library). Gibson clonings were electroporated into Endura Duo electro-competent cells (60240-1, Lucigen) and plated on Bioassay plates with LB-agar supplemented with ampicillin. After 16 hours of growth, all colonies were harvested by scraping and scraped and plasmid DNA was extracted with the NucleoBond Xtra MaxiPrep EF (740424.10, Takara Bio).

[0370] To screen libraries, we co-transformed 50 ng of the pooled ampicillin library and an equimolar amount of the Cas7-11 locus plasmid or pACYC184 plasmid control into NovaBlue(DE3) Gigasingle competent cells (71227-3, EMD Millipore). After transformation, cells were plated on ampicillin and chloramphenicol to select for both plasmids. After 16 hours of growth, all colonies were harvested by scraping and scraped and plasmid DNA was extracted with the NucleoBond Xtra MaxiPrep EF (740424.10, Takara Bio). The target PFS region was PCR amplified and sequenced using a MiSeq (Illumina) with a single-end 150 cycle kit.

Example 27

Computational Analysis of *in vivo* Screens

[0371] To determine enriched spacers from the bacteriophage interference screens, sequenced spacer regions were counted and normalized to total reads for each sample. For enriched spacers, enrichment was measured as the log₂ ratio compared to no phage dilution controls, with a pseudocount adjustment. PFS regions (both at 5' and 3') from spacers above a 1.7 log₂ enrichment threshold in both biological replicates were used to generate sequence logos for the phage dilution samples.

[0372] For transcribed /non-transcribed (beta-lactamase) pUC19 PFS screens, PFS regions were extracted and computationally collapsed to 5nt to have broader coverage. Collapsed PFS were counted and normalized to total reads for each sample. For a given PFS region, enrichment was measured as the log ratio compared to control (pACYC184 control), with a 0.01 pseudocount adjustment. PFSs above a 6 depletion threshold that were enriched in both biological replicates were collected and used to generate sequence logos (Crooks *et al.* 2004).

Example 28

Bacterial Phage Interference Assay for Individual Spacers

[0373] To test individual spacers for MS2 interference with drop plaque assays, complementary oligonucleotides encoding the spacer sequences (see Table 4) flanked by overhangs corresponding to the Cas7-11a Golden Gate acceptor sites were ordered from IDT. Oligonucleotides (final concentration 10M) were annealed in T4 ligase buffer (B0202S, New England Biolabs) supplemented with 5 units of T4 Polynucleotide Kinase (M0201S, New England Biolabs). The oligonucleotides were phosphorylated at 37°C for 30 minutes and annealed by heating to 95°C for 5 minutes followed by a 5°C/minute cool to 25°C. Annealed oligos were then cloned into the locus backbone by Golden Gate cloning. After verification by sequencing, clonal plasmids were transformed into NovaBlue(DE3) GigaSingle competent cells (71227-3, EMD Millipore) made competent with the Mix and Go kit (T3001, Zymo Research). Transformed GigaSingle cells were seeded from an overnight culture grown to OD600 of ~2, at which point they were diluted 1:6 in Top Agar (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl, 7g/L agar) and poured onto warm LB-chloramphenicol plates. Dilutions of MS2 phage in phosphate buffered saline were then spotted on the plates with a multichannel pipette, and plaque formation was recorded after overnight incubation.

Example 29

RFP Targeting and Growth Assay

[0374] Spacers (see Table 4) targeting RFP mRNA were Golden Gate cloned into the DisCas7-11a-only construct (as described above). After verification by sequencing, clonal plasmids were co-transformed with a tetracycline-inducible RFP plasmid on the pBR322 backbone (Addgene 79157). Colonies were picked into TB supplemented with ampicillin, chloramphenicol, and 50 ng/mL anhydrotetracycline final (631310, Takara Bio) and grown to saturation by shaking at 37°C overnight. Cells were separated by centrifugation through a 40µm filter plate (MANMN4010, EMD Millipore) and analyzed by flow cytometry on a FACSCanto II Cell Analyzer (BD Biosciences). Knockdown efficiency was quantified as the percent of RFP positive cells compared to a non-targeting spacer control.

Example 30

Cas7-11 Protein Purification

[0375] The *E. coli* codon-optimized genes for DisCas7-11a and GwCas7-11c, and the human codon-optimized gene for CjcCas7-11b were synthesized by Twist Biosciences, PCR amplified, and Gibson into a Twin-Strep-tag and SUMO tag bacterial expression vector (Addgene 90097). Sequence verified clonal Cas7-11 expression constructs were transformed into either T7 Express lysY Competent *E. coli* cells (C3010I, New England Biolabs), for DisCas7-11a and GwCas7-11c, or Rosetta(DE3)pLysS Competent Cells (70956-3, EMD

Millipore), for CjcCas7-11b. Colonies were picked and grown overnight in TB supplemented with ampicillin. 10 ml of overnight culture was used to inoculate 1L of TB supplemented with ampicillin, and the culture was grown by shaking at 37°C until the OD₆₀₀ reached 1. At this point, cultures were cooled at 4°C for 30 minutes, and induced with IPTG (I6758, Sigma) to a final concentration of 500 µM. Induced cultures were grown overnight (16-24 hours) by shaking at 18°C. After growth, Cells were harvested by centrifuged at 17,700xg for 15 min at 4°C, and the cell pellet was frozen at -80°C for later purification.

[0376] Frozen cell paste was crushed and resuspended to a final concentration of 250mg/mL in lysis Buffer (20 mM Tris-HCl, 500 mM NaCl, 1 mM DTT, pH 8.0) supplemented with Complete Ultra EDTA-free protease inhibitor tablets (5892953001, Sigma), 1mg/mL lysozyme (L6876, Sigma) and 1U/mL Benzonase nuclease (E1014, Sigma). Resuspension was performed by spinning for 30 minutes at 4°C. After resuspension, the solution was sonicated on ice with a 600W Ultrasonic Homogenizer (U.S. Solid) for 15 minutes with a 50% duty cycle. Sonicated lysate was cleared by centrifugation at 38,400xg for 1 hour at 4°C. The supernatant was supplemented with 1mL of Strep-Tactin XT superflow high capacity resin (2-4030-010, IBA Life Sciences) and incubated with rotation for 1 hour at 4°C. Bound resin was loaded onto a glass Econo-Column (7371507, Bio-Rad) and washed with three column volumes of lysis buffer supplemented with Complete Ultra EDTA-free protease inhibitor tablets. To cleave protein off of bead-bound Twin-Strep-tag and SUMO tag, the resin was resuspended in 3mL of SUMO digest buffer (30 mM Tris- HCl, 500 mM NaCl, 1 mM DTT, 0.15% NP-40, pH 8.0) supplemented with 100 ug of SUMO protease (purified in-house). Cleavage occurred on-column overnight at 4°C with rotation, after which supernatant was drained from the column by gravity filtration and concentrated to a volume of 1mL with a 100kDa centrifugal filter (UFC810024, Sigma). Concentrated protein was loaded onto a gel filtration column (Superdex 200 Increase 10/300 GL, GE Healthcare Life Sciences) via an ÄKTA pure FPLC machine. The resulting fractions from gel filtration were analyzed by SDS-PAGE Stain-Free gels (4568096, Bio-Rad), and fractions containing protein were pooled, buffer exchanged into Storage Buffer (600 mM NaCl, 50 mM Tris-HCl pH 7.5, 5% glycerol, 2mM DTT), quantified via comparison SDS-PAGE with a Bovine Serum Albumin (BSA) ladder, aliquoted, and frozen at -80°C for storage.

Example 31

Nucleic Acid Target and Cas7-11 pre-crRNA Preparation

[0377] RNA targets for testing cleavage and pre-crRNA targets were synthesized by *in vitro* transcription. Single stranded DNA oligonucleotide templates with a T7 RNA polymerase

promoter appended at the 5' end were synthesized by IDT, and converted to dsDNA by PCR and gel extraction. Gel extracted dsDNA products (~1µg/reaction) were used as input for the HiScribe T7 Quick High Yield RNA Synthesis kit (E2050S, New England Biolabs), with the 30uL reaction size recommended for transcripts <300nt. Transcription occurred overnight at 30°C, after which targets were purified with RNA Clean and Concentrator columns (R1017, Zymo Research). Labeling was performed with the 3' EndTag End Labeling System (MB-9002, Vector Laboratories) and 5' EndTag End Labeling System (MB-9001, Vector Laboratories), per the manufacturer's instructions, which was followed by incubation with 100x molar ratio of Cyanine5.5 maleimide (17080, Lumiprobe) at 65°C for 30 minutes. We confirmed that this labeling was not 5' or 3' specific, but rather body labeling, via Rnase H gel assays (fig. S21), likely due to the intercalating or groove-binding properties of Cyanine5.5 (Biver *et al.* 2005). After labeling, RNA was purified with RNA Clean and Concentrator columns (R1017, Zymo Research). All targets used for *in vitro* cleavage are listed in supplemental Table 5.

Example 32

Cas7-11 crRNA Preparation

[0378] crRNA designs were ordered from IDT as reverse complement DNA oligonucleotide templates with a T7 RNA polymerase promoter appended at the 5' end. These templates were annealed with a T7 promoter oligonucleotide, generating a dsDNA promoter region. Annealing occurred in Taq Buffer (B9014S, New England Biolabs) with both oligonucleotides at 10µM. 1uL of the anneal reaction was used as input for the HiScribe T7 Quick High Yield RNA Synthesis kit (E2050S, New England Biolabs), with the 30uL reaction size recommended for transcripts <300nt. Transcription occurred overnight at 37°C, after which targets were purified with RNAClean XP beads (A63987, Beckman Coulter); beads were used at 3.3x volume ratio and supplemented with 3x volume of isopropanol. All crRNA used for *in vitro* cleavage are listed in supplemental Table 6.

Example 33

In vitro Nuclease Assays

[0379] Unless otherwise indicated, *in vitro* nuclease assays were performed with 233nM purified Cas7-11, 30nM of labeled ssRNA target and 200 nM crRNA in nuclease assay buffer (40mM Tris- HCl, 60mM NaCl, 6mM MgCl₂, pH 7.5) supplemented with 4U of RNase Inhibitor, Murine (M0314S, New England Biolabs). For pre-crRNA processing reactions, crRNA was omitted and pre-crRNA was used in place of labeled ssRNA target. Reactions

were incubated for 1 hour at 37°C (unless otherwise indicated) and then quenched with addition of proteinase K, EDTA, and urea (final concentrations 1mg/mL proteinase K, 6mM EDTA, and 400uM Urea) for 15 minutes at 50°C. To prepare for gel electrophoresis, reactions were denatured with 4.5M urea denaturing buffer at 95°C for 5 minutes and loaded onto a 10% Novex PAGE TBE-Urea gel (EC6885BOX, Invitrogen), which was run at 235V for 25 minutes at 60°C. Gels were imaged using an Odyssey scanner (LI-COR Biosciences).

Example 34

Experimental and Computational Analysis of *in vitro* PFS Screens

[0380] *In vitro* PFS screens were reverse transcribed with a target specific primer using the qScript cDNA SuperMix (95048-025, Quanta Bio) and, the target PFS region was PCR amplified and sequenced using a MiSeq (Illumina) with a single-end 150 cycle kit. PFS regions were computationally extracted and collapsed to 5nt to have broader coverage, and collapsed PFS were counted and normalized to total reads for each sample. For a given PFS region, enrichment was measured as the log ratio compared to non-targeting control, with a 0.01 pseudocount adjustment. PFSs above a 5 depletion threshold that were enriched in both experimental replicates were collected and used to generate sequence logos

Example 35

In vitro Collateral Activity Assays

[0381] *In vitro* collateral activity assays were performed with 233nM purified Cas7-11 or LwaCas13a, 10nM of unlabeled ssRNA target and 200 nM crRNA in nuclease assay buffer (40mM Tris-HCl, 60mM NaCl, 6mM MgCl₂, pH 7.5) supplemented with 4U of RNase Inhibitor, Murine (M0314S, New England Biolabs) and 250nM final concentration RnaseAlert V2 (4479768, Thermo Fisher Scientific). Reactions were incubated at 37°C and activity was read continuously in the FAM channel of a Synergy Neo2 plate reader (BioTek).

Example 36

Electrophoretic Mobility Shift Assay (EMSA)

[0382] EMSA reactions were performed with dilutions of Cas7-11-crRNA complex with 10nM of labeled ssRNA target in EMSA buffer (5mM EDTA, 2.5% glycerol, 20mM Tris, 5μg/mL heparin, pH 7.5) supplemented with 4U of RNase Inhibitor, Murine (M0314S, New England Biolabs). Binding occurred by incubation at 37°C for 20 minutes, after which reactions were mixed with 2x loading buffer (0.5x TBE buffer, 10% Ficoll, 18mM EDTA) and run on a 6% Novex PAGE TBE gel (EC6265BOX, Thermo Fisher Scientific) at 180V for 30 min at 4°C. Gels were imaged using an Odyssey scanner (LI-COR Biosciences).

Example 37

Next-generation sequencing of *in vitro* cleaved RNA

[0383] *In vitro* nuclease assays were performed and samples were quenched as described above using unlabeled ssRNA targets. After quenching, samples were treated with alkaline phosphatase (EF0651, Thermo Fisher Scientific) and purified with RNA Clean and Concentrator columns (R1017, Zymo Research). Libraries were then treated with the same small RNA sequencing pipeline applied to *in vivo* crRNA species, with the RiboMinus depletion step omitted.

Example 38

Design and cloning of mammalian constructs

[0384] To generate vectors for testing DisCas7-11a in mammalian cells, a mammalian codon optimized DisCas7-11a sequence was ordered from Twist Biosciences, PCR amplified, and cloned into a mammalian expression vector containing combinations of modifications, including with and without NLS tags or N- or C-terminal msfGFP fusions, all under the control of a CMV promoter.

[0385] The dual luciferase reporter was cloned by PCR amplifying *Gaussia* and *Cypridinia* luciferase coding DNA, the EF1alpha and CMV promoters and assembled using the NEB Gibson Assembly (E2611S, New England Biolabs).

[0386] For mammalian DisCas7-11a guide expression, we synthesized both the full-length and mature DR sequences and cloned with golden-gate acceptor sites under U6 expression via restriction digest cloning. Individual guides were then cloned into the corresponding expression backbones for DisCas7-11a by golden-gate cloning. All mammalian DisCas7-11a plasmids are listed in Table 7. All DisCas7-11a guide sequences for knockdown or RNA editing experiments are listed in Table 8.

Example 39

Mammalian cell culture

[0387] All mammalian experiments were performed using the HEK293FT cell line (American Type Culture Collection (ATCC)). HEK293FT cells were grown in Dulbecco's Modified Eagle Medium with high glucose, sodium pyruvate, and GlutaMAX (Thermo Fisher Scientific), additionally supplemented with 1× penicillin–streptomycin (Thermo Fisher Scientific) and 10% fetal bovine serum (VWR Seradigm). We maintained cells below a confluency of 80%.

[0388] Transfections were performed using Lipofectamine 2000 (Thermo Fisher Scientific) in 96-well plates coated with poly-D-lysine (BD Biocoat). For transfections, cells were plated 16 hours prior to transfection at seeding densities of ~20,000-30,000 cells per well, allowing cells to reach 90% confluency by transfection. For each well on the plate, transfection

plasmids were combined with Opti-MEM I Reduced Serum Medium (Thermo Fisher) to a total of 25 μ l. Separately, 24.5 μ l of Opti-MEM was combined with 0.5 μ l of Lipofectamine 2000. Plasmid and Lipofectamine solutions were then combined and incubated for 5-10 minutes, after which they were pipetted onto cells.

5

Example 40

Mammalian cell RNA knockdown assays

[0389] To assess RNA knockdown in mammalian cells with reporter constructs, 100 ng of DisCas7-11a expression vector was co-transfected with 150 ng of guide expression plasmid and 40 ng of the knockdown reporter construct. After 48 hours, we harvested media
10 containing secreted luciferase and measured luciferase activity using the Gaussia Luciferase Assay reagent (GAR-2B) (Targeting Systems) and Cypridina (Vargula) luciferase assay reagent (VLAR-2) (Targeting Systems) kits. Assays were performed in white 96 well plates on a plate reader (Biotek Synergy Neo 2) with an injection protocol. All replicates performed are biological replicates. Luciferase measurements were normalized by dividing the Gluc
15 values by the Cluc values, thus normalizing for variation from well to well.

[0390] For targeting of endogenous genes, 100 ng of DisCas7-11a expression vector was co-transfected with 150 ng of guide expression plasmid. After 48 hours, we lysed cells and harvested RNA using a method previously described (Joung *et al.* 2017) with a gene-specific reverse transcription primers. Using Fast Advanced Master Mix (Thermo Fisher Scientific),
20 we measured gene expression using the cDNA via qPCR and TaqMan qPCR probes for the KRAS, PPIB, CXCR4, and MALAT1 transcripts (Thermo Fisher Scientific) as well as the GAPDH control probe (Thermo Fisher Scientific). qPCR reactions were read out on a Bio-Rad CFX384 Touch Real-Time PCR Detection System, with four 5 μ l technical replicates in 384-well format.

25

Example 41

RNA editing in mammalian cells

[0391] To assess dDisCas7-11a-ADAR2dd editing activity in mammalian cells, we transfected 100 ng of dDisCas7-11a-ADAR2dd expression vector, 150 ng of guide expression plasmid, and 25 ng of the RNA editing reporter. After 48 hours, we harvested
30 media from the cells, which should contain the secreted Gluc (normalization control) and Cluc (protein targeted for correction of W85X) proteins. If planning to sequence the RNA for the precise RNA editing levels, we also harvested RNA using a method previously described (Joung *et al.* 2017) with a gene-specific reverse transcription primer.

[0392] For luciferase measurements, we used undiluted media and the Gaussia Luciferase Assay reagent (GAR-2B) (Targeting Systems) and Cypridina (Vargula) luciferase assay reagent (VLAR-2) (Targeting Systems) kits. Assays were performed in white 96 well plates on a plate reader (Biotek Synergy Neo 2) with an injection protocol. All replicates performed are biological replicates. Luciferase measurements were normalized by dividing the Cluc values by the Gluc values, thus normalizing for variation from well to well. This ratio measurement could further be normalized by dividing by the non-targeting guide ratios.

[0393] For sequencing of RNA editing rates, we performed two rounds of PCR (NEBNext High-Fidelity 2X PCR Master Mix from New England Biolabs) using the extracted cDNA to add Illumina adaptors and sample barcodes. Illumina next generation sequencing was used to sequence editing rates in our prepared libraries on a MiSeq instrument. RNA editing rates were evaluated at the targeted adenosine in the W85X pre-termination codon.

Example 42

DisCas7-11-mediated RNA-guided RNA interference against MS2 phage and RNA knockdown in *E. coli*

[0394] DisCas7-11-mediated RNA-guided RNA interference against MS2 phage and RNA knockdown in *E. coli* was investigated. The potential role of PFS in the function of the III-E systems was assessed by performing randomized PFS screens on both a DNA-only target and an expressed target at the 5' end of an ampicillin gene. It was found that there was no signal on the DNA-only target on either side of the protospacer (FIG. 28A) and only a weak sequence preference when cleaving the expressed target (FIG. 28B). Furthermore, it was found that there were no observable trends in the flanking sequences of enriched MS2 spacers (FIGS. 28C and 28D), suggesting a lack of targeting constraints.

Example 43

Regulation of DisCas7-11 interference activity by accessory proteins

[0395] The regulation of DisCas7-11 interference activity by accessory proteins was assessed. Whether accessory proteins in the *D. ishimotonii* subtype III-E CRISPR locus could modulate RNA interference activity was investigated (FIG. 28E). Expression of the full *D. ishimotonii* CRISPR locus resulted in reduced MS2 interference compared to DisCas7-11 alone (FIG. 28F), prompting us to test the deletion of each putative accessory gene. Deletion of Csx29 was found to restore the interference activity. (FIG. 28G).

Example 44

Mapping DisCas7-11 a Cleavage Mechanism

[0396] The mapping of the DisCas7-11a cleavage mechanism via in vitro cleavage of ssRNA of a 31nt target was investigated. The DisCas7-11a cleavage of synthetic 31nt MS2 ssRNA with a 31nt crRNA completely duplexed to the target shows two cleavage fragments that are generated in the targeting condition when protein is generated (indicated by red triangles) (FIG. 29). Other bands present in all conditions indicate incomplete products generated when in vitro transcribing the 31nt target.

Example 45

DisCas7-11a-mediated knockdown of mRNA in mammalian cells

[0397] The effects of DisCas7-11a-mediated knockdown of mRNA in mammalian cells by guide variants or inactivating mutations was measured. The knockdown of Gaussia luciferase (Gluc) mRNA in mammalian cells by DisCas7-11a unmodified or with N-/C- terminal msfGFP fusions, with guides containing the full DR sequence was assessed. The guides are designed to be tiled across the Gluc transcript. The knockdown of Gluc mRNA in mammalian cells by DisCas7-11a unmodified or with N-/C- terminal msfGFP fusion, with guides containing the mature DR sequence was assessed. Guides are designed to be tiled across the Gluc transcript. The comparison of knockdown activity of Gluc mRNA in mammalian cells between active DisCas7-11a, catalytically inactive D429A/D654A DisCas7-11a, and GFP is illustrated in FIG 31A. Guides are the same as in a and b above and have the full DR sequence. The comparison of knockdown activity of endogenous mRNA in mammalian cells between active DisCas7-11a, catalytically inactive D429A/D654A DisCas7-11a, and GFP is illustrated in FIG. 31B. Here the guides have the full DR sequence.

Example 46

RNA editing efficiency with DisCas7-11a-NES-ADAR2 constructs

[0398] The RNA editing efficiency with active and inactive DisCas7-11a-NES-ADAR2 constructs was measured. The RNA A-to-I editing of Cypridinia luciferase (cluc) mRNA W85X mutation in mammalian cells by active DisCas7-11a-NES-ADAR2 or dead DisCas7-11a-NES-ADAR2 is shown in FIG. 32A. The guides are designed with mismatch distances between 2-50 nt. Editing is measured by restoration of Cluc luciferase activity and is normalized to the non-targeting guide condition. The RNA A-to-I editing of Cypridinia luciferase (cluc) mRNA W85X mutation in mammalian cells by active DisCas7-11a-NES-ADAR2 or dead DisCas7-11a-NES-ADAR2 is shown in FIG. 32B. The guides are designed with mismatch distances between 2-50 nt, and the editing is measured by amplicon sequencing and measuring the percent correction of the pre-termination stop codon.

Example 47

RNase H cleavage of target showing body labeling

[0399] The RNase H cleavage of target showing body labeling was investigated. Cleavage of 3' (FIG. 33A) and 5' (FIG. 33B) labeled targeted by incubation with RNase H and indicated oligonucleotides show similar band patterns and double bands indicative of body labeling.

5

Example 48

Mechanism of the programmable RNA cleavage activity of DisCas7-11

[0400] The mechanism of the programmable RNA cleavage activity of DisCas7-11 was assessed. The type III-E system from *D. ishimotonii* and corresponding effector DisCas7-11 were chosen, as a representative of type III-E systems to characterize in detail. The small RNA from *D. ishimotonii* were harvested and sequenced. Mature crRNA species indicative of pre-crRNA processing were identified (FIG. 2C), a hallmark of an active CRISPR-Cas locus. Heterologous reconstitution of the *D. ishimotonii* type III-E locus in *E. coli* demonstrated similar processing (FIG. 9A), and expression of the DisCas7-11 protein and pre-crRNA were both necessary and sufficient for mature crRNA expression (FIG. 9B and FIG. 9C), suggesting that Cas7-11 possesses pre-crRNA processing activity. Processing occurred though cleavage in the second hairpin of the direct repeat (DR) at residue 21, generating a mature rRNA with spacers of ~18-24 nt and 14 nt DRs lacking any predicted secondary structure, similar to the mature DR tags of other type III systems (FIG. 2D). To generalize the observation of the processing activity of Cas7-11 beyond *D. ishimotonii*, the *Candidatus* *Jettenia caeni* type III-E locus through heterologous expression in *E. coli* was assessed. Robust processing of the *C. Jettenia caeni* pre-crRNA that resulted in mature crRNAs 37 nt in length was observed (FIG. 20), suggesting that pre-crRNA processing activity is common among type III-E systems.

10

15

20

25

[0401] One skilled in the art will appreciate further features and advantages of the invention based on the above-described embodiments. Accordingly, the invention is not to be limited by what has been particularly shown and described, except as indicated by the appended claims. All publications and references cited herein are expressly incorporated herein by reference in their entirety.

What is claimed is:

1. A composition that cleaves an RNA target comprising a guide RNA that specifically hybridizes to the RNA target, and a polypeptide comprising an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-32 and 359.
2. The composition of claim 1, wherein the amino acid sequence is SEQ ID NO: 1.
3. The composition of claim 2, wherein the polypeptide comprises a glutamate at a position corresponding to position 70 of SEQ ID NO:1; aspartate at a position corresponding to position 429 of SEQ ID NO:1; aspartate at a position corresponding to position 487 of SEQ ID NO:1; and/or an aspartate at a position corresponding to position 654 of SEQ ID NO:1.
4. The composition of claims 2 or 3, wherein the polypeptide comprises an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO:1
5. The composition of claims 4, wherein the polypeptide comprises an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO:1.
6. The composition of claim 5, wherein the polypeptide comprises an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO:1.
7. The composition of claim 6, wherein the polypeptide is the amino acid sequence of SEQ ID NO:1.
8. The composition of claim 1, wherein the guide RNA comprises a mismatch distance that is 20-65% of the length of the guide.
9. The composition of claim 1, wherein the guide RNA has a sequence with a length of from about 20 to about 53 nucleotides (nt), preferably from about 25 to about 53 nt, more preferably from about 29 to about 53 nt or from about 40 to about 50 nt.
10. The composition of claim 1, wherein the guide RNA is a pre-crRNA.
11. The composition of claim 1, wherein the guide RNA is a mature crRNA.
12. The composition of claim 1, wherein the RNA target is a single-strand RNA (ssRNA).
13. The composition of claim 1, wherein the RNA target is in a cell.
14. The composition of claim 13, wherein the cell is a prokaryotic cell.
15. The composition of claim 13, wherein the cell is a eukaryotic cell.
16. The composition of claim 15, wherein the eukaryotic cell is a mammalian cell.
17. The composition of claim 16, wherein the mammalian cell is a human cell.
18. The composition of claim 1, wherein the polypeptide further comprises a deaminase domain for programmable RNA base editing.

19. The composition of claim 18, wherein the deaminase is selected from the group consisting of an adenosine deaminase, a cytidine deaminase, and a catalytic domain thereof.
20. The composition of claim 18, wherein the deaminase is an adenosine deaminase.
- 5 21. The composition of claim 18, wherein the deaminase domain comprises an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 25-31.
22. The composition of claim 18, wherein the deaminase domain comprises an amino acid sequence 85% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 32-42.
- 10 23. The composition of claim 1, wherein the guide RNA comprises a mismatch that is about 20 to about 30 nucleotides from the non-pairing C of the guide RNA.
24. A nucleic acid molecule encoding the polypeptide of any of the preceding claims.
25. The nucleic acid molecule of claim 24, wherein the nucleic acid molecule encodes the guide RNA.
- 15 26. The nucleic acid molecule of claim 24, further comprising nucleic acid molecule that encodes the guide RNA.
27. A vector comprising the nucleic acid molecule of claims 24-26.
28. A cell comprising the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
- 20 29. The cell of claim 28, wherein the cell is a prokaryotic cell.
30. The cell of claim 28, wherein the cell is a eukaryotic cell.
31. The cell of claim 30, wherein the eukaryotic cell is a mammalian cell.
32. The cell of claim 31, wherein the mammalian cell is a human cell.
- 25 33. A method of cleaving an RNA target in a cell comprising providing to the cell the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
34. A method of stabilizing an RNA target in a cell comprising providing to the cell the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
- 30 35. A method of affecting translation of an RNA target in a cell comprising providing to the cell the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.

36. The method of any of claims 33-35, wherein the RNA target is an ssRNA.
37. A method of treating a genetically inherited disease in a subject in need thereof comprising administering to the subject an effective amount of the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27,
 5 wherein the genetically inherited disease involved a guanosine to adenosine change in the genome of the subject.
38. The method of claim 37, wherein the genetically inherited disease is selected from the group consisting of Meier-Gorlin syndrome; Seckel syndrome 4; Joubert syndrome 5; Leber congenital amaurosis 10; Charcot-Marie-Tooth disease, type 2;
 10 leukoencephalopathy; Usher syndrome, type 2C; spinocerebellar ataxia 28; glycogen storage disease type III; primary hyperoxaluria, type I; long QT syndrome 2; Sjögren-Larsson syndrome; hereditary fructosuria; neuroblastoma; amyotrophic lateral sclerosis type 9; Kallmann syndrome 1; limb-girdle muscular dystrophy, type 2L; familial adenomatous polyposis 1; familial type 3 hyperlipoproteinemia; Alzheimer's
 15 disease, type 1; metachromatic leukodystrophy; and cancer.
39. A method of treating a genetically inherited disease in a subject in need thereof comprising administering to the subject an effective amount of the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27, wherein the genetically inherited disease is a pre-termination disease.
- 20 40. A method of altering splicing of a pre-mRNA in a cell comprising administering to the cell an effective amount of the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
41. A method of changing microRNA targets in a subject in need thereof comprising administering to the subject an effective amount of the composition of claims 1-23,
 25 the nucleic acid molecule of claims 24-26 or the vector of claim 27.
42. A method of increasing RNA stability in a cell comprising administering to the cell an effective amount of the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
43. A method of modulating translation in a cell comprising administering to the cell an
 30 effective amount of the composition of claims 1-23, the nucleic acid molecule of claims 24-26 or the vector of claim 27.
44. A method of detecting a bacterium or derivative thereof in a sample, the method comprising:

adding to the sample an effective amount of the composition of claims 1-23,
the nucleic acid molecule of claims 24-26 or the vector of claim 27; and

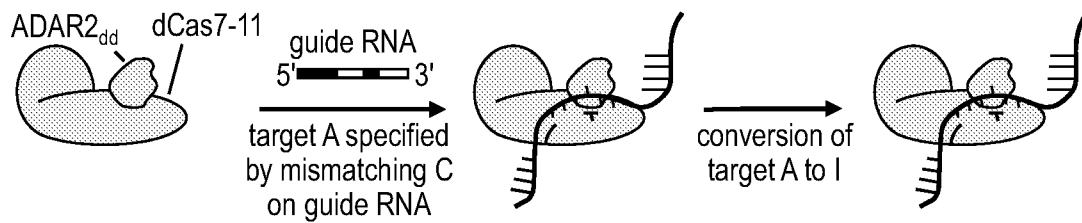
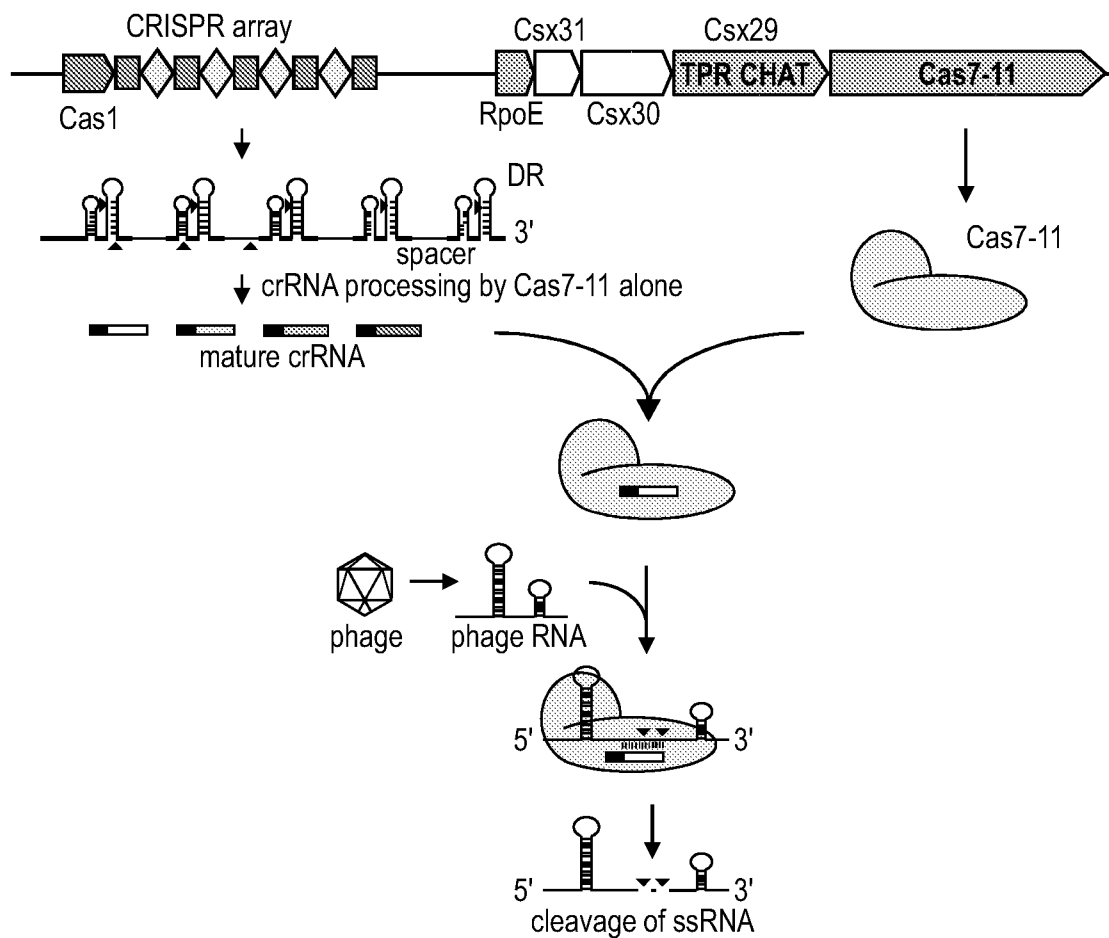
detecting a reporter specific to the bacterium or derivative thereof.

45. A method of detecting a virus or derivative thereof in a sample, the method
5 comprising:

adding to the sample an effective amount of the composition of claims 1-23,
the nucleic acid molecule of claims 24-26 or the vector of claim 27; and

detecting a reporter specific to the virus or derivative thereof.

1/142

**FIG. 1A****FIG. 1B**

2/142

	type III-A/B complex	Cas7-11	Cas13
Property			
Complex identity	multi component	single effector	single effector
processing activity	accessory protein (Cas6)	in effector	in effector
Rnase/Dnase	Rnase and DNase activity	RNase only	RNase only
mammalian activity	no	yes	yes
DR structure	short, unstructured	short, unstructured	hairpin
Cleavage mechanism	multiple cuts <i>in cis</i>	multiple cuts <i>in cis</i>	collateral activity <i>in trans</i>
catalytic domains	RAMP (acidic)	RAMP (acidic)	HEPN (basic)
Sequence constraint	no PFS	no PFS	PFS constraint

FIG. 1C

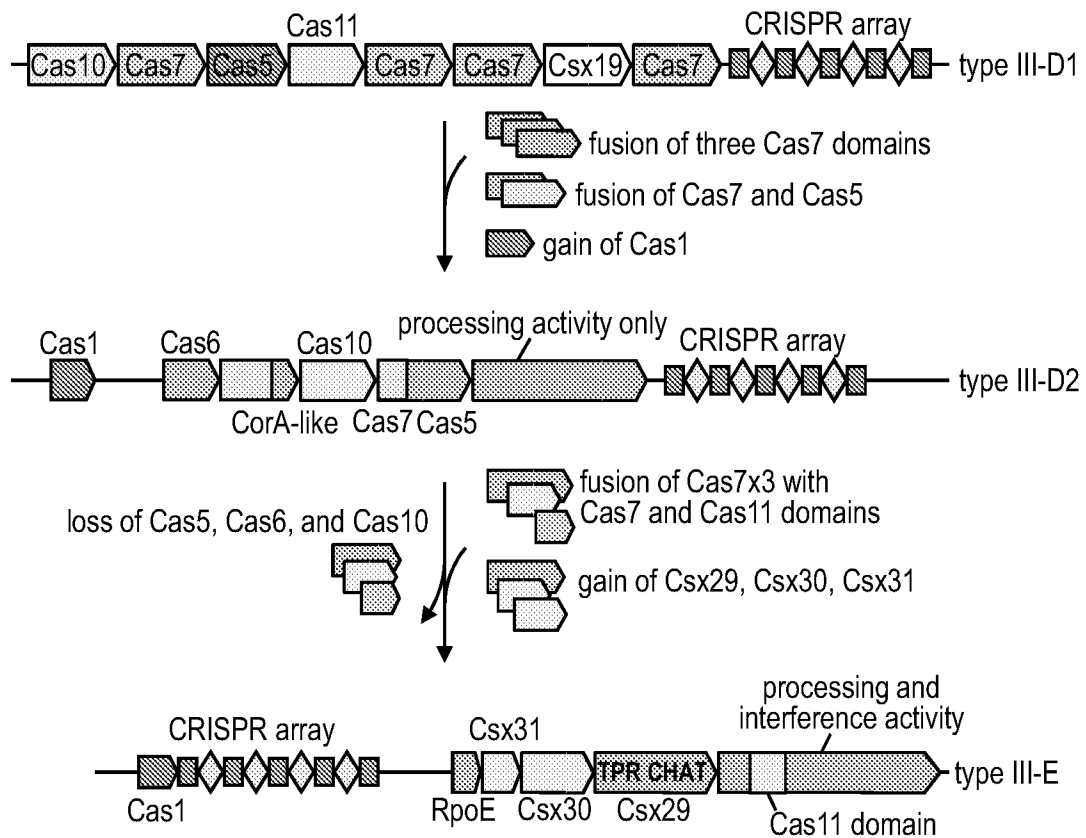


FIG. 1D

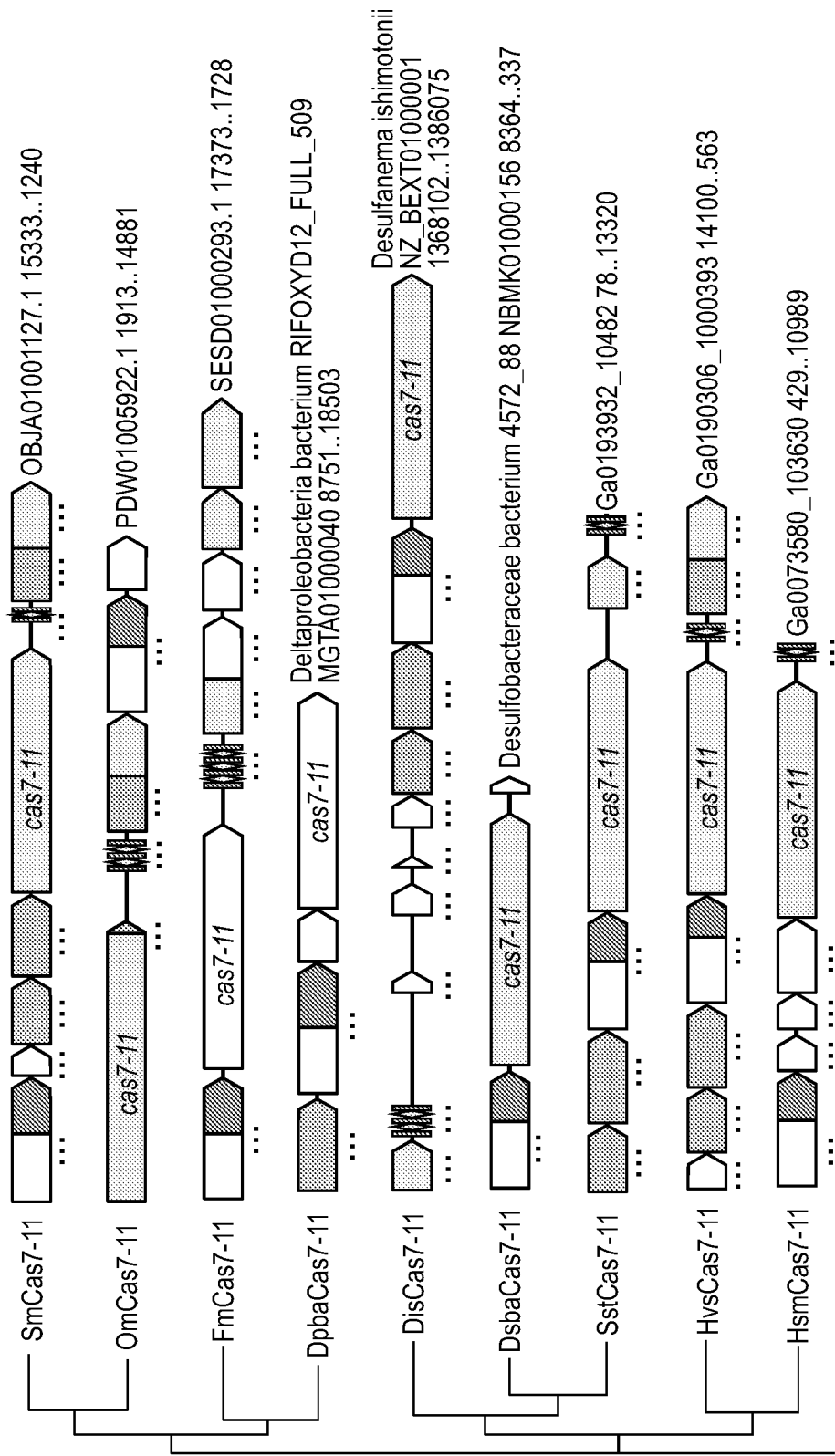


FIG. 1E

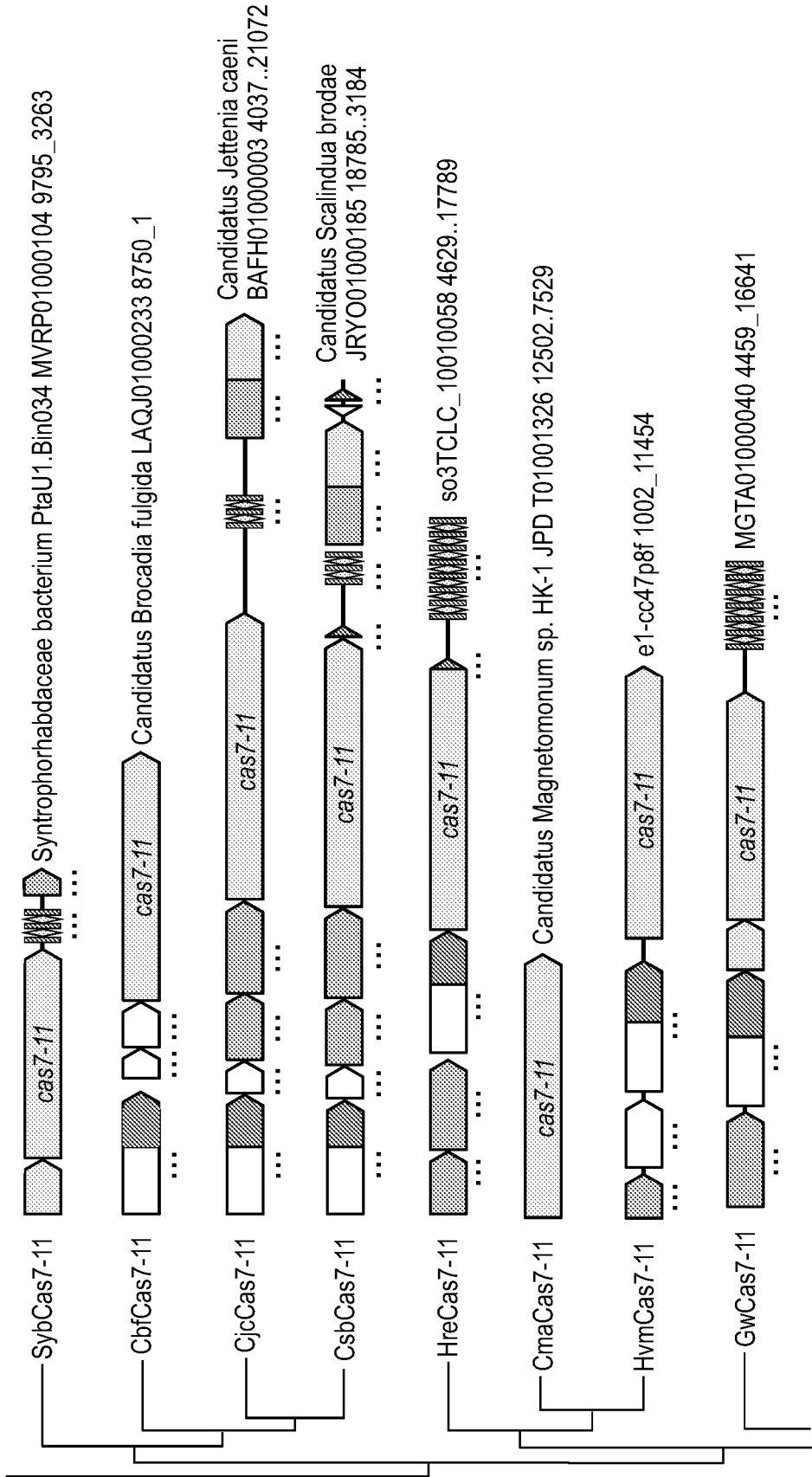


FIG. 1E (cont.)

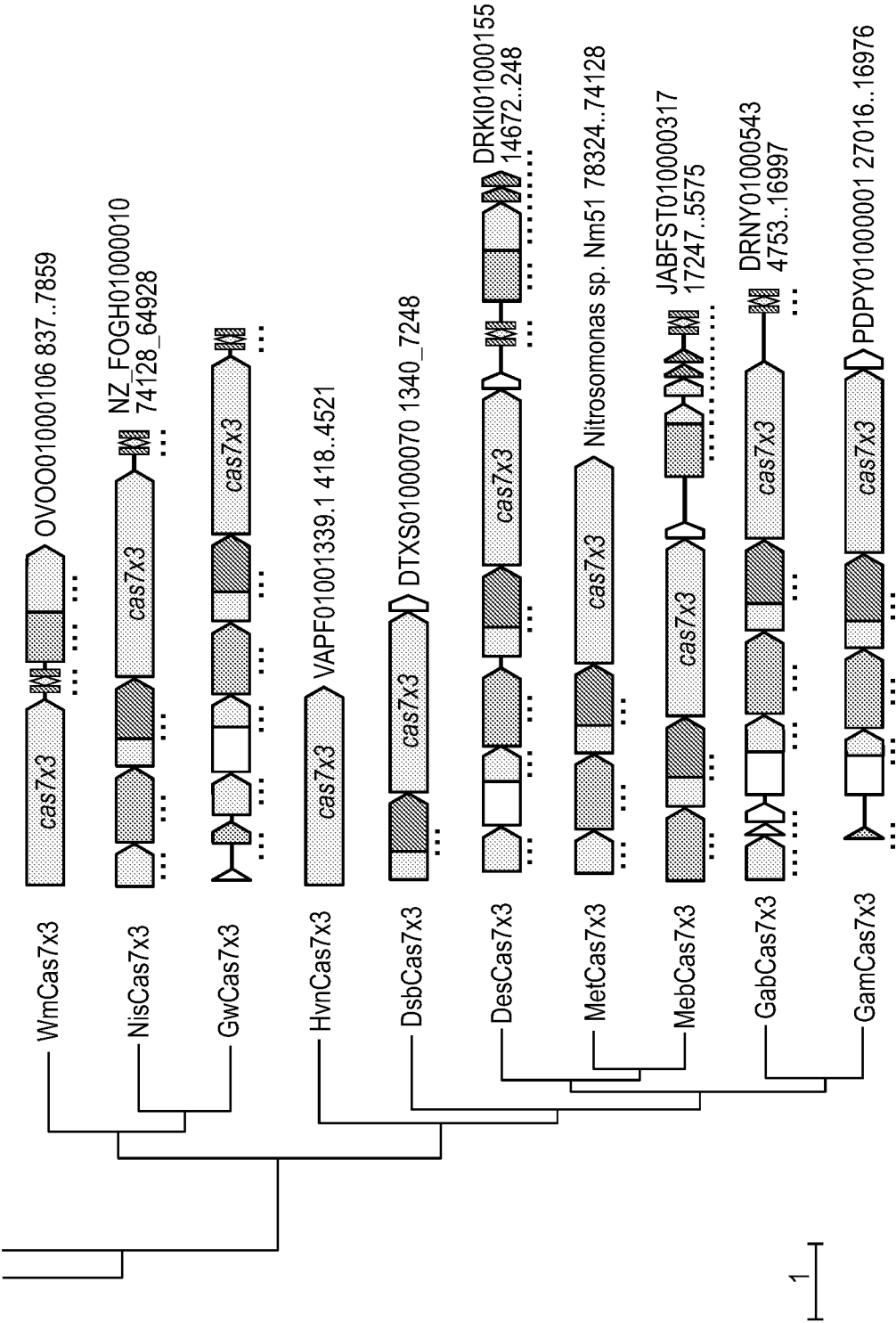


FIG. 1E (cont.)

MBD3181016.1	RDMATSAHTTIDKETKRTKEHLINIEVVRIEDN-RYLKGS	EMIN-
MBS8521845.1	DQVRTSAHNKINRATGRVEKDLLFSYELG--ANT-VPLEGTIERVNP	
RCV63412.1	DMVRTSTHNKISRATGRVEKDHLFSYELG--ANT-TPFEGMIEPVNH	
MBC8461765.1	GTTRVSLHNRVDPTTGRVPEETFFSFEVG--SAE-K-FTGRTERS	GP
HHR29383.1	EPIRISTHNRIDPATGRVPKDLLFTFEMG--SPA-V-FRCYIERIGI	
WP_129889723.1	LQTFLYTQARIDRPRTAVPGALYSSEFG--LRE-ITFKGS	TGWLE
WP_174676620.1	NQKDVYTQVRIDRPRTAVEGALYTSEFG--TRG-LTFHGDITGSLT	
TFH44231.1	LFLNTRTGTKIDRRRTGTQEDHLFTTESG--IKG-ITLTSKIYGRGN	
MAT45258.1	YGQSIRSNRNPRTAANDHLFFTECV--DPD-TYFICEIRSSSN	
WP_075063448.1	YGRLLRTGIGIDRKLRVVPDRLFTNECV--GTK-IPFTGRIGEFP	
HreCas7-11	EIAVSRVVRVDPNSGKA-KDFFRVWEID--HKLCPNFICKMSISLS	
CmaCas7-11	DIVQKRILNRVDQTCGKA-KDFFKVCQVD--HIACTLNGIIRINDE	
HvmCas7-11	ETAIERIVNRVDPSSGKA-KDYMRIWEID--PLVCSQFNGLITIN--	
SmCas7-11	EIGSRRTINRVDDFTGKA-HDFFSIWEVD--A--VREFQGEIVLAAD	
OmCas7-11	EIGRLRTINRIDRLTTKA-QDFFRIYEVD--Q--VRDFFGTLTLAGD	
DisCas7-11	AIGSORVINRVDFKSGKA-HDFFKAYEVD--HTRFPRFEGETIDNK	
DsbaCas7-11	EIAVRRVVRVDPYASGKA-HDFFRIFEVD--HIAFPCHFGEIAFGEN	
SstCas7-11	DIGTQRTINRVDPYATGKA-HDFFKVCWEID--HSLLSVFQCKISADN	
CsbCas7-11	DIASGRILNRVDFDTGKA-KDYFRTWEAD--YETYGTYTGRITLRNE	
CjcCas7-11	DIASERILNRVDYTTGKA-HDYFKVWEVD--DDQWQQTGTITMHDD	
CbfCas7-11	DVCKERVILNRVDFRTGKA-QDYFRVFEID--HEDWGVYTGELTITEP	
HvsCas7-11	ELAPKRAINRIEQYTGKA-QDYFTVYEAL--NKEFWTFKGRIRIKED	
HsmCas7-11	EIGIRRIILNRVDFTTGKA-QDYFYVWEVE--HSMCPKFQGT	KINED
FmCas7-11	DIAQKRTSNRVDFFSKKA-HDHYGVWEVT--A--VKNLIGYTIYISDA	
DpbaCas7-11	QLIRPRIILNRMDPGTSKA-RDYFRVFEIE--NQLCSQFRGWLWLSGD	

FIG. 1F

WmCas7x3	QRI-----TPDAPDADP-YWQADIALDADGRPYLPG
NisCas7x3	EEMDKQSG-----ESVDKKQNN-SWIQAIALDLNKKPYIPG
GwCas7x3	E-----KERDENKET-RWLEAVALDHKGQPYIPG
HvmCas7x3	VTRP-----ELTVADRDELVDINAVTDTYTKPYLPG
DsbCas7x3	TDRPGLEI-----EQKDGSKVK-AEINAFIKDSNGKPYLPG
GamCas7x3	IPVT--I-----KDTQGNRE-VEVNSVITGKAALPIIPG
DesCas7x3	VEHD-----LIKNDGTP-VQINALITGAGGLPIIPG
GabCas7x3	AEIE--I-----NDSQGDRRQ-VQANAIIRGKDDKPIIPG
MetCas7x3	CEHD-----DVKNNDGEP-VKINACIKGSKGRPIIPG
MebCas7x3	CYRP-----ELTNADQKP-VDINACIKGANLPIIPG
GwCas7-11	-----EGQKQTDQAE-SLHLRTIRDGHGRFRIPF
HreCas7-11	-----SEGQTS-----QAIILCPDGSYRIPR
CmaCas7-11	-----KTSHTR-----STIFLNMGQFCIPR
HvmCas7-11	-----QEGQTS-----ATILNIRDGYFRIPR
SmCas7-11	-----PIDDDQTS-----SRTIVDRDGRYRIPR
OmCas7-11	-----GSHDHTS-----RKILLTRDFYRIPR
DisCas7-11	-----EDAKQTD-----LQVLLTPDNKYRIPR
DsbaCas7-11	-----KEKKQTD-----LMLLLDGQNHYRIPR
SstCas7-11	-----EDAKHTN-----LKVLLDRQNRYPRI
SybCas7-11	-----GAGGHTD-----LSIILLGKDGHYRVPR
CsbCas7-11	SDSIPGKEK-----KSEDSLIVINEHTS-----FNIILLDKENRYRIPR
CjcCas7-11	FDSTQDDLDLVPDIVNTDEKLEANEQTS-----FRIIMDKKGRYRIPR
CbfCas7-11	-----EQTEEVEHTS-----LRLVMDKKGRFRIPR
HvsCas7-11	-----DEEKAE-----GAILLTPDNRFRIPR
HsmCas7-11	-----KTEEQID-----MQIILLTKDGRYRIPR
FmCas7-11	-----QAGKQTS-----MPTLQDSNDHFRIPR
DpbaCas7-11	-----ATEHTD-----LPILLTSDRRHFRIPR

FIG. 1G

WmCas7x3	HRKPQBNAPDGTPRQRGDRA-----LLPGASIRGRIRS
Ni sCas7x3	KKDEAKNEADAKPRTNHQGV-----ILPASSIRGRIRA
GwCas7x3	KADDADAV---PRRTHDDKI-----VLPASSIRGRIRT
HvmCas7x3	KKKRRSNTPNLRPLRDAIGRP-----CIPESSVRGALRA
DsbCas7x3	KKRKED--IDHQPLRDSAGNA-----RLPAKSIRGAMRS
GamCas7x3	DKCKAEDTPDIYPLEEKNGVP-----AFPVRSFRGAIRS
DesCas7x3	NTPDITQAPDMVPLVDEGPN-----MLPASSFRGALRA
GabCas7x3	SKGDGBQPAVHQPLTDRSNNP-----ILPARSFRGALRA
MetCas7x3	LSSNEAKTDHYPLLDKRNRP-----YLPVSSFRGVIRS
MebCas7x3	LEADPKTKIDHYPLLDNHKKP-----RLPASIRGVIRS
GwCas7-11	ESDAPBNVAYKKPVVQYDETGRLRTDPPGVEMLTCLKGEGVRGVWAY
HreCas7-11	DHDNKDAVMVQKTVLFDVDSG-----NYSQMPHHFIKSGIRGACRF
CmaCas7-11	DNRNPDAVMVKKTILVYEQDSSSTHKNPKEVPKY-FIKSETIRGLIRS
HvmCas7-11	DPRNTBAIMVRKTVFCPD PNA--KNRPAPATVY-MIKGESIRGLIRS
SmCas7-11	NSDTABIIISFRRTVVDNGEVL-----REP-----VLRGEGIRGLIRT
OmCas7-11	-RHEEDSVYFQKRIFTSDGRV-----VLVP-----AIRGEGIRGLIRT
DisCas7-11	DKRGTDVTVFKYKAEGEEAK-----PVC-----AYKAESFRGVIRS
DsbaCas7-11	TEDVADIVSFKKYTQGGEKITY-----AYKSESFRGVVRT
SstCas7-11	EKDGSDIVSFRKYADDGSGEVY-----AYKAESFRGVVRA
SybCas7-11	ERIGFDNIAYEKRRYNGETNT-----TESIP-----AVKGETFRGVVRT
CsbCas7-11	EPGNRDAIAYKKRVYNDGNNA-----IEPEPRF-AVKSETHRGIFRT
CjcCas7-11	NKDNIDCIAYEKKRWENGKIK-----FVP-----TIKGETIRGVIRM
CbfCas7-11	DPNNVDAIVFEKMKLDGDQVK-----YLP-----AIKGETIRGVIRT
HvsCas7-11	EGKAPDAVFFKKYVFFENGKIE-----EKP-----CFKAESIRGIFRT
HsmCas7-11	DSSNTBLVTFKKYKLEESKEVF-----AIKGESIRGVVRT
FmCas7-11	-HGNTDSVFYKKPILKSKEKE-----PSYQW-----AIKSDTVRGLIRS
DpbaCas7-11	-RNESDAVFYQKS VAGEKGPVY-----AIKGEGLRGIVSS

FIG. 1H

9/142

WmCas7x3	PATLV	HDM	AI	DR	FT	GC	K	GAK	FK	---	RYA	EC	P	---	TL	EC	QL	S	DL
NisCas7x3	LATLR	HEM	VA	ID	DR	FT	GC	K	GAK	FK	---	DYI	EC	P	---	TL	TA	AI	Y
GwCas7x3	SIKTR	RRHEM	AI	DR	FT	GC	K	GAK	FK	---	DYV	EC	P	---	TL	AC	KL	S	DL
HvmCas7x3	EDCD	NIQE	VA	ID	DR	FT	GC	K	KAK	FK	---	EYI	GS	P	---	RF	GT	IA	DK
DsbCas7x3	TRQET	VQDF	VA	ID	DR	FT	GC	K	TAK	FD	---	SFS	WR	P	---	QY	S	LMH	PS
GamCas7x3	LKPF	IIQE	VA	ID	DR	FT	GC	K	EAK	FK	---	AHY	QA	P	---	VF	KG	KVR	SQ
DesCas7x3	FRREQ	ETFE	VA	ID	DR	FT	GC	K	GAL	T	---	RHA	ES	P	---	RF	EC	HL	DP
GabCas7x3	YRPA	KTQ	QF	VA	ID	DR	FT	GC	K	GAL	FS	---	KYF	ER	P	---	VL	KG	IS
MetCas7x3	KSKKT	KQD	F	VA	ID	DR	FT	GC	K	GAK	FD	---	THF	ER	P	---	EF	EC	AI
MebCas7x3	ANEL	KTQE	F	VA	ID	DR	FT	GC	K	GAK	FK	---	KHS	ER	P	---	YF	QC	RI
GwCas7-11	TPHAM	RS	DR	VA	ID	DR	FT	GC	K	GAK	FD	---	AS	PK	P	---	LN	FK	ST
HreCas7-11	EVEA	IKCD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	DD	Y	P	---	LP	GS	PN
CmaCas7-11	NVSD	CCID	H	VA	ID	DR	FT	GC	K	GAK	FD	---	DY	P	---	LS	AS	PK	NC
HvmCas7-11	SVSD	KKMD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	DD	Y	P	---	LP	GS	CA
SmCas7-11	TVAD	KRLD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---	---
OmCas7-11	TWND	KKID	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---	---
DisCas7-11	DPEP	VT	FD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
DsbaCas7-11	EPEP	RR	FD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
SstCas7-11	QPEP	MI	FD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
SybCas7-11	AWTR	KHL	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
CsbCas7-11	EKLE	KHL	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
CjcCas7-11	SYKK	KL	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
CbfCas7-11	KDIA	KK	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
HvsCas7-11	GQEE	KFF	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
HsmCas7-11	DSAP	KRL	D	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
FmCas7-11	EIKT	YR	MD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---
DpbaCas7-11	EVRE	KL	FD	H	VA	ID	DR	FT	GC	K	GAK	FD	---	K	---	---	---	---	---

FIG. 11

WmCas7x3	PVAG A QNQPAIP TS RGL S FES GSNIR LHPTPYSIR TT
NisCas7x3	EARP K NGKHAIP TS RGM S FES NSNFR LHPEHYSVR SL
GwCas7x3	SLHP K NNGLAIP TS RGM S FES NSNFR LD KTYSMR TM
HvmCas7x3	EILP T GGKPAIP TS RGM S AE A NSSIR LE KTLSYR SM
DsbCas7x3	QVDH TENGEIAP TT RGL S SE A NSSMR LD GMSYRQPV
GamCas7x3	LRHP Q DGEPAIP STS RGL S MTE ANCAMR LDSEIISYR PM
DesCas7x3	RIEN R GNRIAP AS RGM S AE A NSAMR LH GILSYR KA
GabCas7x3	LKEH Q NNKLAIP TS RGL S AE A NSAIR LD GVLSYR PA
MetCas7x3	EINN R NGELAIP TS RGM S AE A NSAMR LD GLLSYR TA
MebCas7x3	ELDN R NGQLAIP TS RGM S AE A NSAMR LD GLLSYR DA
GwCas7-11	TQRFHQ NDEIG P AS RGM L NYQ NSCYRNLKATEEITR MP
HreCas7-11	NYRF R NDELAIP SE RGM S YE A NSCFR ME GRYLSR MG
CmaCas7-11	NYAF R NDHIAIP AS RGM S FET HSCFR MD KKYLTR VI
HvmCas7-11	NYGF R NGNVAIP SS RGM S FE A NSCFR FD ERYLSRSEK
SmCas7-11	NFPF K NDEIM P AP W A SQ YE A NSCFR MK KRFLSW ME
OmCas7-11	EYPS R NNTPMIP AG R A SQ YE A NSCFR MD GQTLWS MS
DisCas7-11	SYAF R HKQIMIP SE RGM S YET NSCFR FD TKRLSW MD
DsbaCas7-11	SYRF T NRVPMIP SE RGM S FE A NSCFR FD KYRLSW MD
SstCas7-11	SYAF S NGDIM P SE RGM S YE A NSCFR FD GYRLSW ME
SybCas7-11	MCKF S AGKPMIP SE RGM S YE A NSCFR FD EKYLTR VQ
CsbCas7-11	NYKF N NGELMIP SE RGM R HFE A KSCFA FG DSTLSW MN
CjcCas7-11	NYQF H NDEIM P SE RGM S YE A NSCFR YD TKYITR LS
CbfCas7-11	TYDF QMNNAIMIP SE RGM SA YE M NSCFR FH KQYLTR IS
HvsCas7-11	YYHFLK DNKPIIP AE RGA S YE A NSCFR FG KKVLSW ME
HsmCas7-11	VYKF R GGHLCIP AE RGM S YE A NSCFR FD KRLISW MT
FmCas7-11	SFRF R DDEVLIIP SE RGM S FE A NSCFR IN KAHLWS IN
DpbaCas7-11	RFKFMRMGSQAIP SA RSMTS FE A NSCFR LD KSHLSW ME

FIG. 1J

11/142

WmCas7x3	E-----	ALSA--	IGRI--	VE---	RNGELK	YPL
NisCas7x3	DYV-----	ALSA--	MGRI--	VD---	DQGEIK	QPL
GwCas7x3	Q-----	SLSA--	MGRI--	VR---	HDQKLY	LPL
HvmCas7x3	ANRN-----	EDKPLSA--	IGMI--	KKIETGDKVEYR	LPL	
DsbCas7x3	SG-----	SLSA--	IGMV--	VI---	RDGKKF	YPL
GamCas7x3	PSH-----	ILSA--	IGMV--	TK---	RGEDFW	IPL
DesCas7x3	N-----	ALRE--	IGMI--	VL---	RDGKRF	LPL
GabCas7x3	R-----	ALRK--	IGIL--	FK---	REFQWR	VQM
MetCas7x3	R-----	ALRK--	VGMY--	IY---	VDNKSF	IKL
MebCas7x3	L-----	ALSK--	IGITFIN--	---	RQGQWQ	IPM
GwCas7-11	DEA-----	KYRK--	AGR--	TVSGDGAQKKYS	QEM	
HreCas7-11	D-----	EFKDFHPCI	---	---	VBGAK	REM
CmaCas7-11	ESETTQ--	KRKSGRYQVEESDPDLF--	PCR--	QK--	KGNKYK	EKM
HvmCas7-11	DPT-----	ELTKYYPCK--	KR--	DGNKFF	LKM	
SmCas7-11	E-----	DYKDFYPCR--	---	---	LDGGKQ	KKM
OmCas7-11	SEHK-----	DYQ--	PGK--	---	TDNGRK	QPM
DisCas7-11	DHQN-----	VLQDFLPCR--	---	---	TADGKH	QKF
DsbaCas7-11	DVK-----	ELEQFKPCR--	---	---	ADDGKR	EEM
SstCas7-11	DRN-----	VLMQFKPCR--	---	---	TDNGLR	EEM
SybCas7-11	KKGA-----	KSELVPCI	---	VWG--	QNGGLAVQV	
CsbCas7-11	DEKDYKIDSNSIRKMESQORNPYRI--	PDE--	QKELRNSGNGLFNRLLY			
CjcCas7-11	EKKD--	ESNDKNKSQDDASQKIR--	KGL--	KK--	TDEGFS	IEV
CbfCas7-11	EDK-----	ELREFIPCI	---	RI---	INGDVI	EKA
HvsCas7-11	KDAK-----	EFM--	PCR--	SK--	KKGKLYMVKM	
HsmCas7-11	EEAK--	RPDPKKSEEQNRMRFR--	PCR--	IK--	KDKKFF	QEM
FmCas7-11	DMAK-----	HYR--	PCR--	IQ--	NNEKMF	QPY
DpbaCas7-11	DDAG-----	DYK--	PCR--	EK--	KDDKAV	RKF

FIG. 1K

12/142

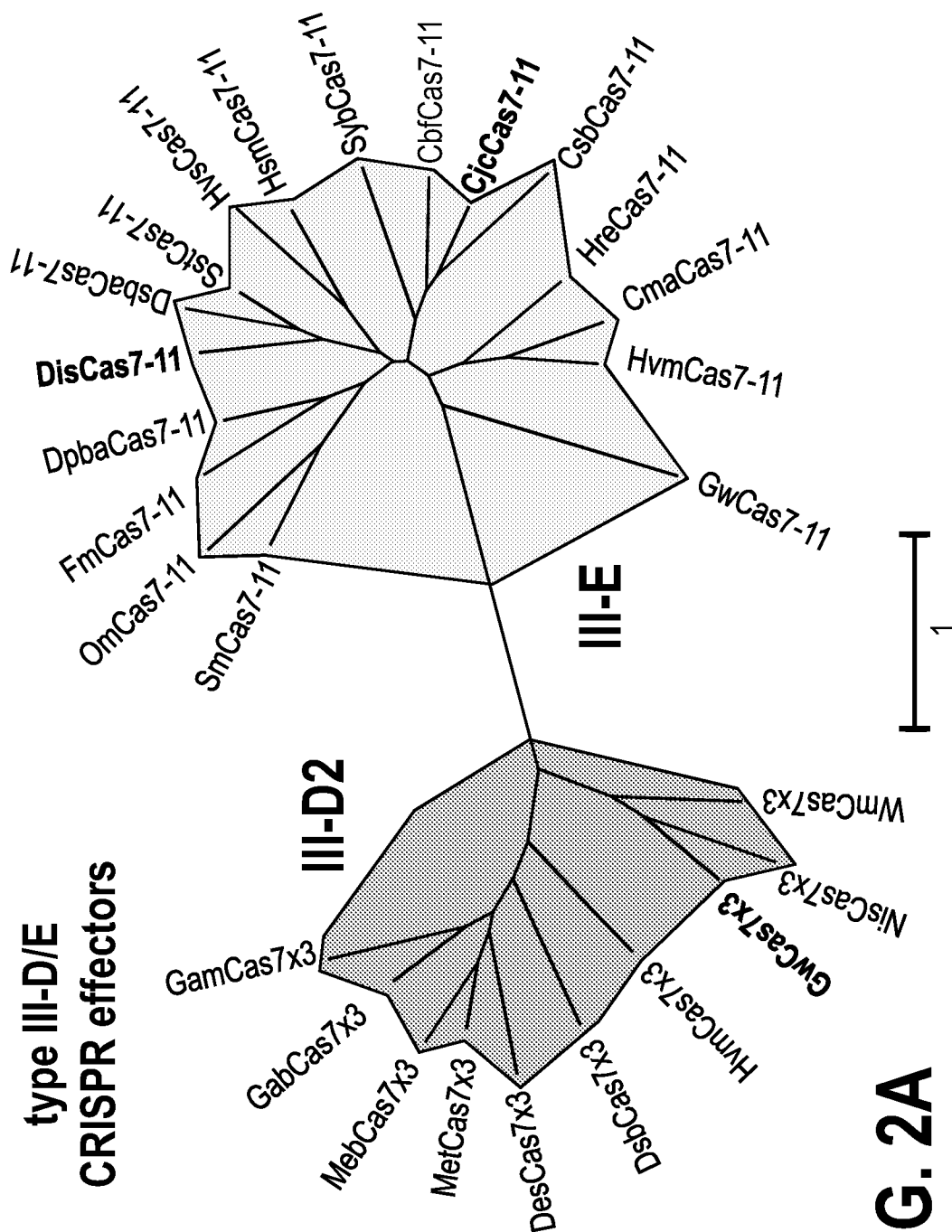


FIG. 2A

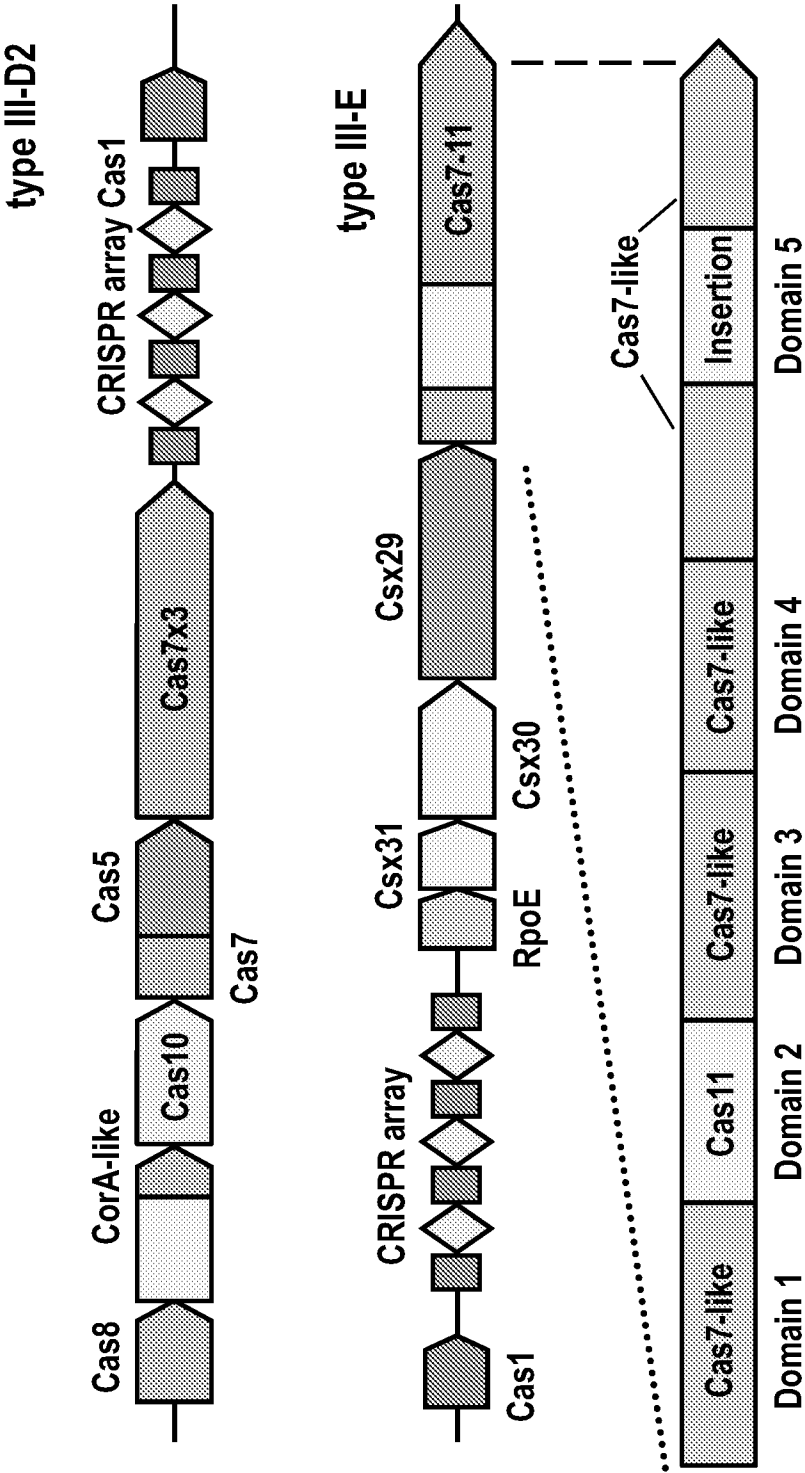
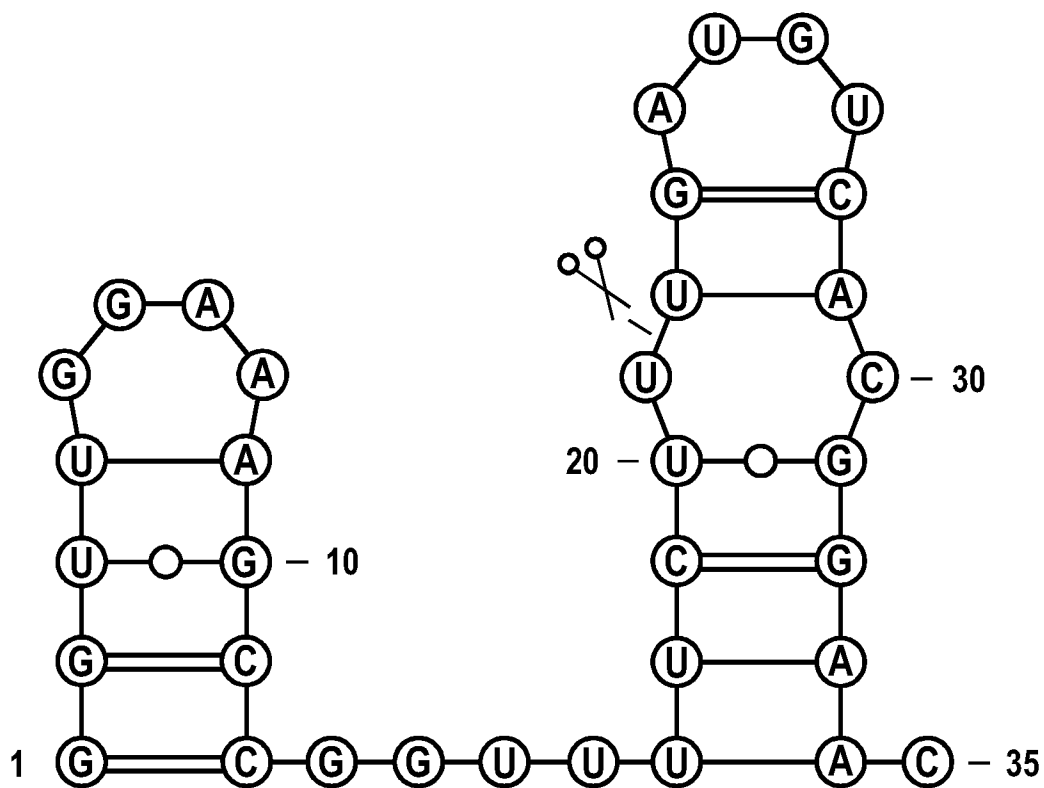


FIG. 2B





DR processing

FIG. 2D

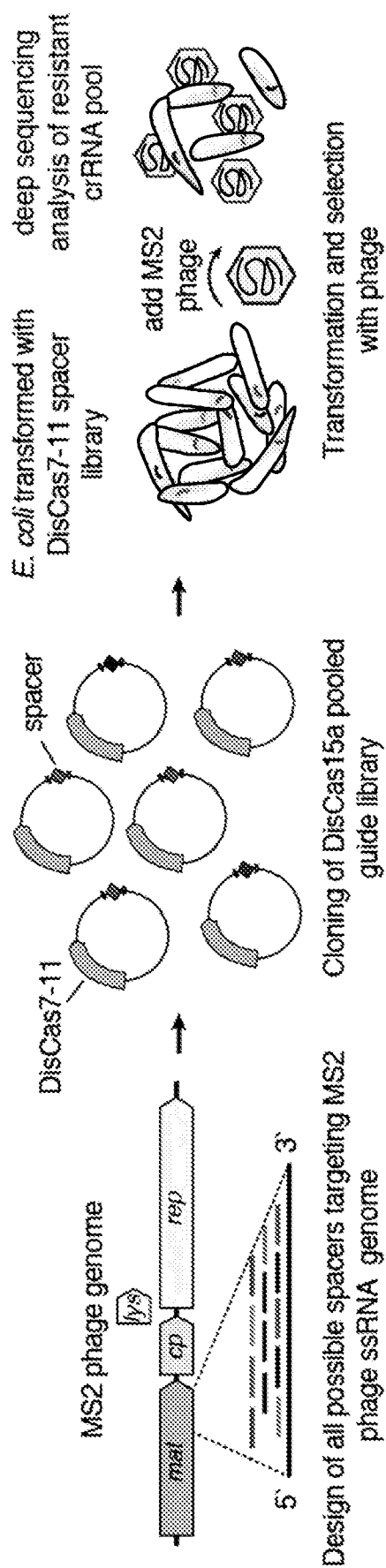
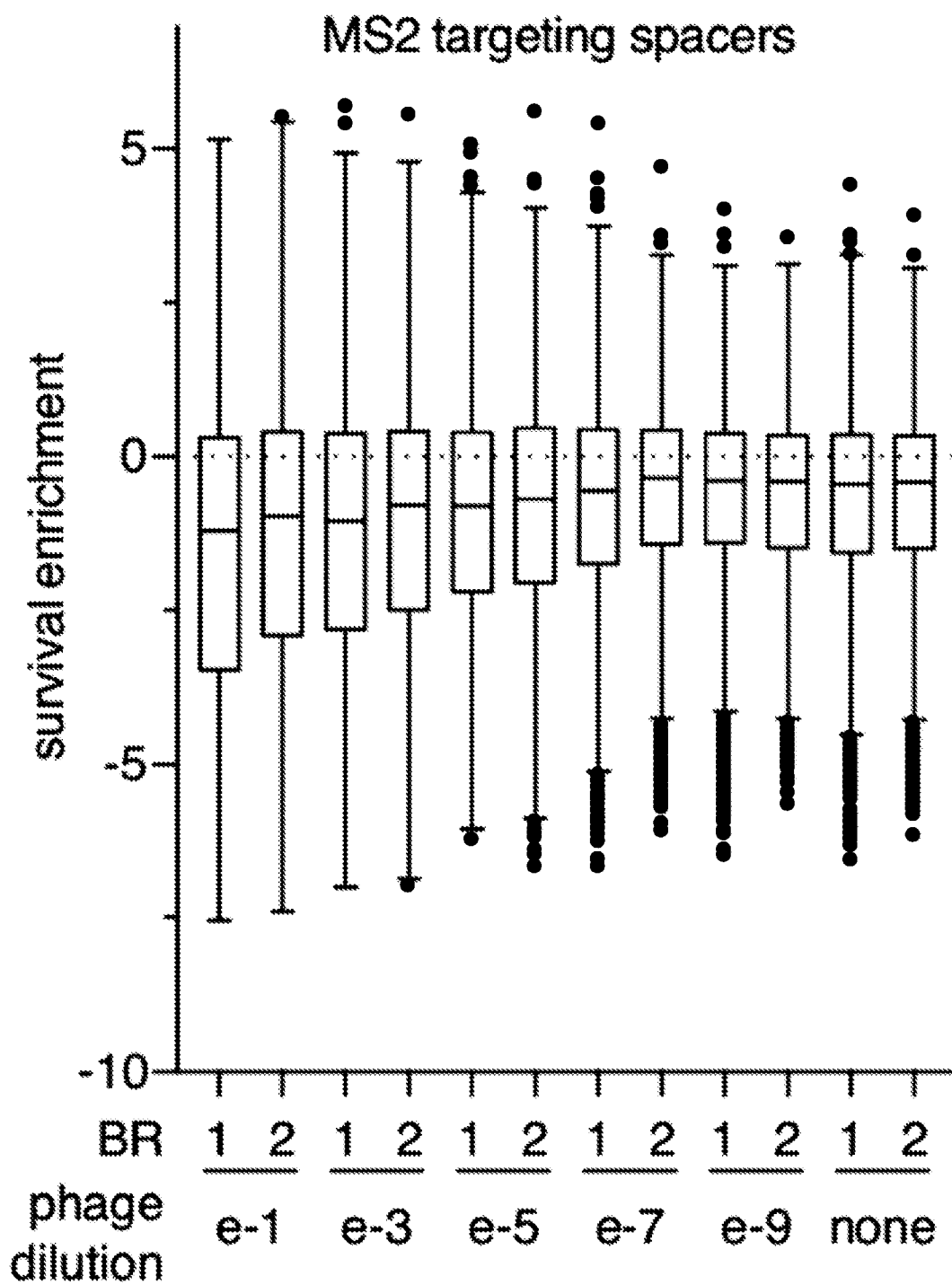
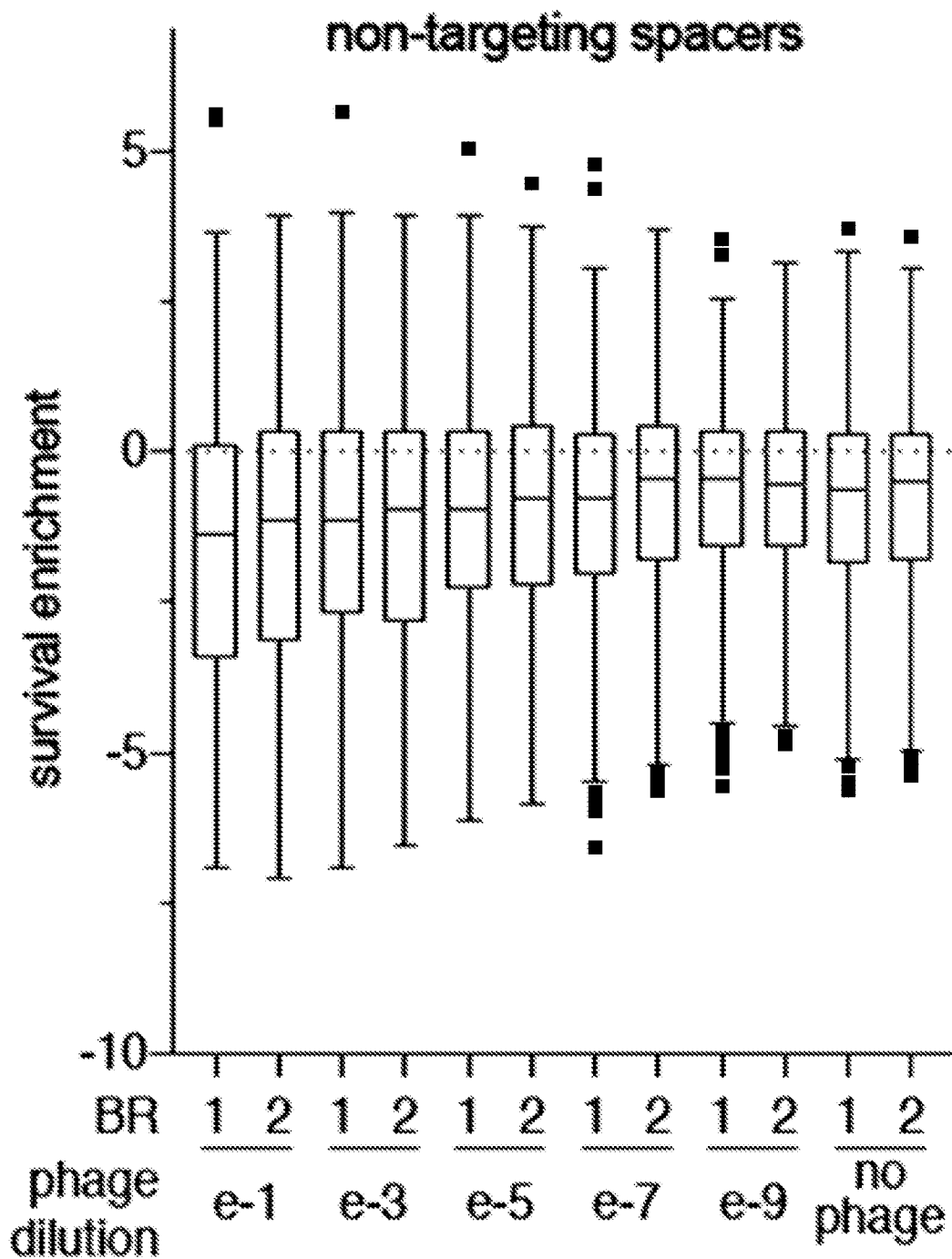


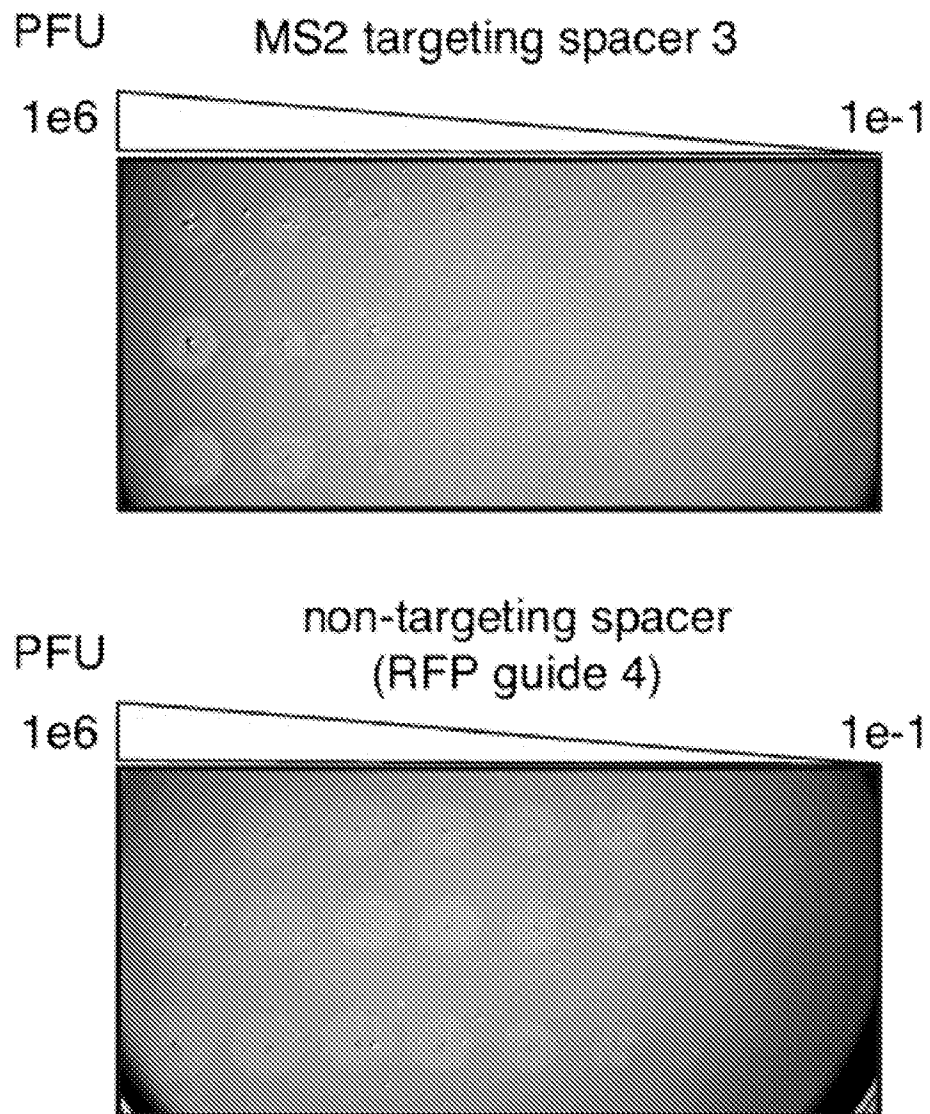
FIG. 3A

17/142

**FIG. 3B**

18/142

**FIG. 3C**

**FIG. 3D**

20/142

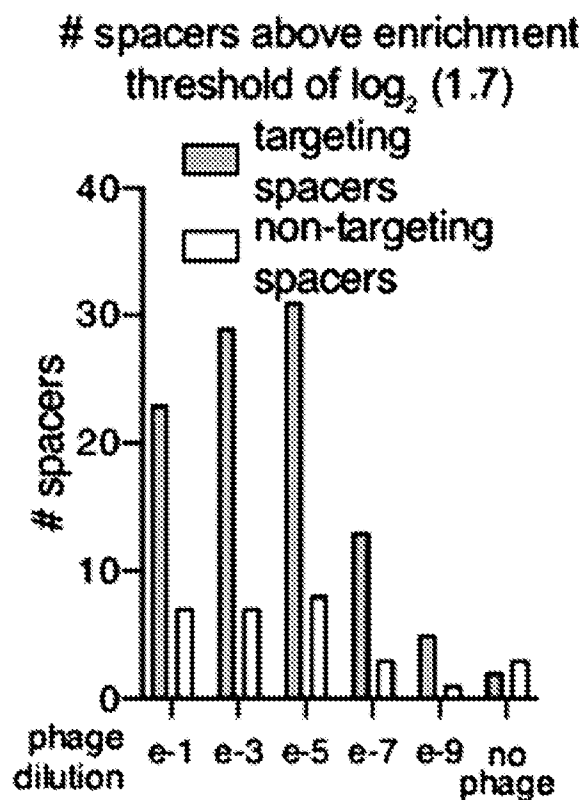


FIG. 3E

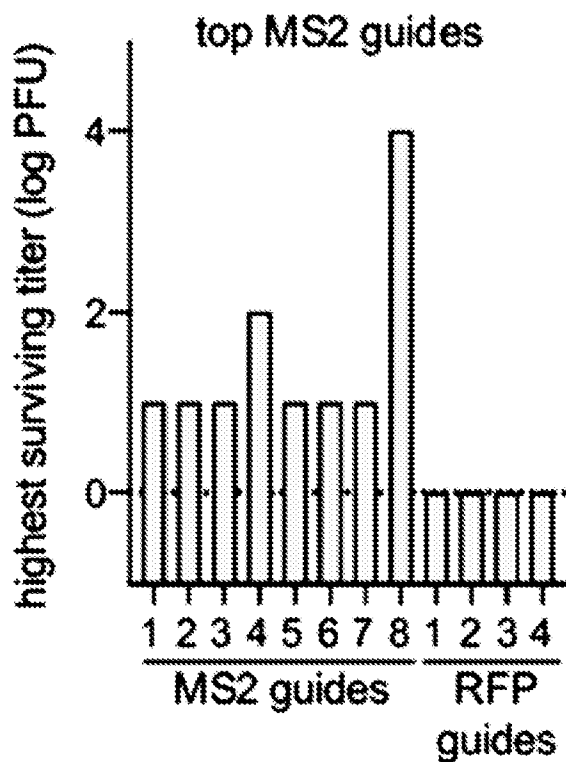
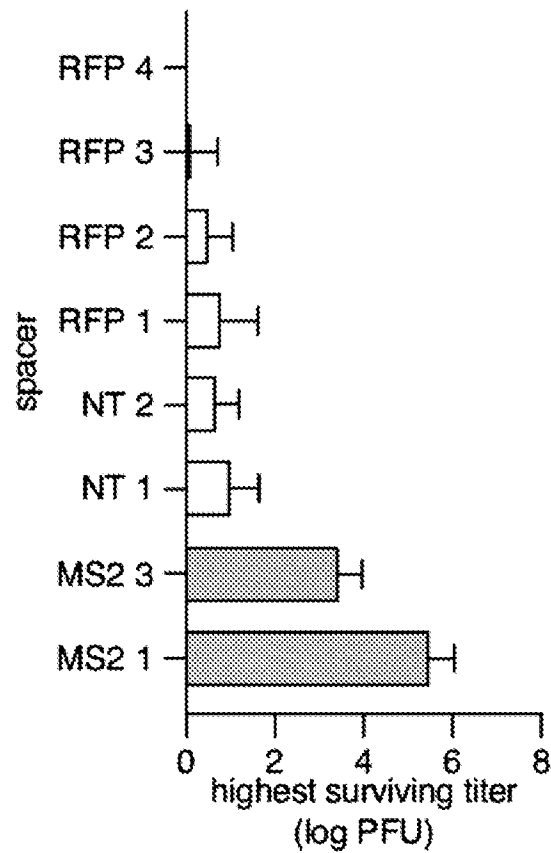
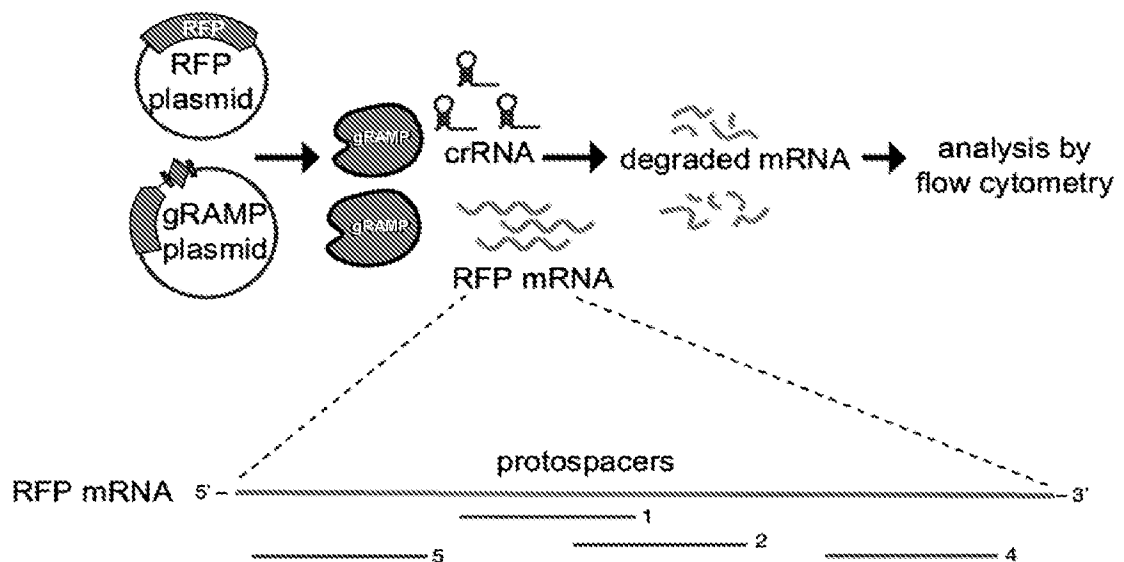
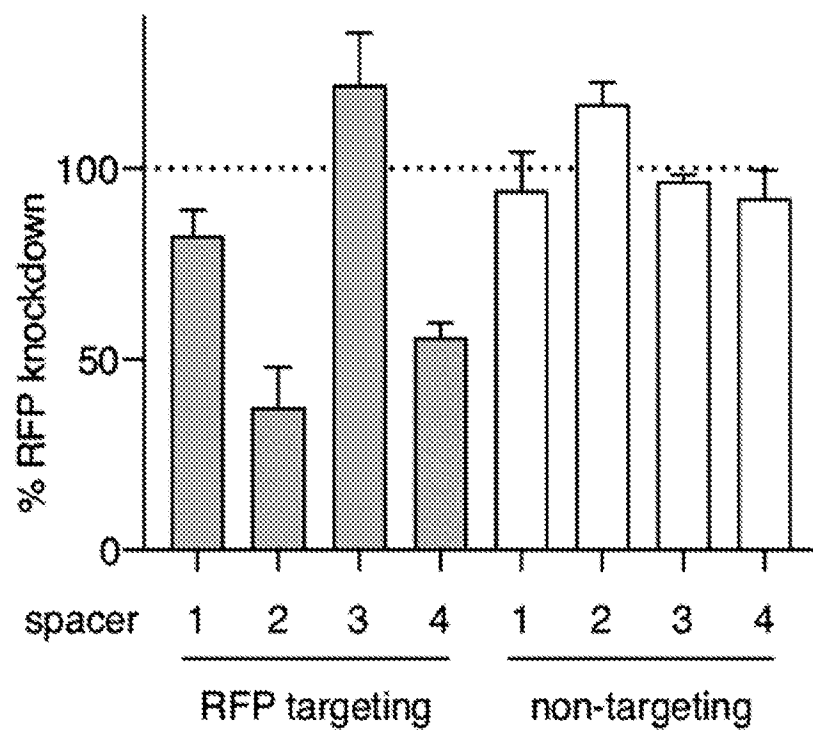
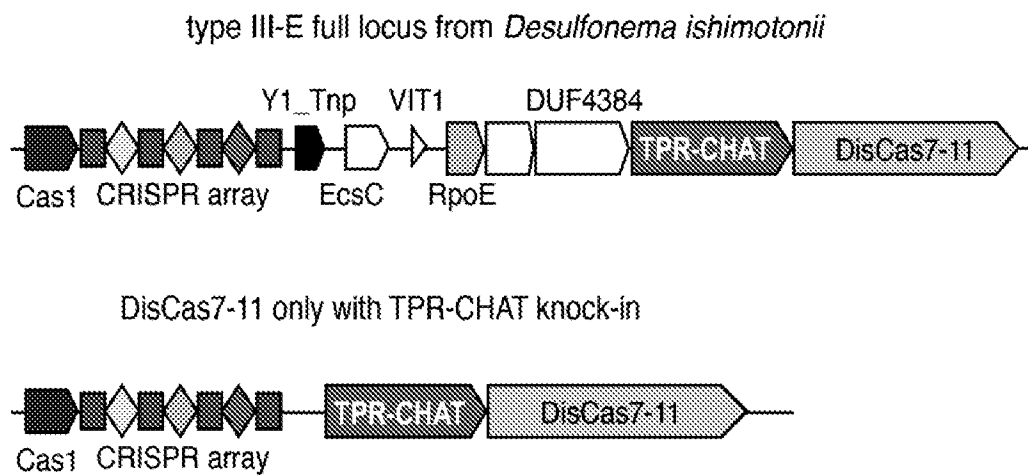


FIG. 3F

21/142

**FIG. 3G****FIG. 3H**

22/142

**FIG. 3I****FIG. 3J**

23/142

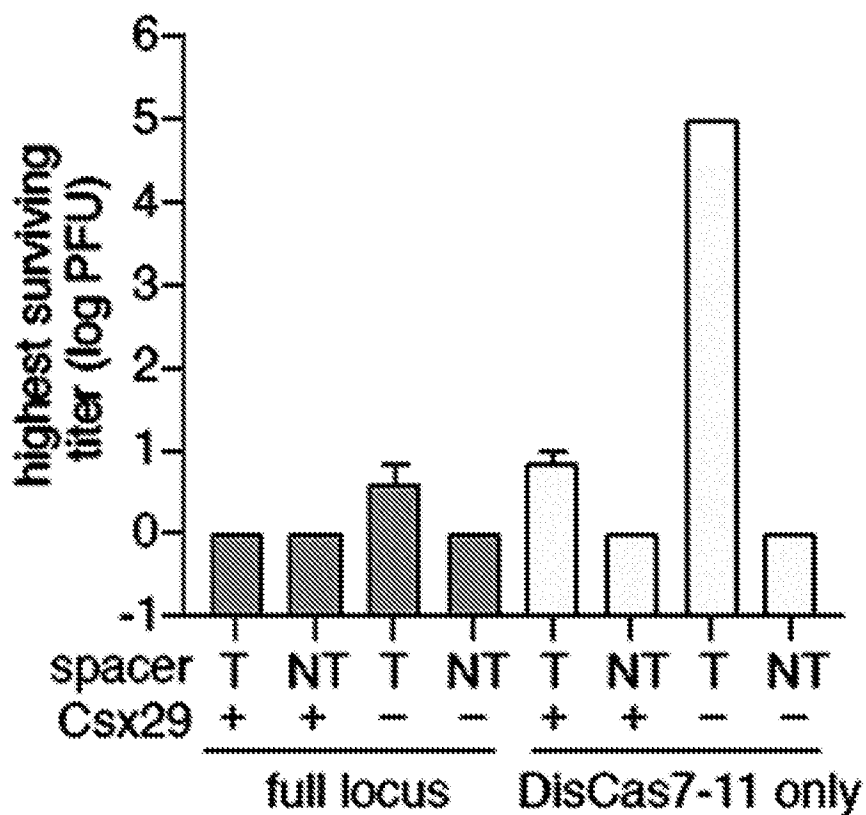
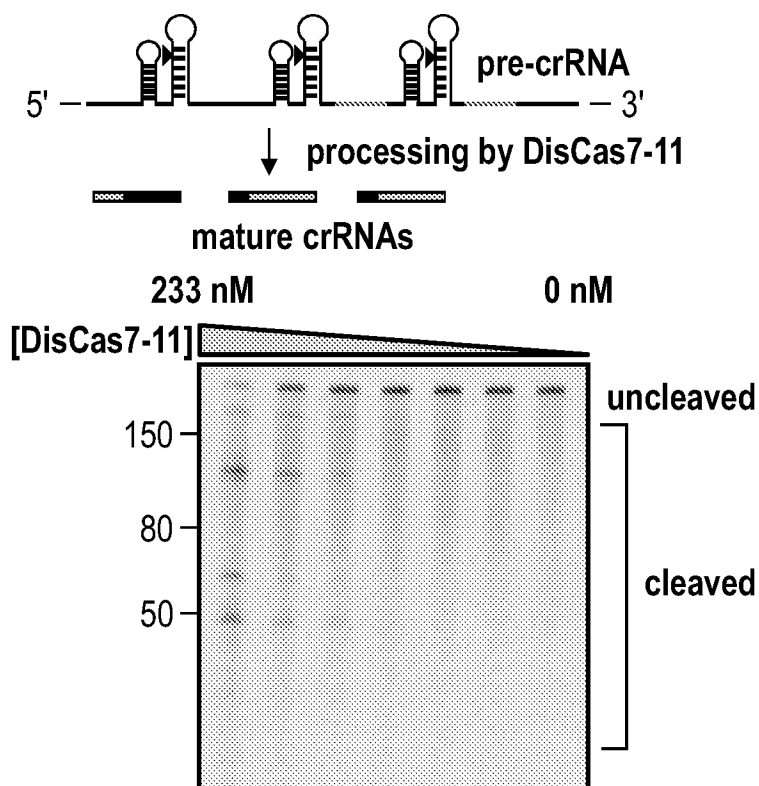
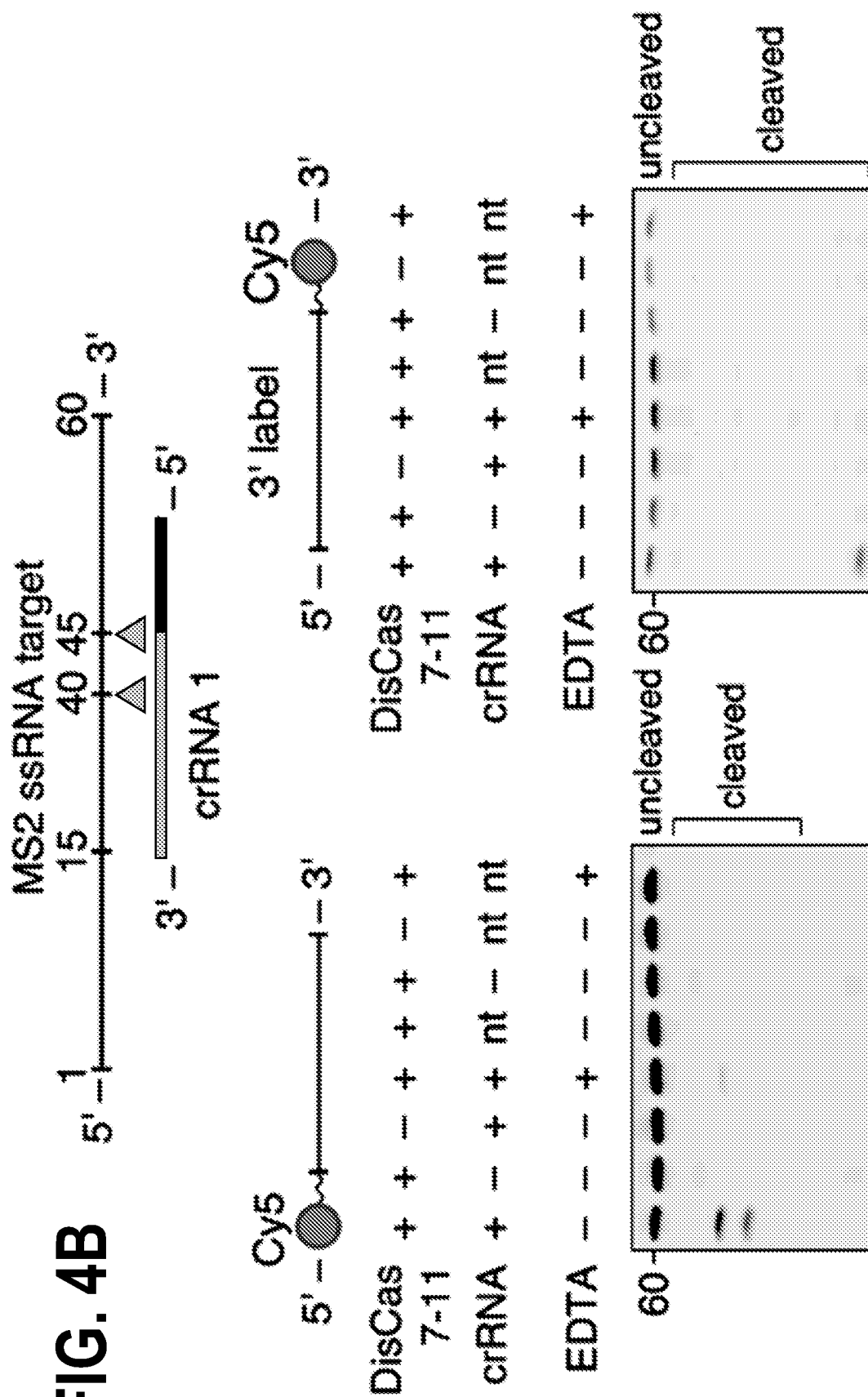
**FIG. 3K****FIG. 4A**

FIG. 4B



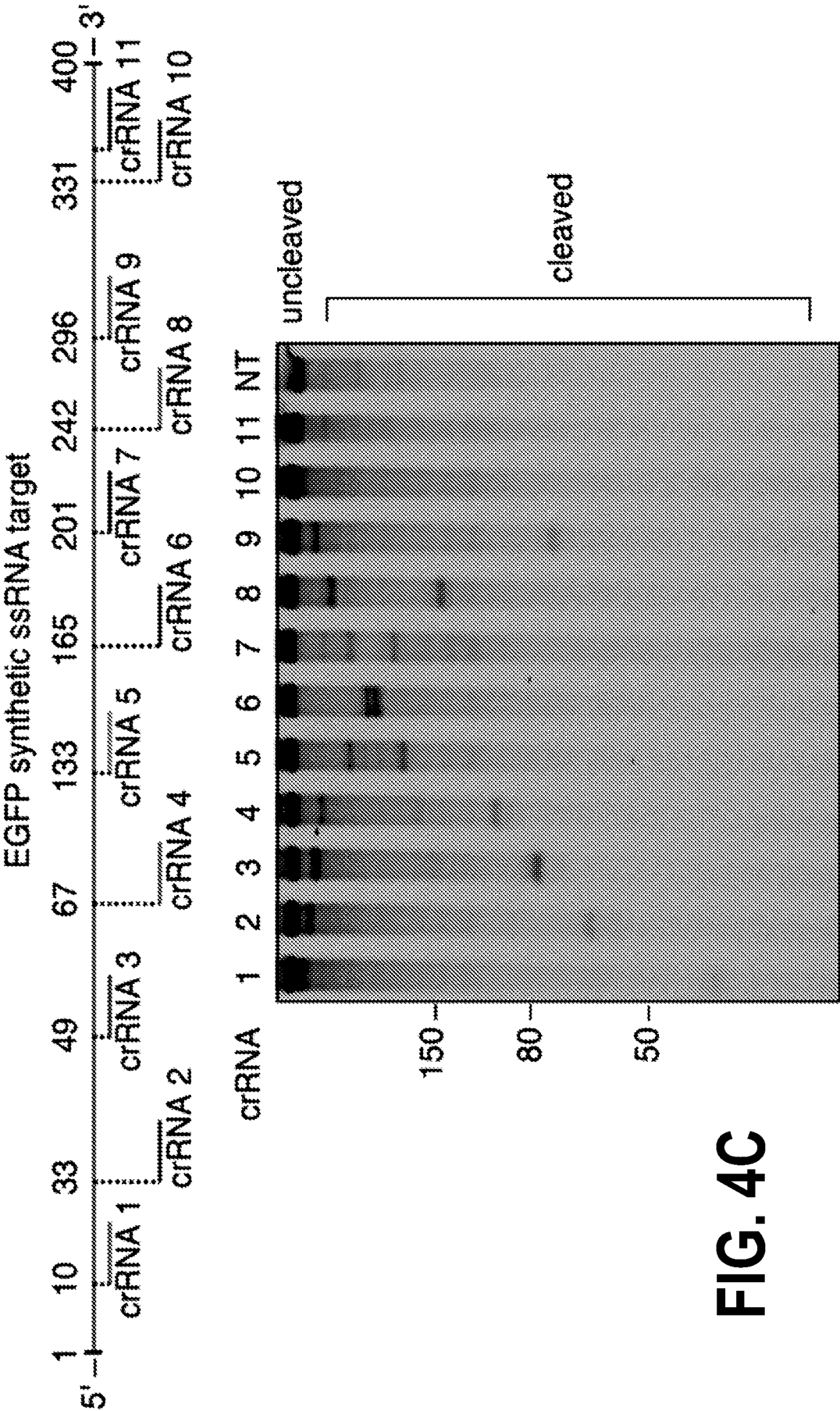


FIG. 4C

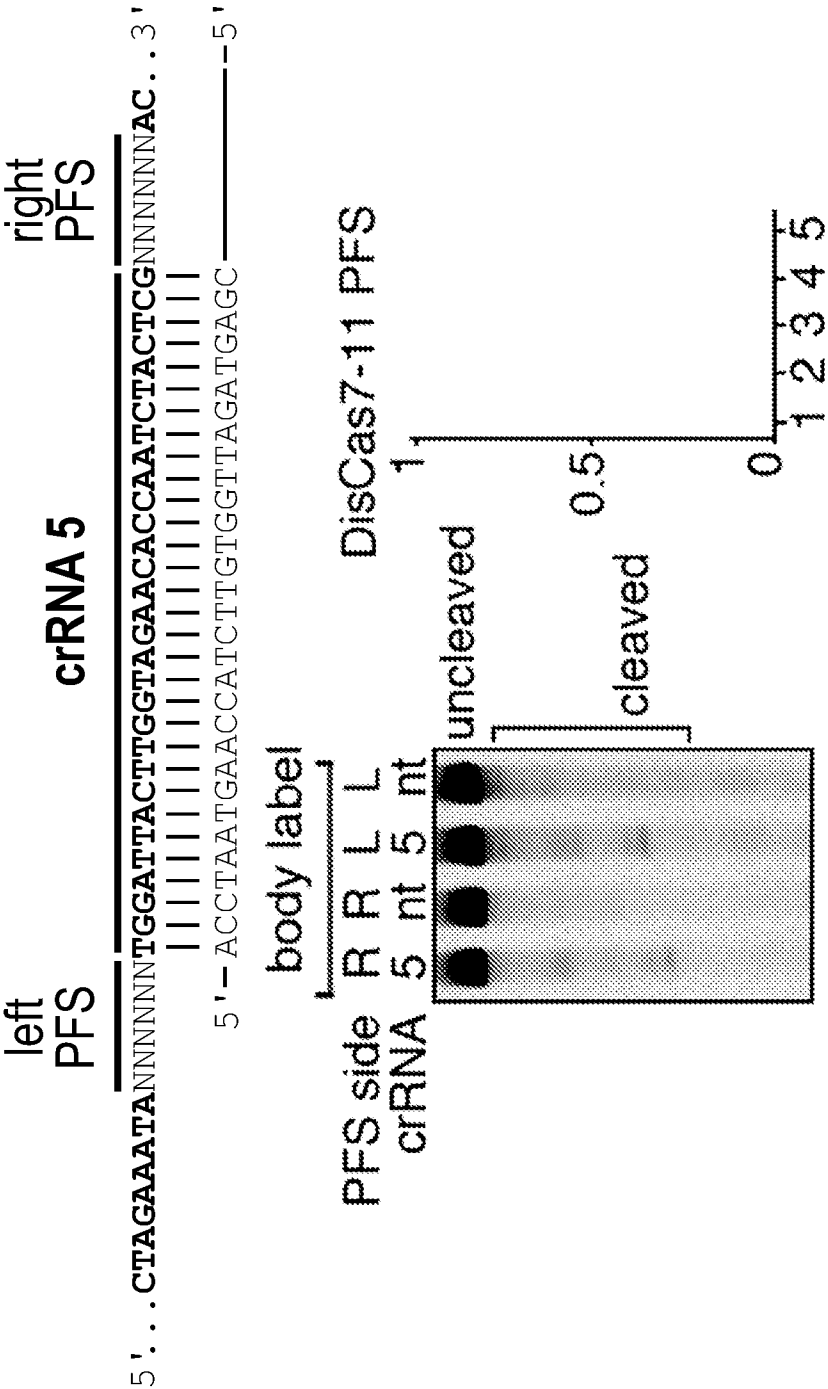
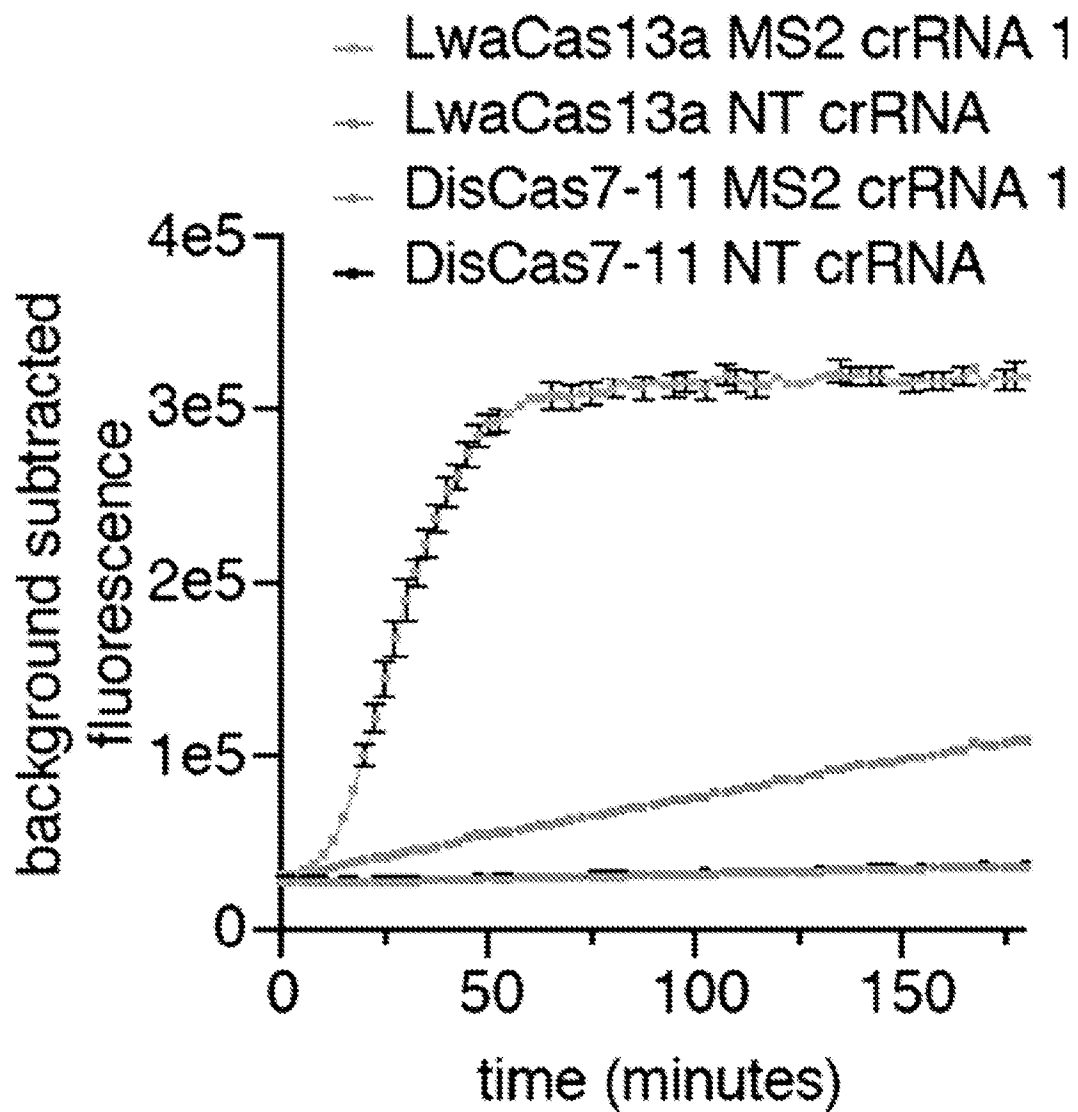
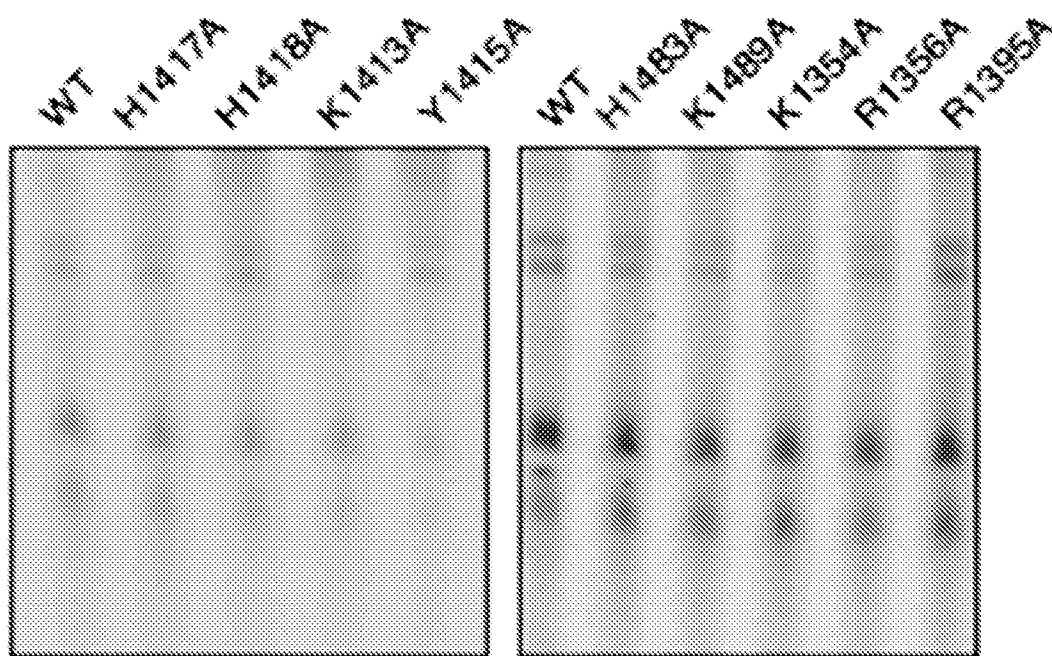
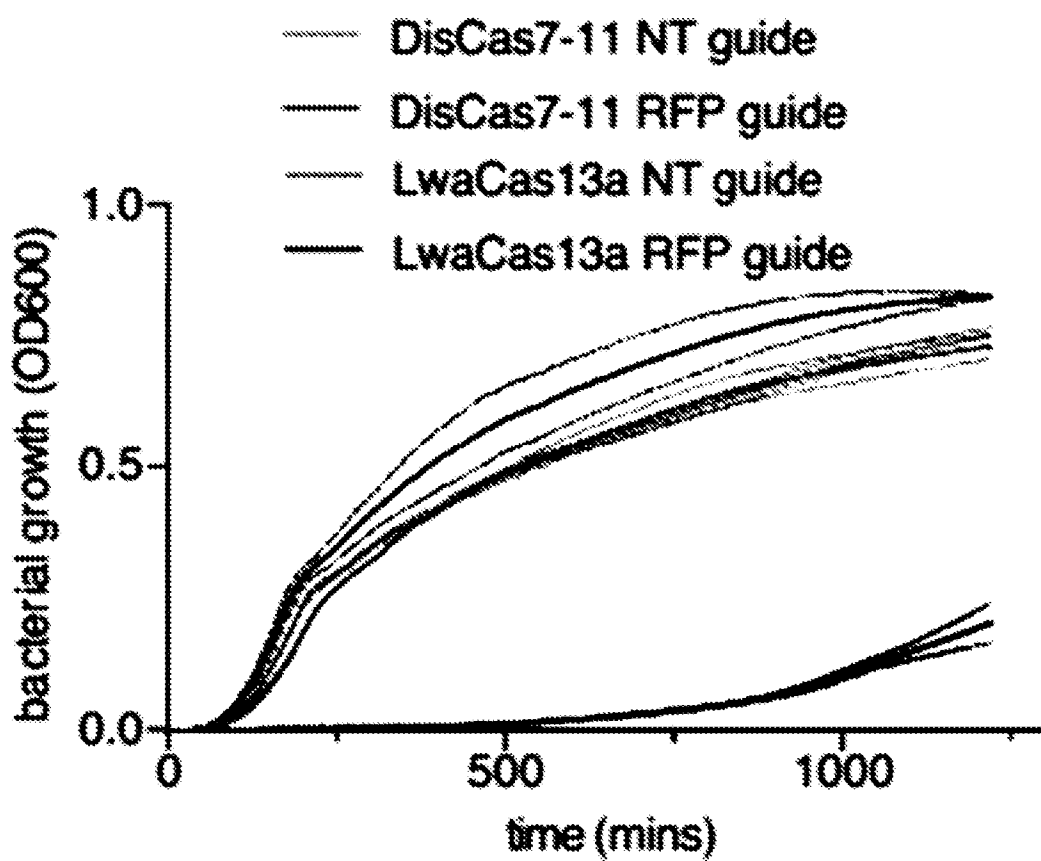


FIG. 4D

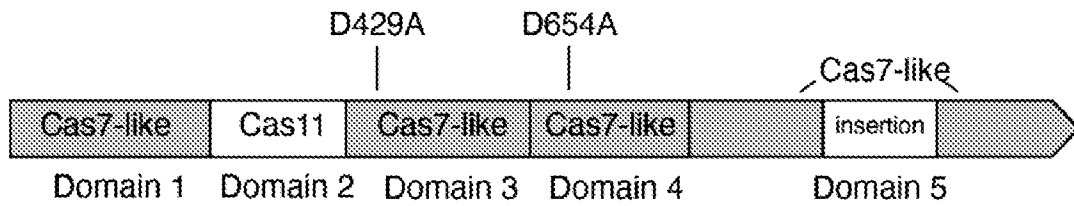
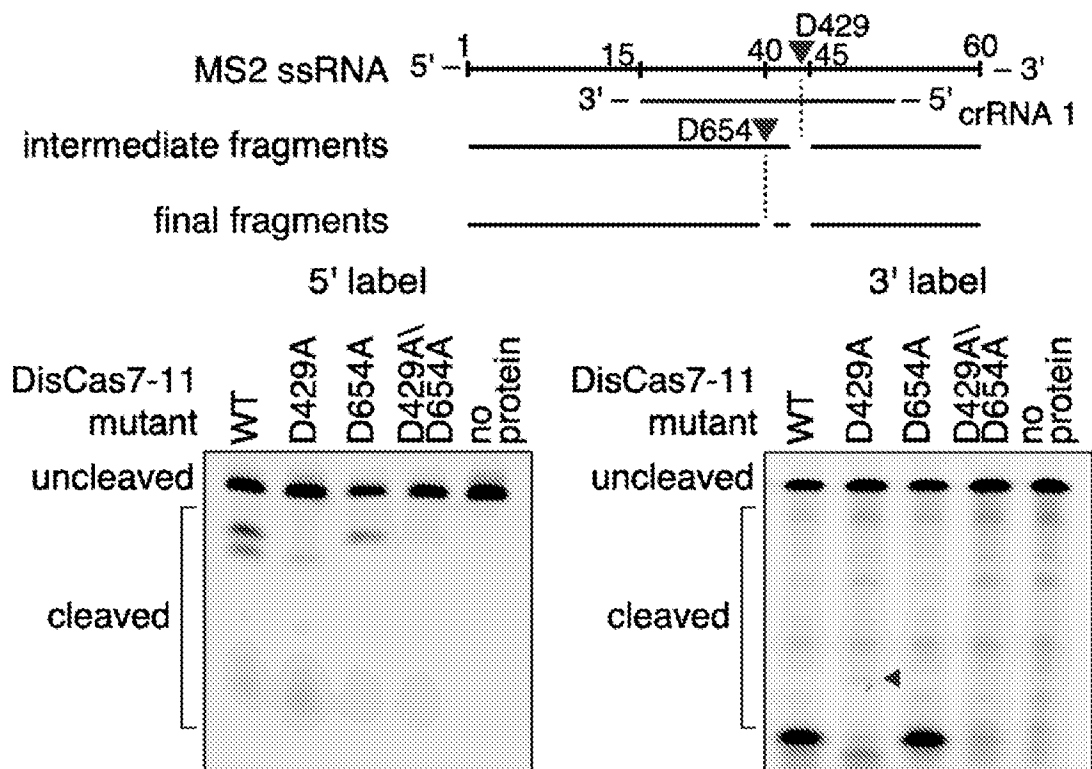
27/142

**FIG. 4E**

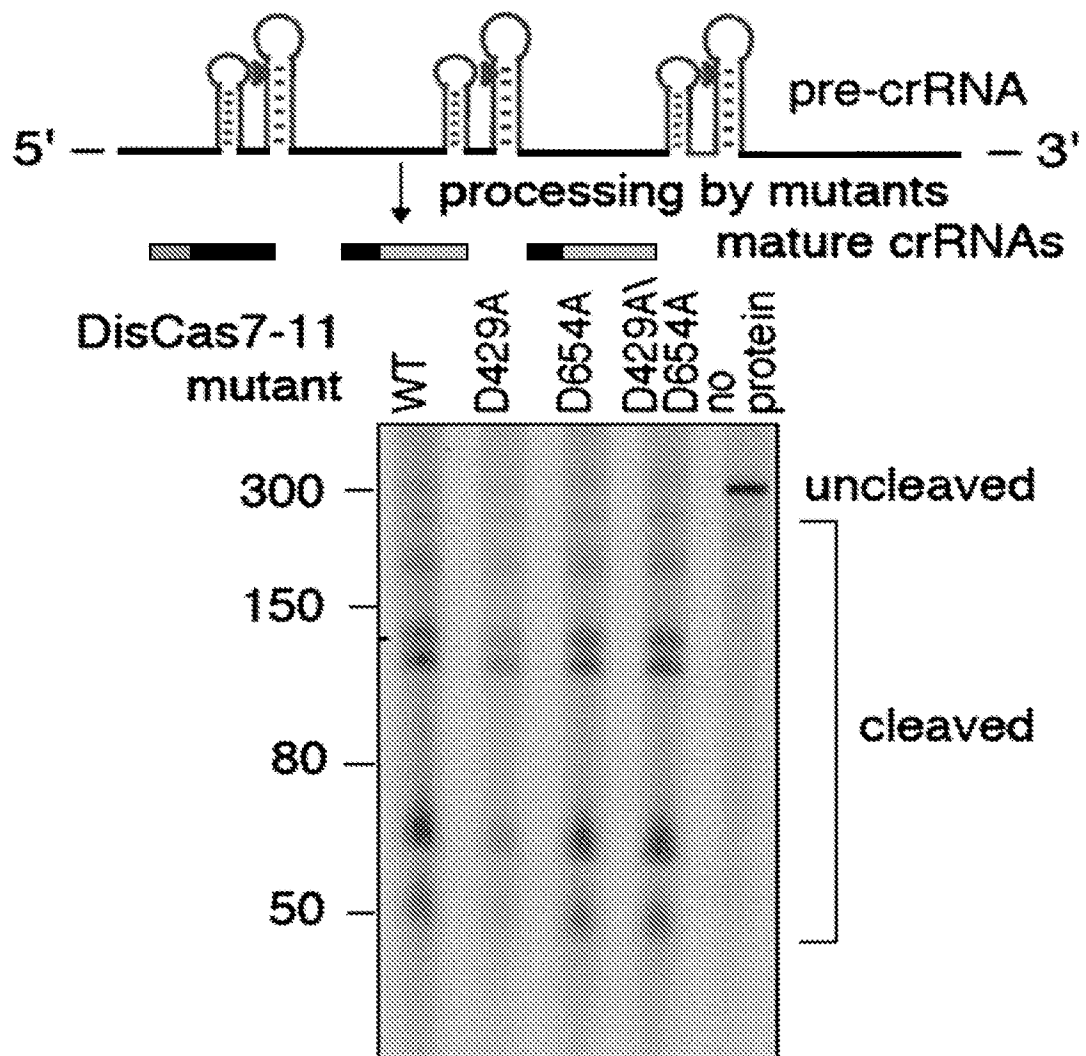
28/142

**FIG. 4F****FIG. 4G**

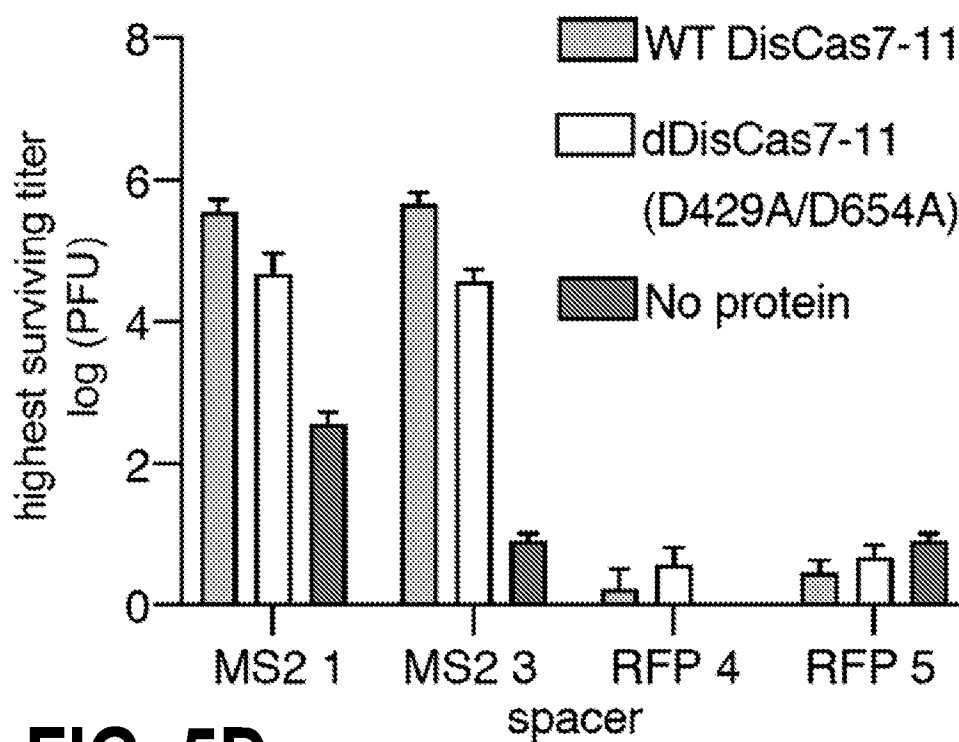
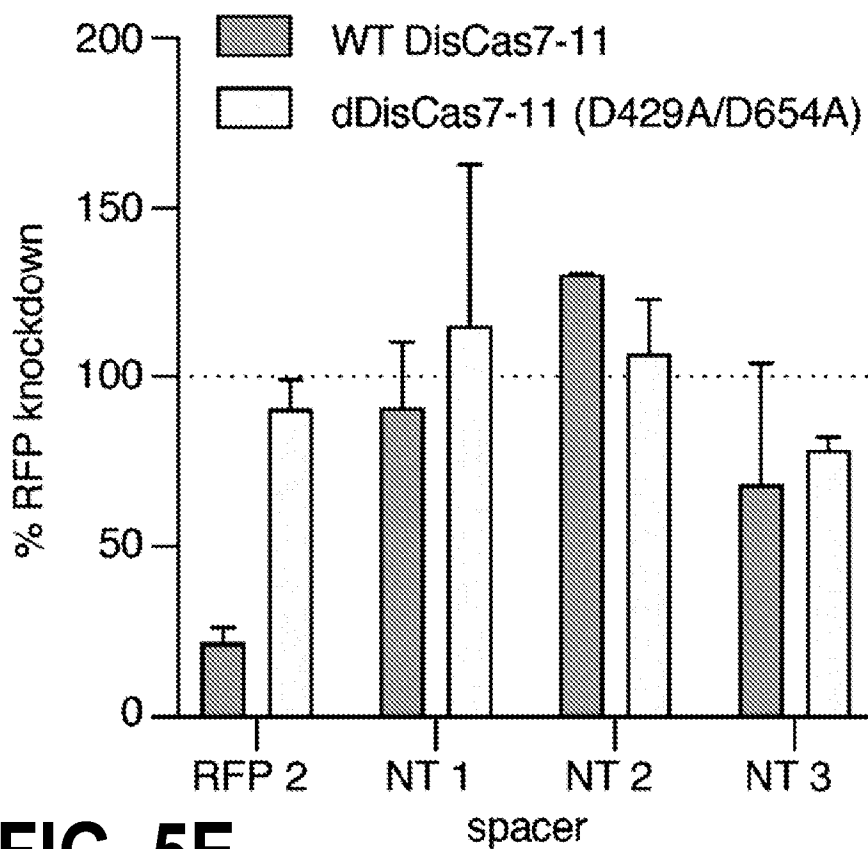
29/142

**FIG. 5A****FIG. 5B**

30/142

**FIG. 5C**

31/142

**FIG. 5D****FIG. 5E**

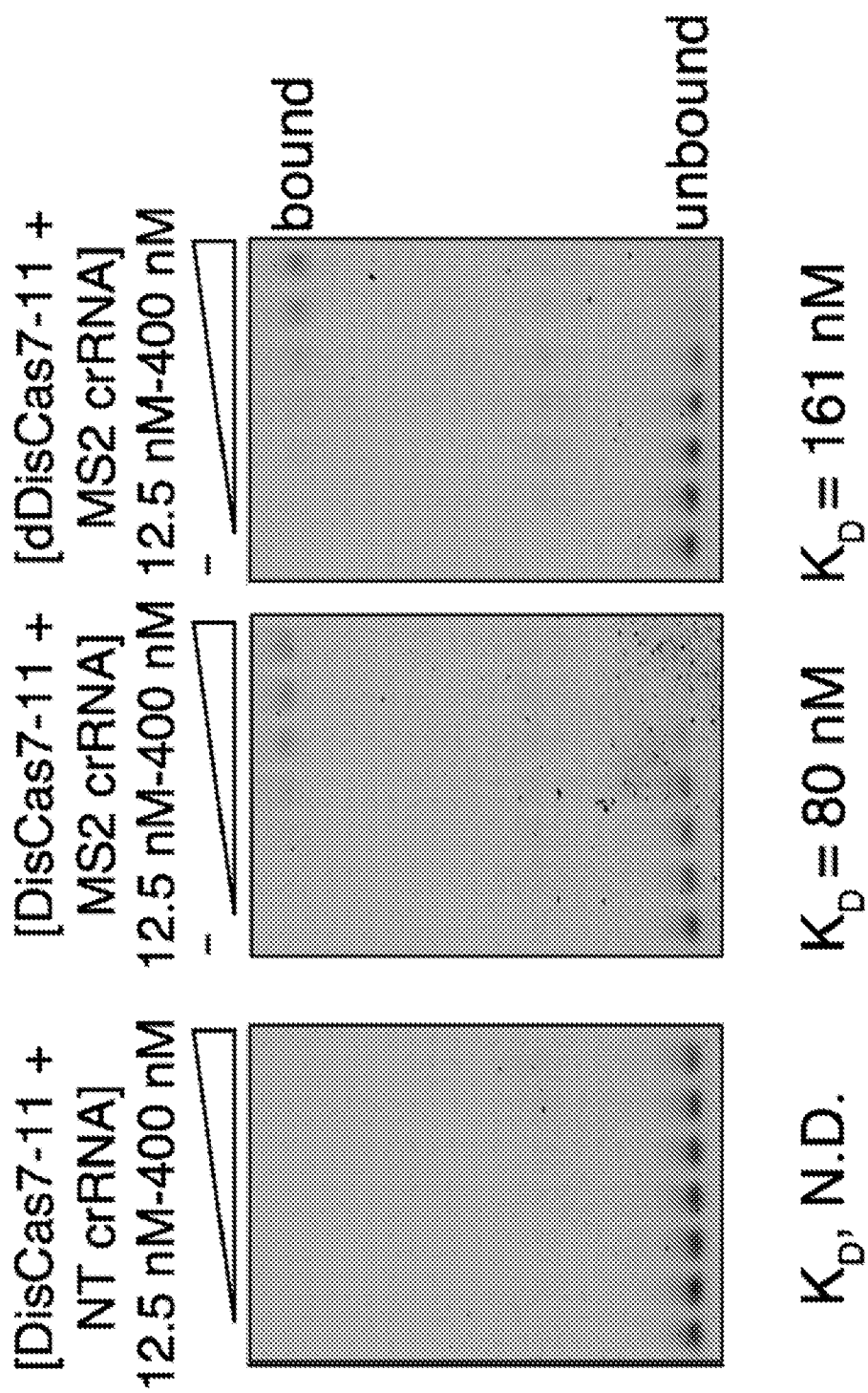


FIG. 5F

33/142

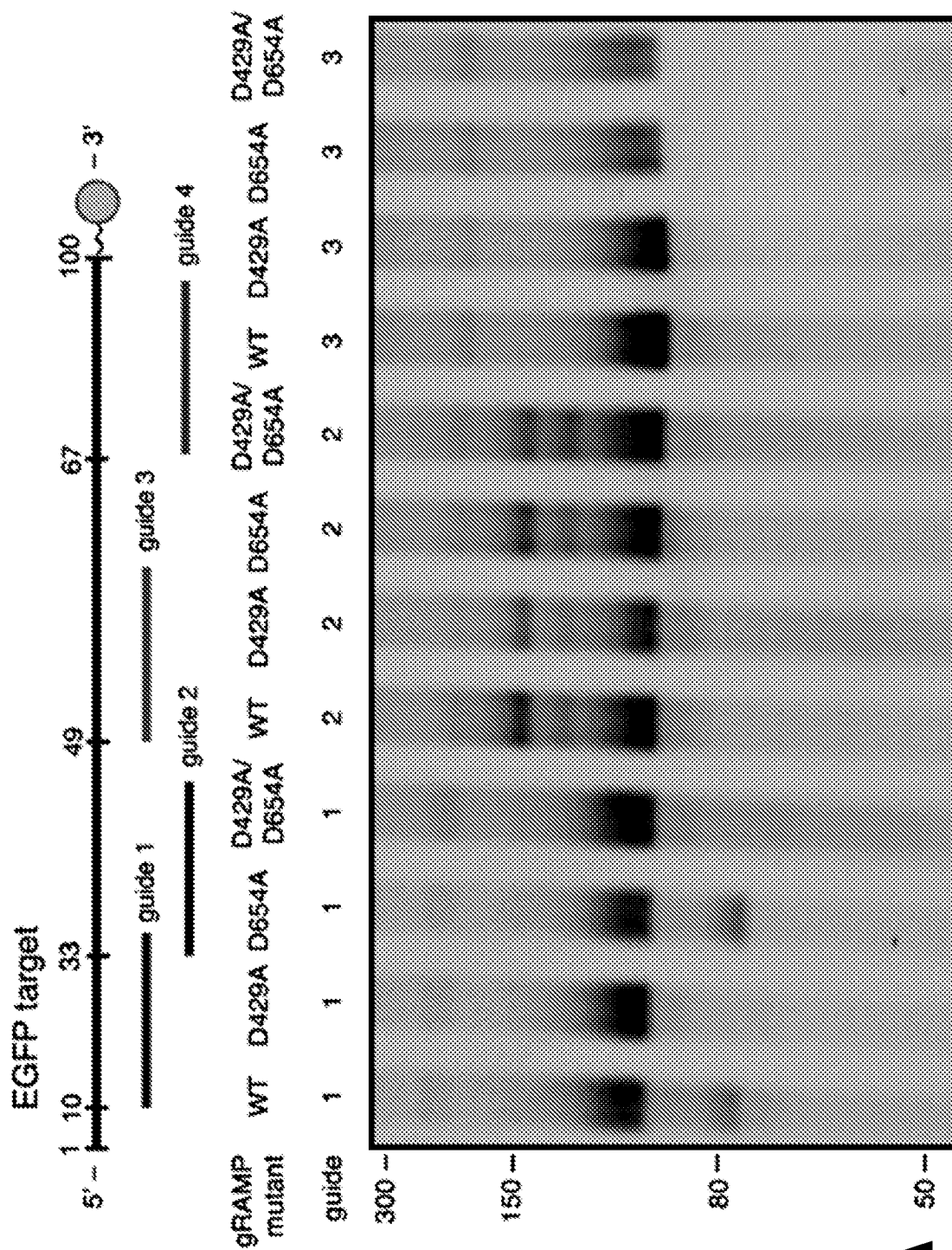


FIG. 6A



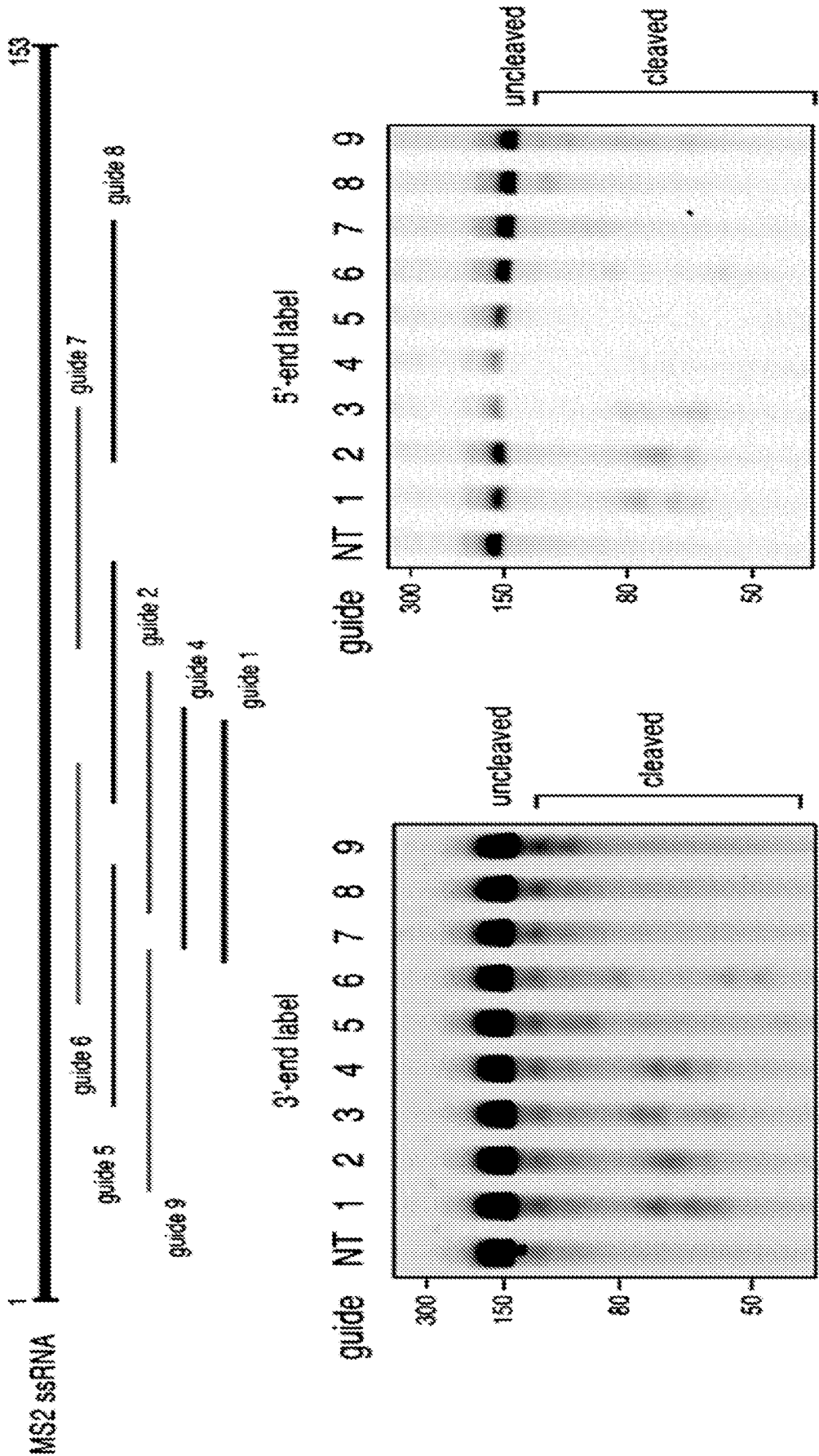


FIG. 6C

36/142

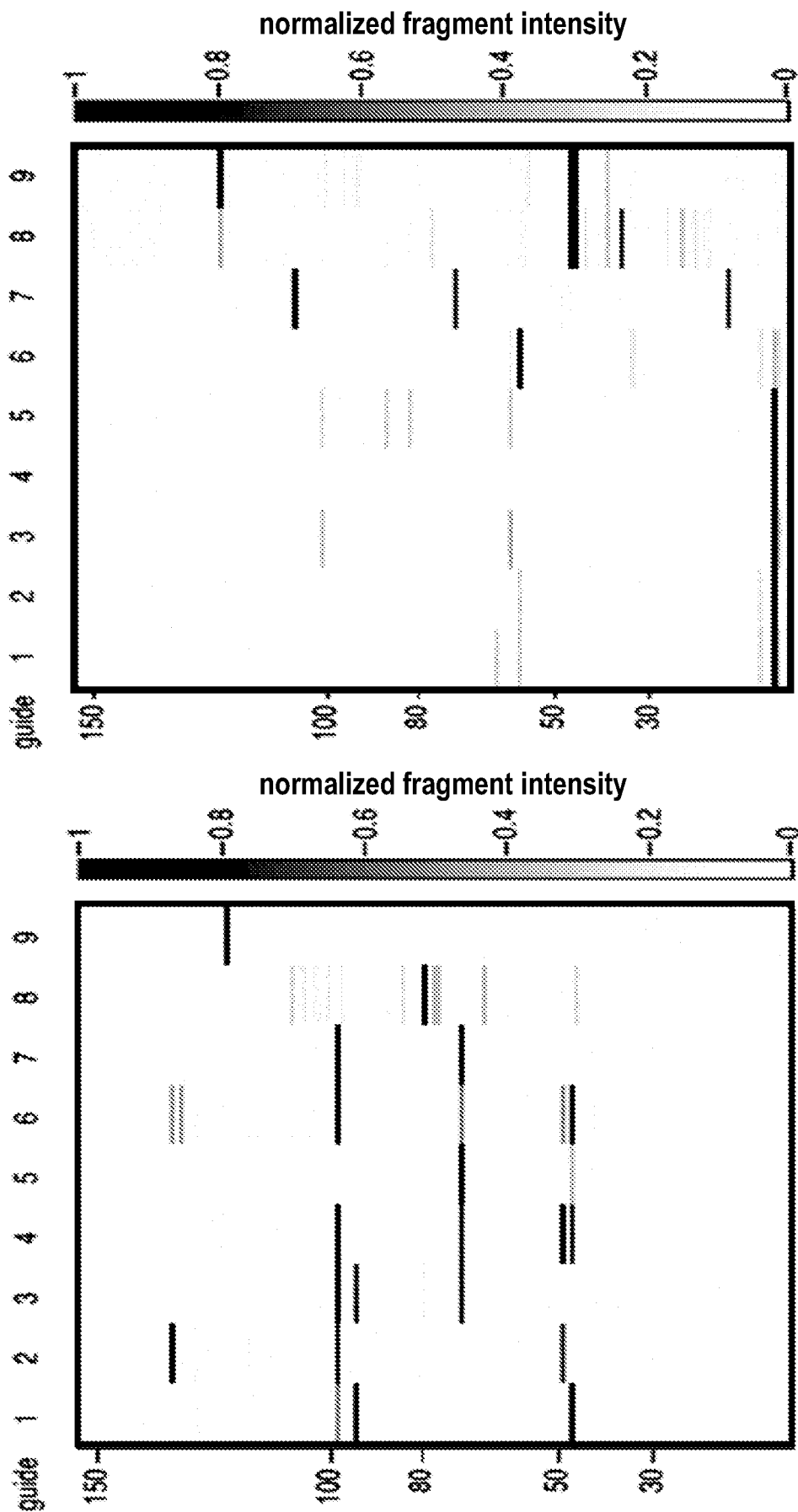


FIG. 6D

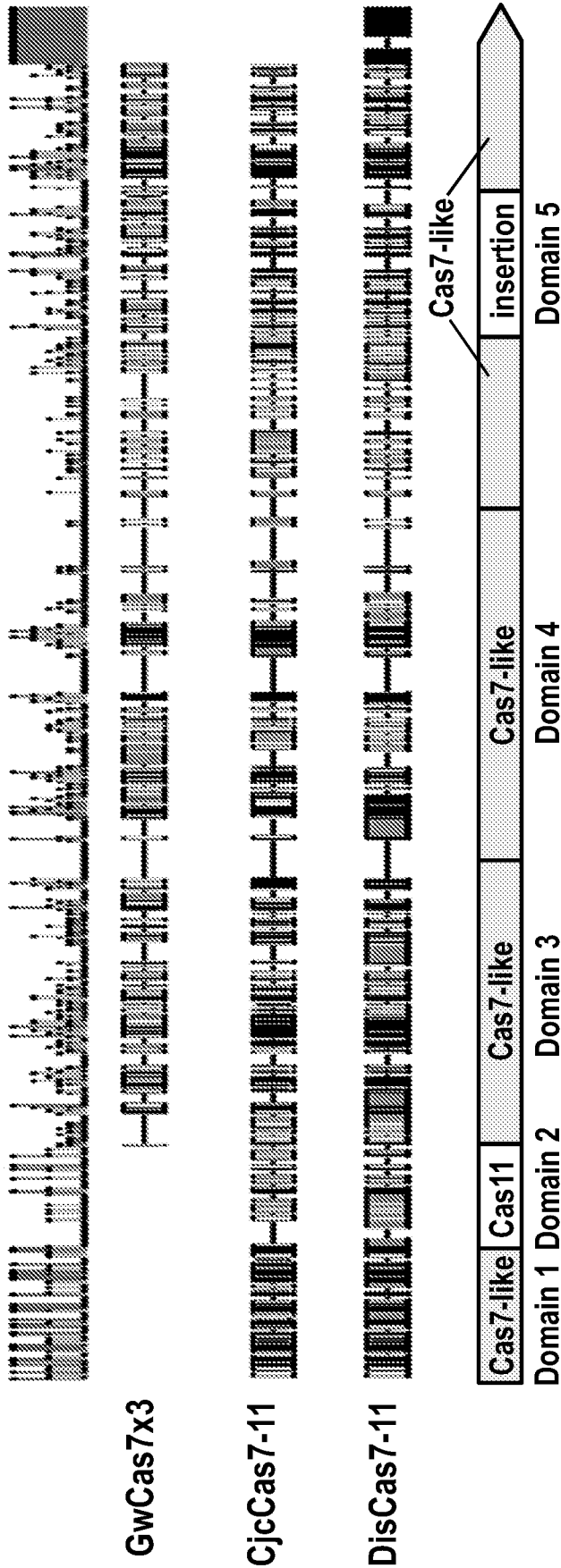


FIG. 7A

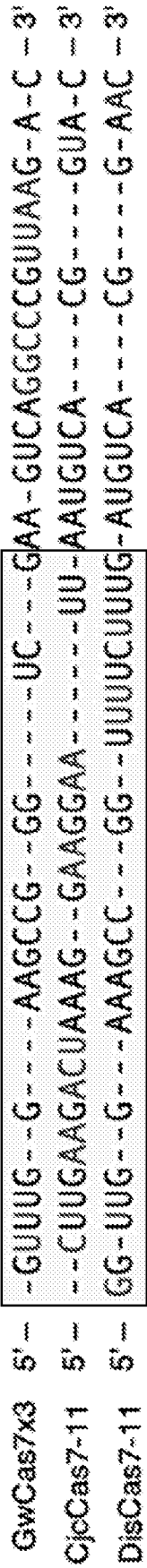
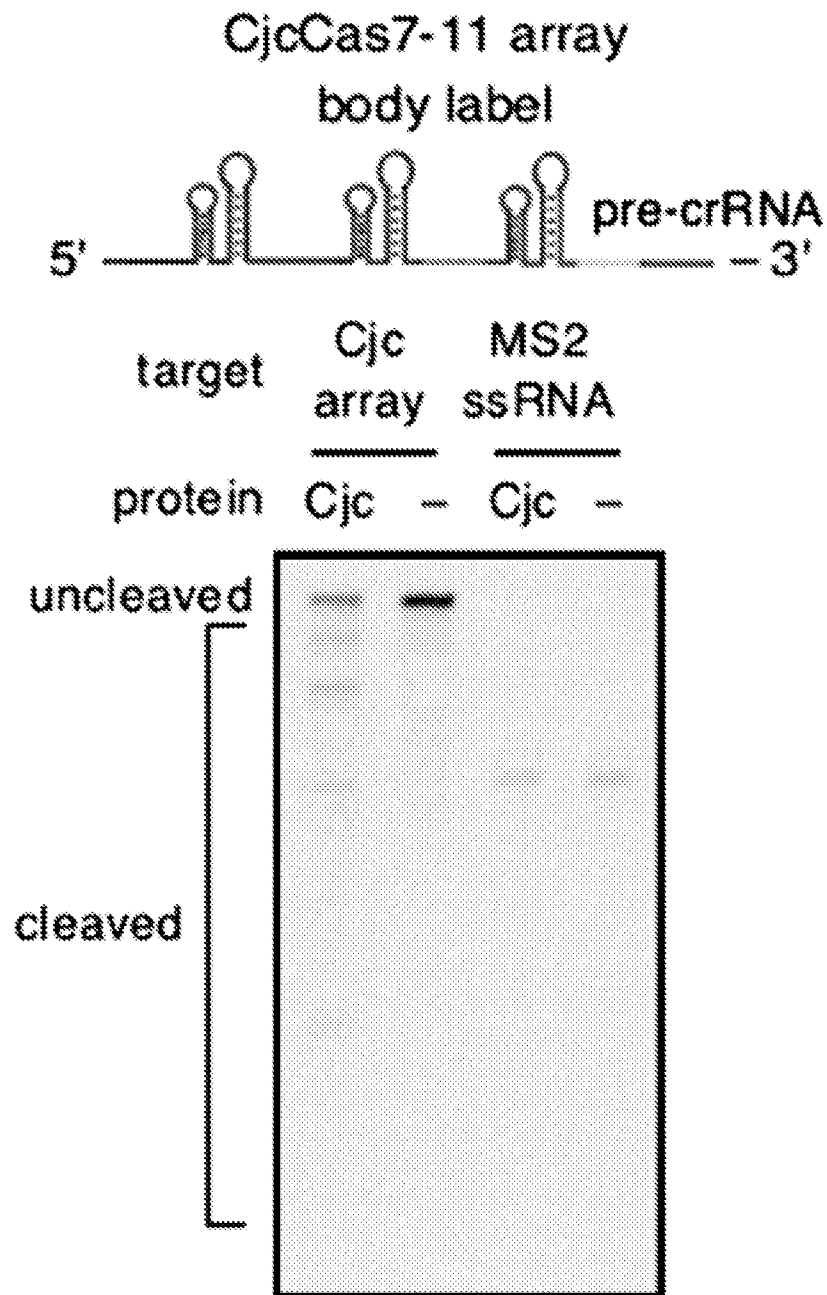


FIG. 7B

38/142

**FIG. 7C**

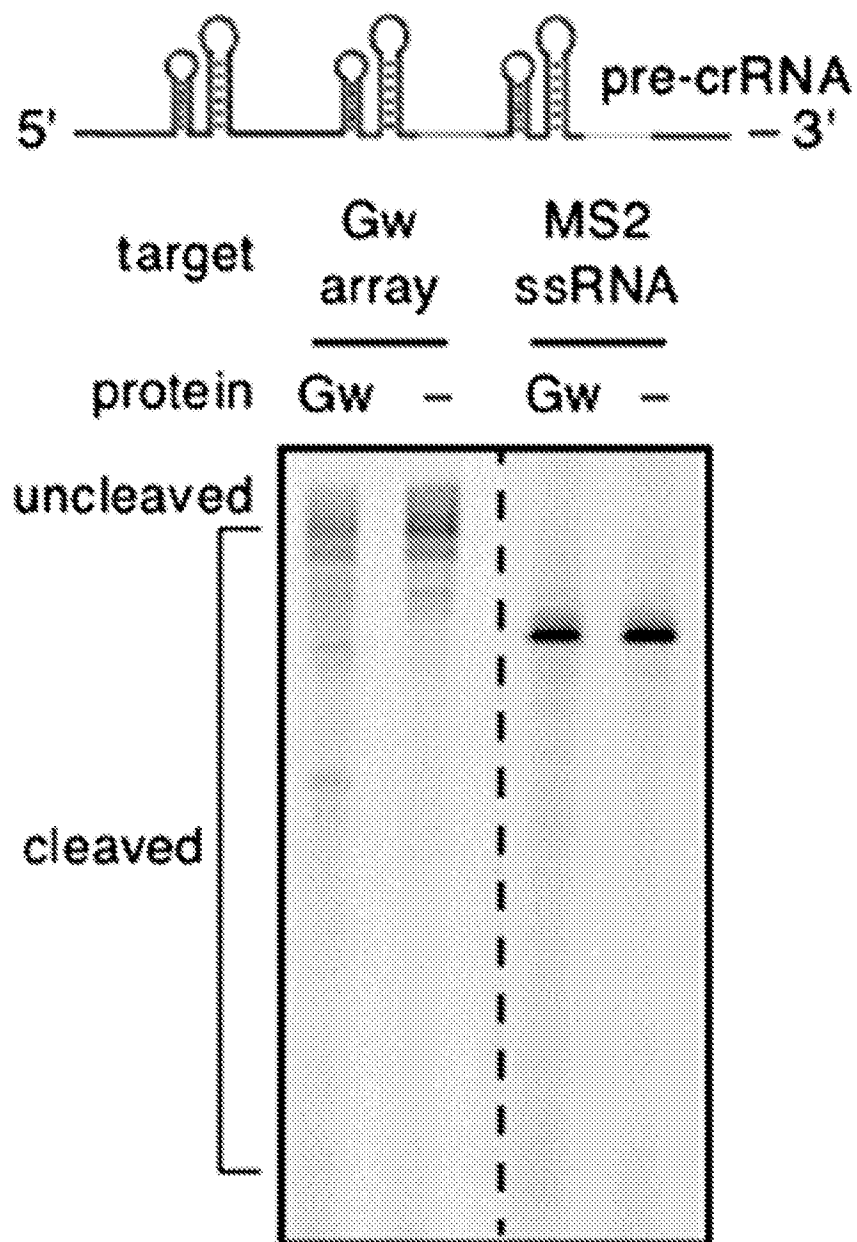
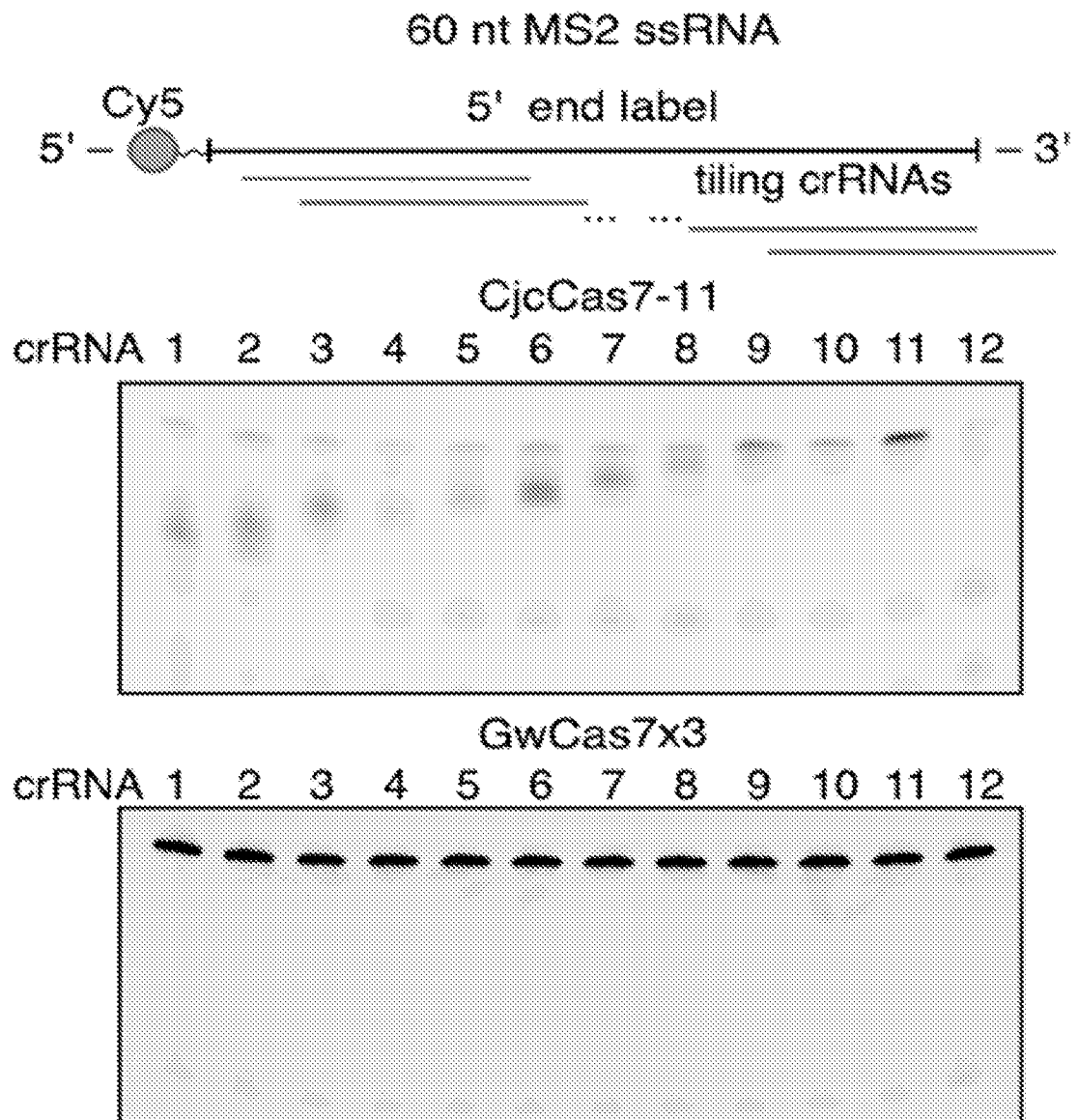
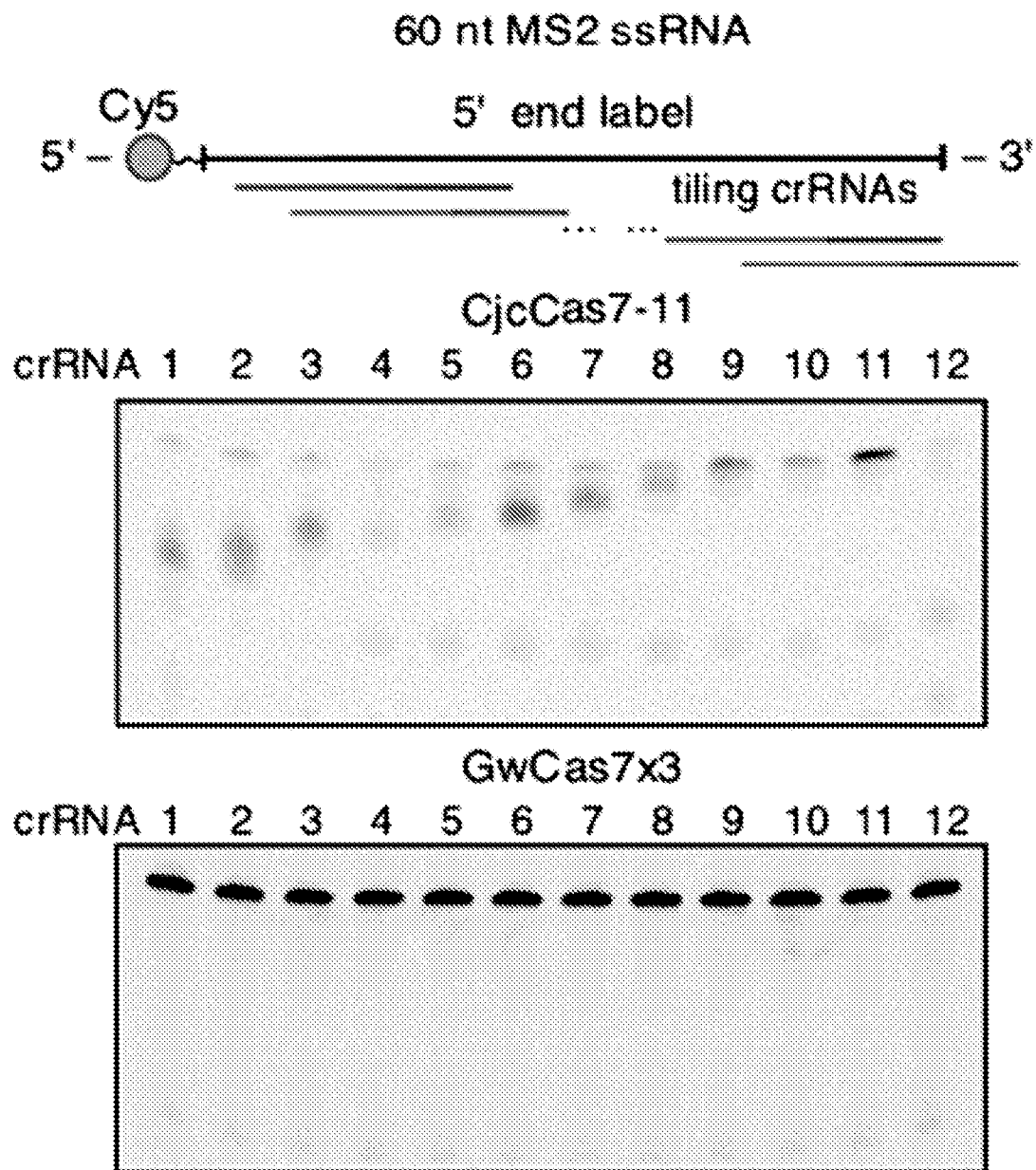


FIG. 7D

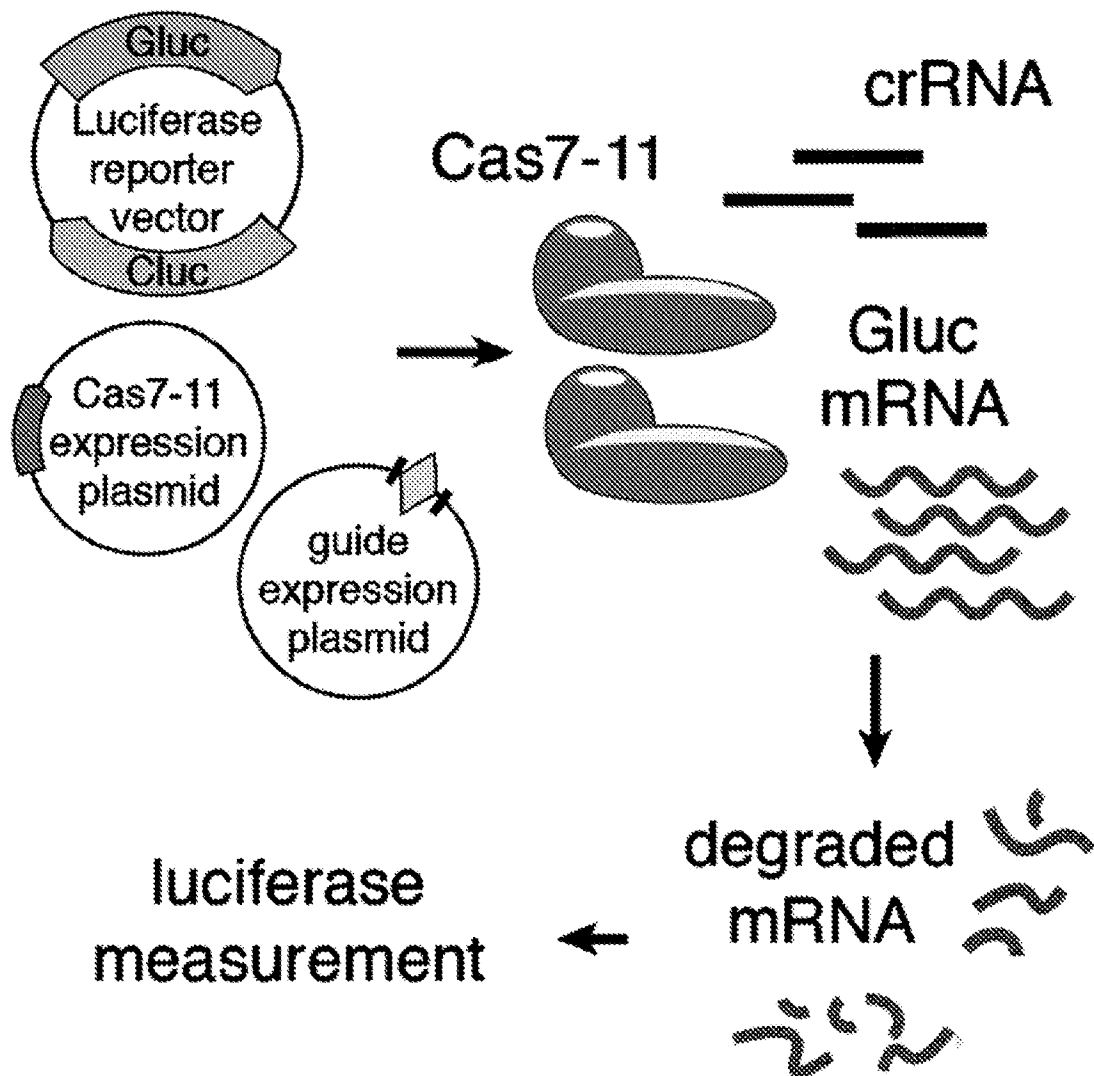
40/142

**FIG. 7E**

41/142

**FIG. 7F**

42/142

**FIG. 8A**

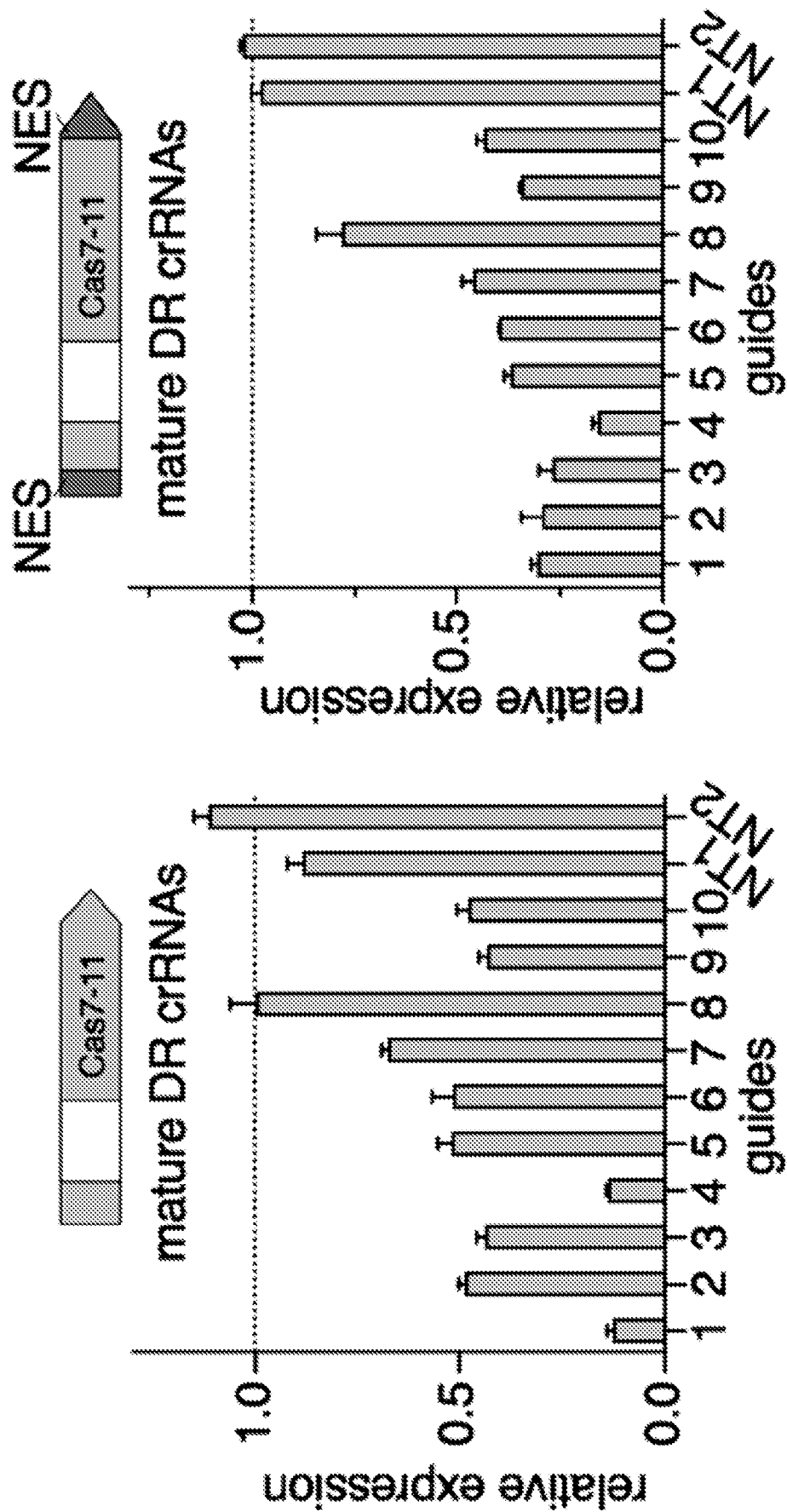


FIG. 8B

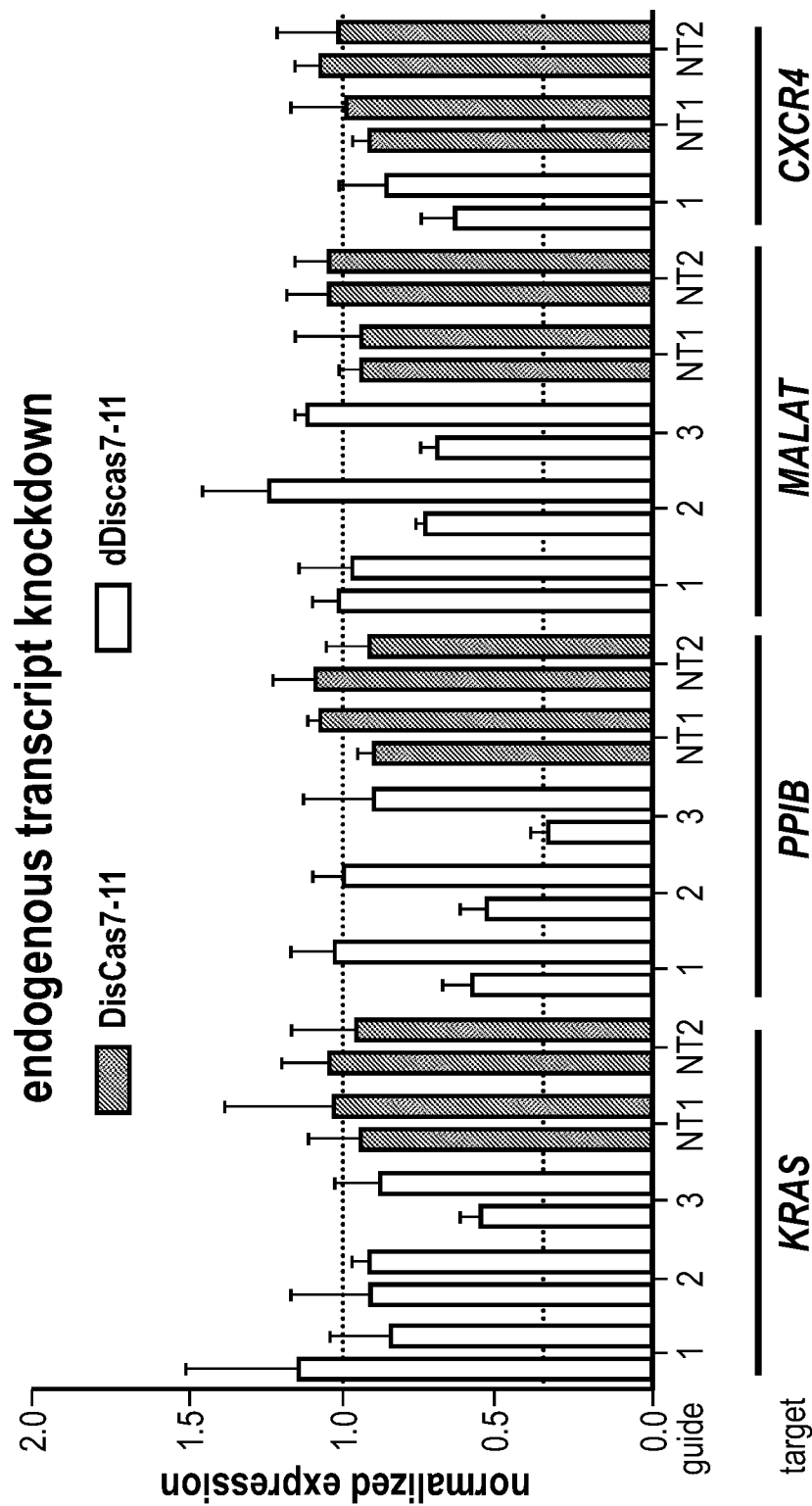


FIG. 8C

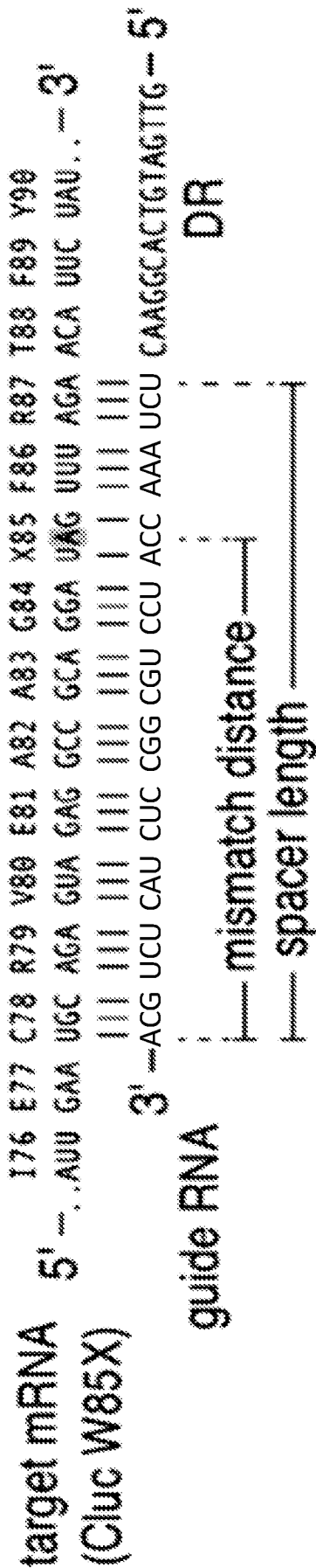
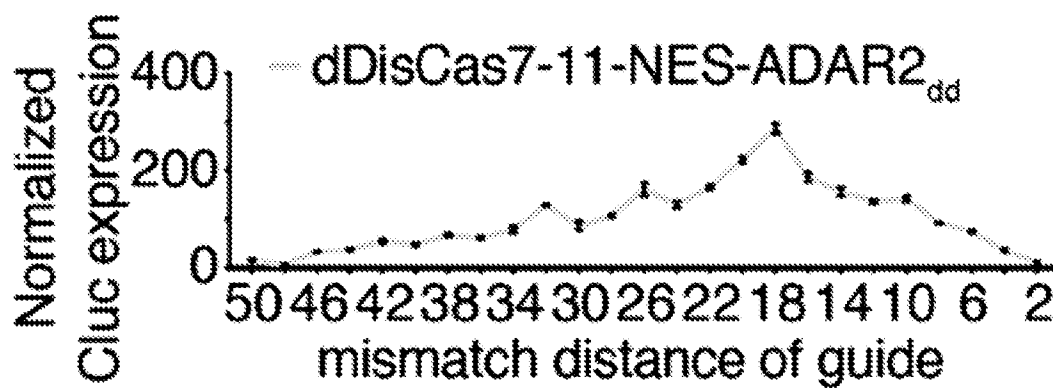
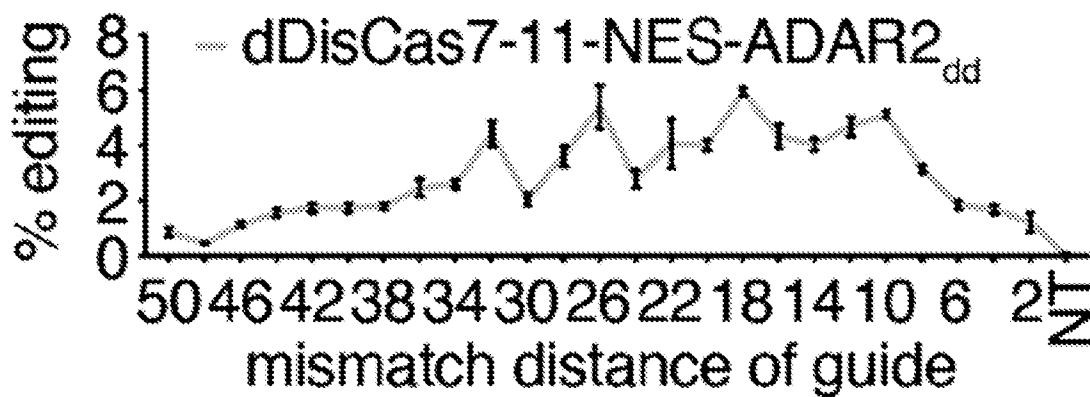
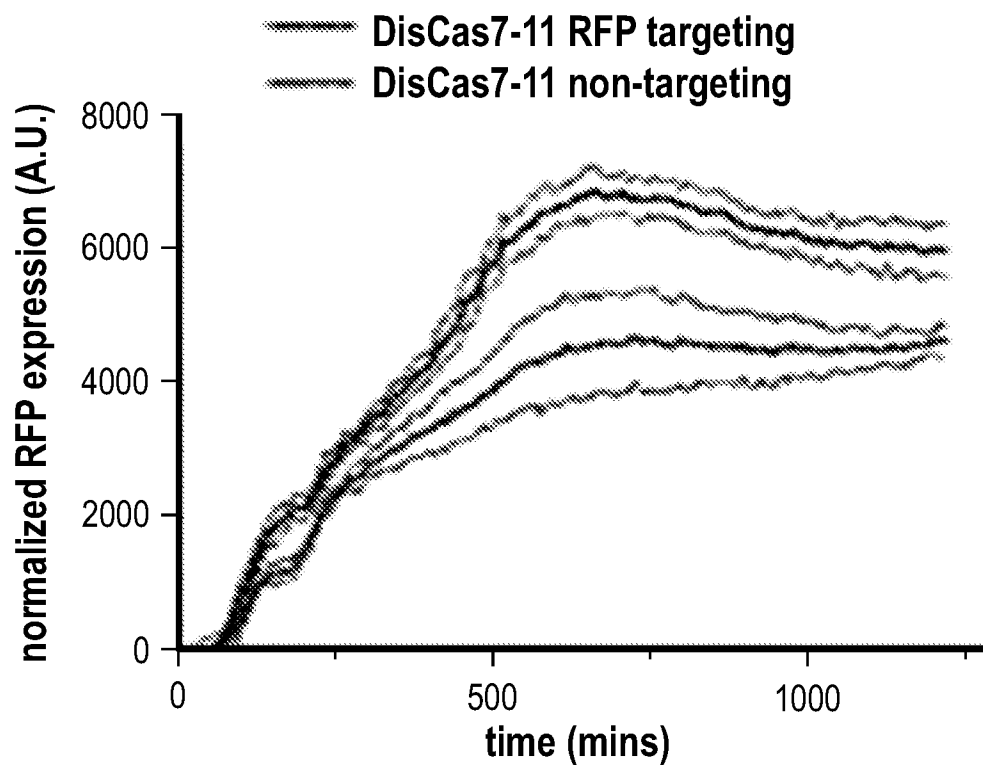
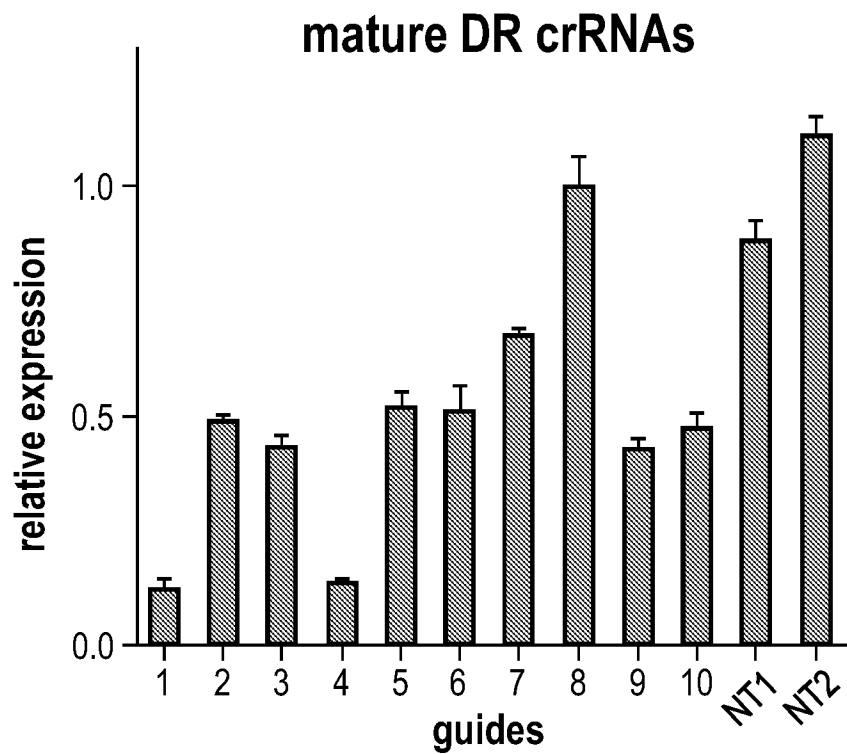


FIG. 8D

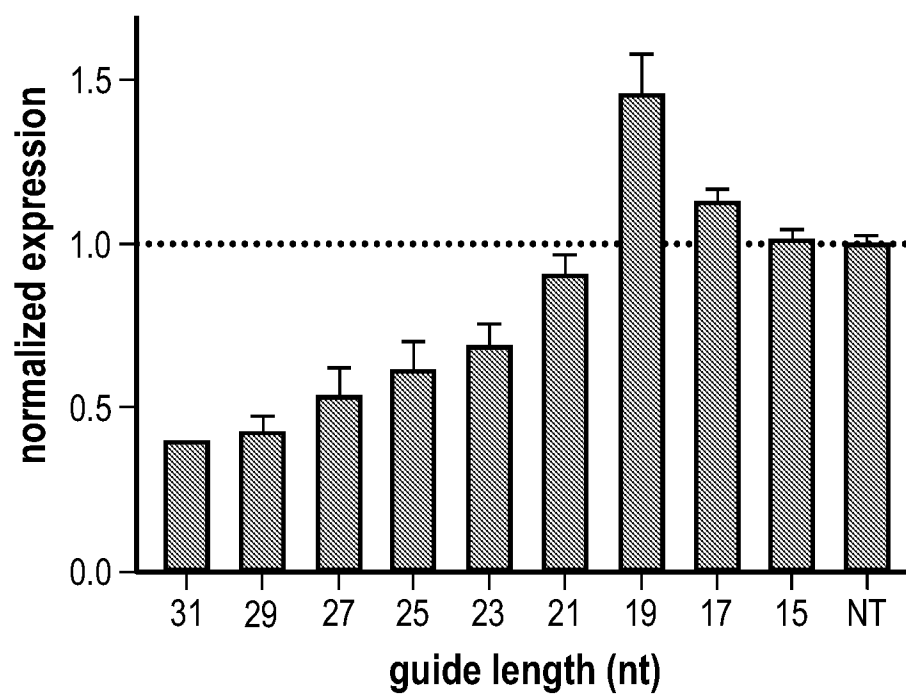
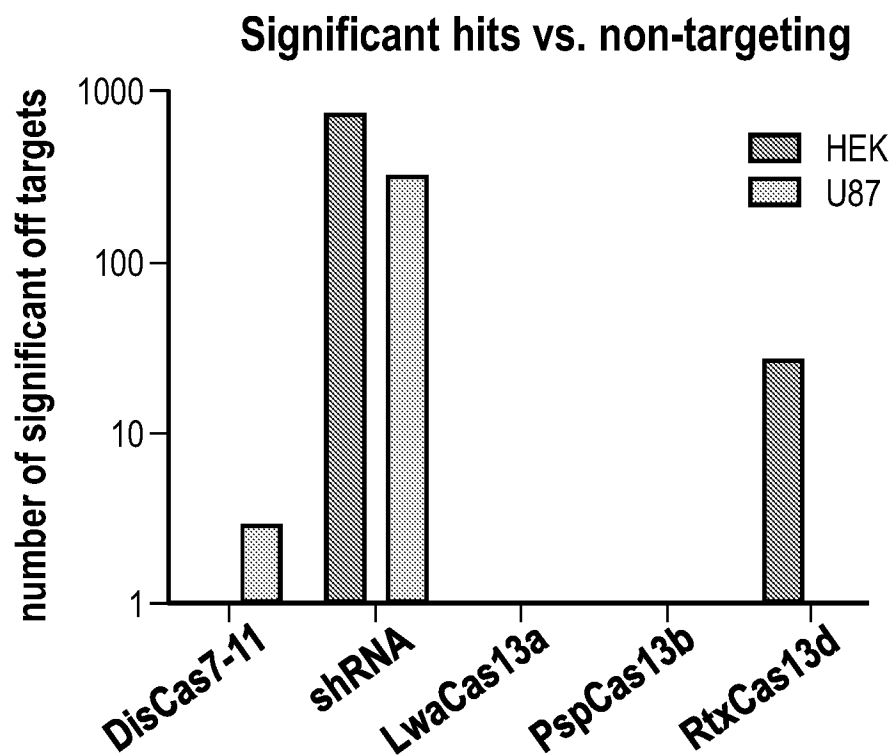
46/142

**FIG. 8E****FIG. 8F**

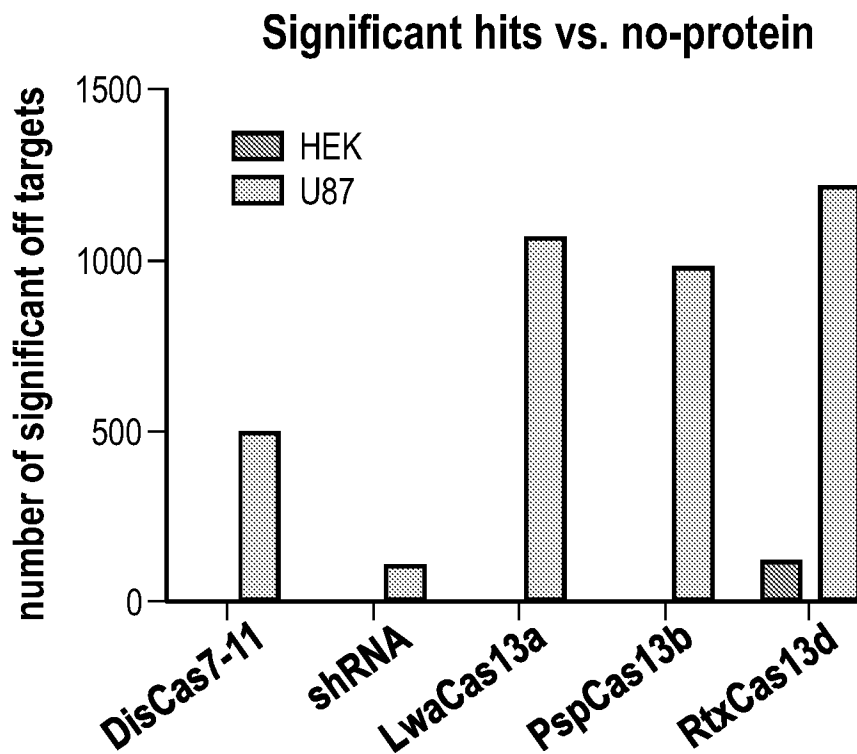
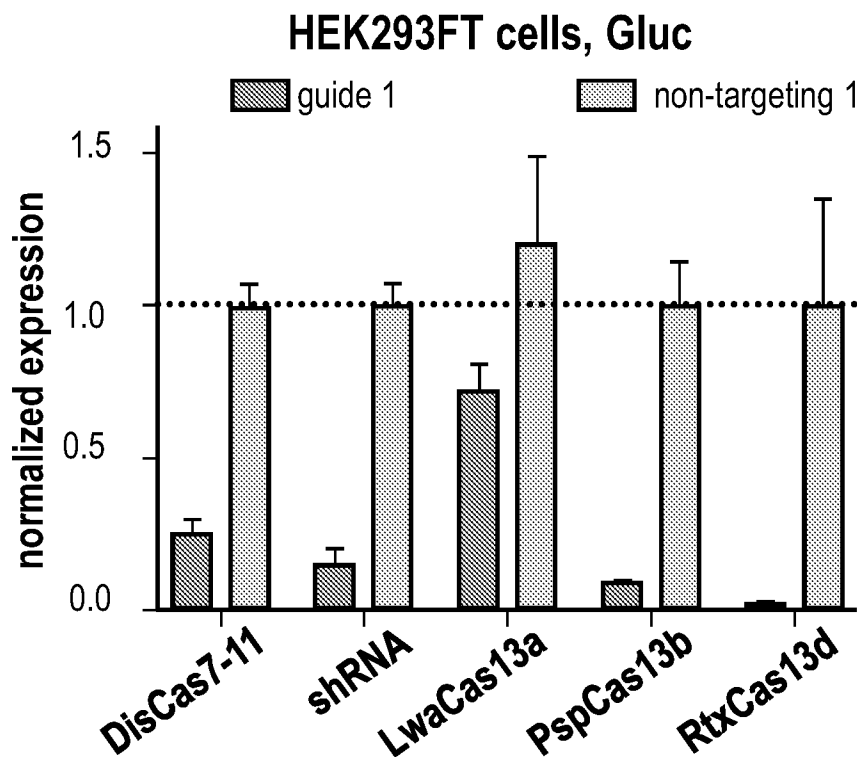
47/142

**FIG. 8G****FIG. 8H**

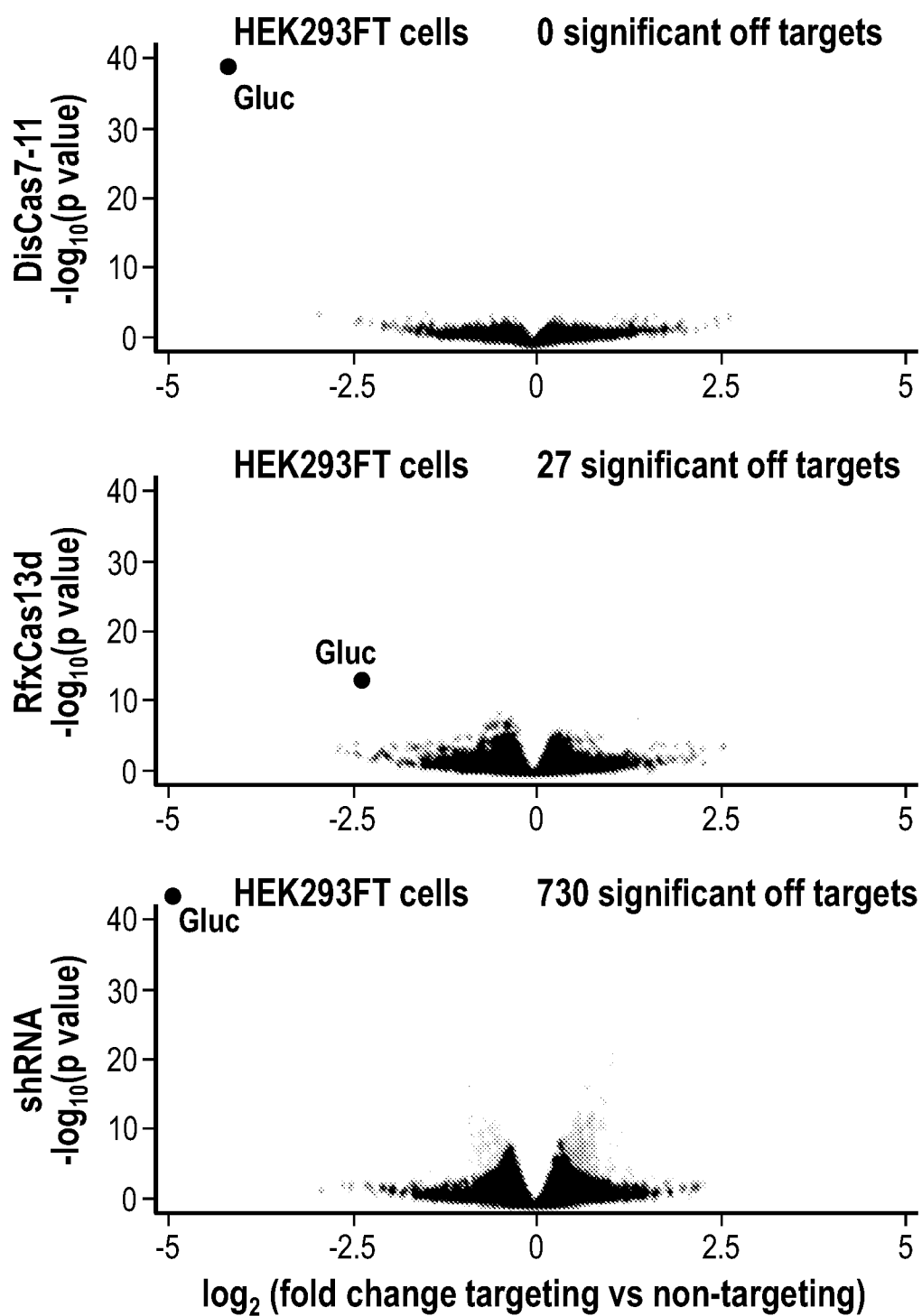
48/142

**FIG. 8I****FIG. 8J**

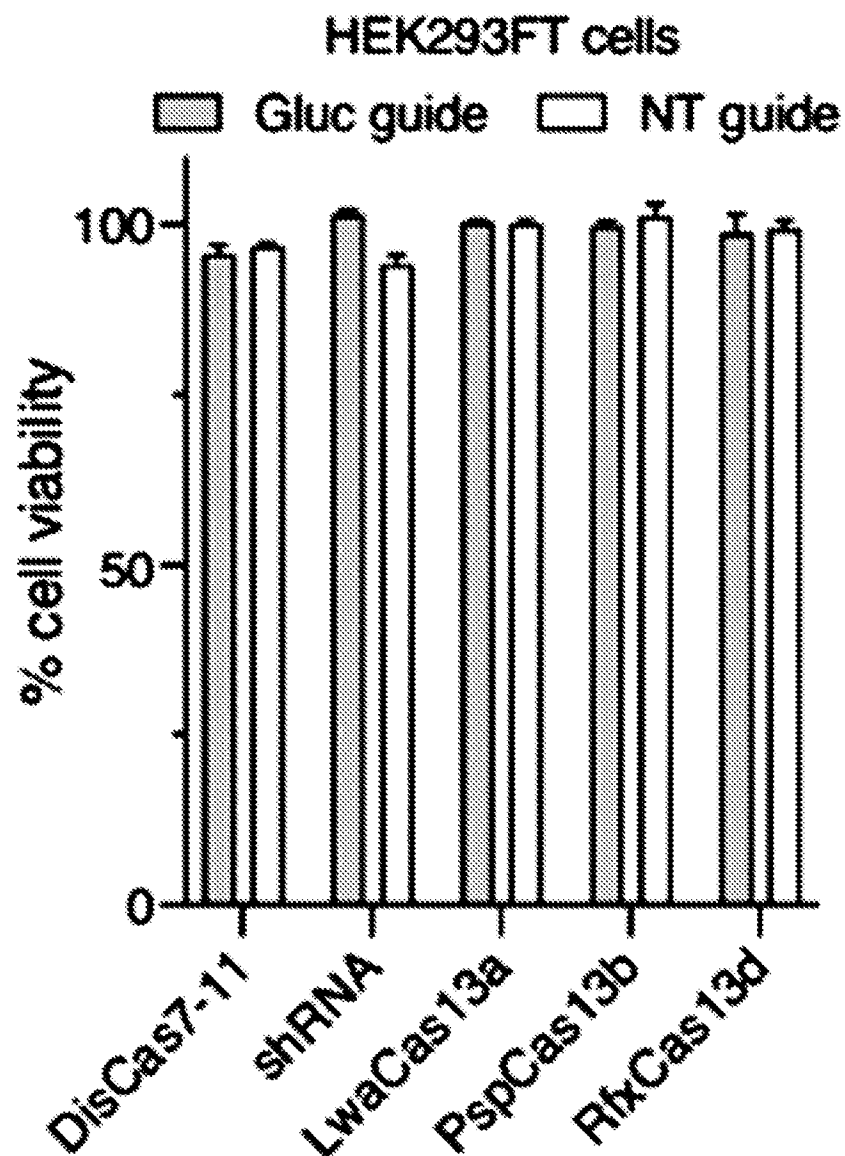
49/142

**FIG. 8K****FIG. 8L**

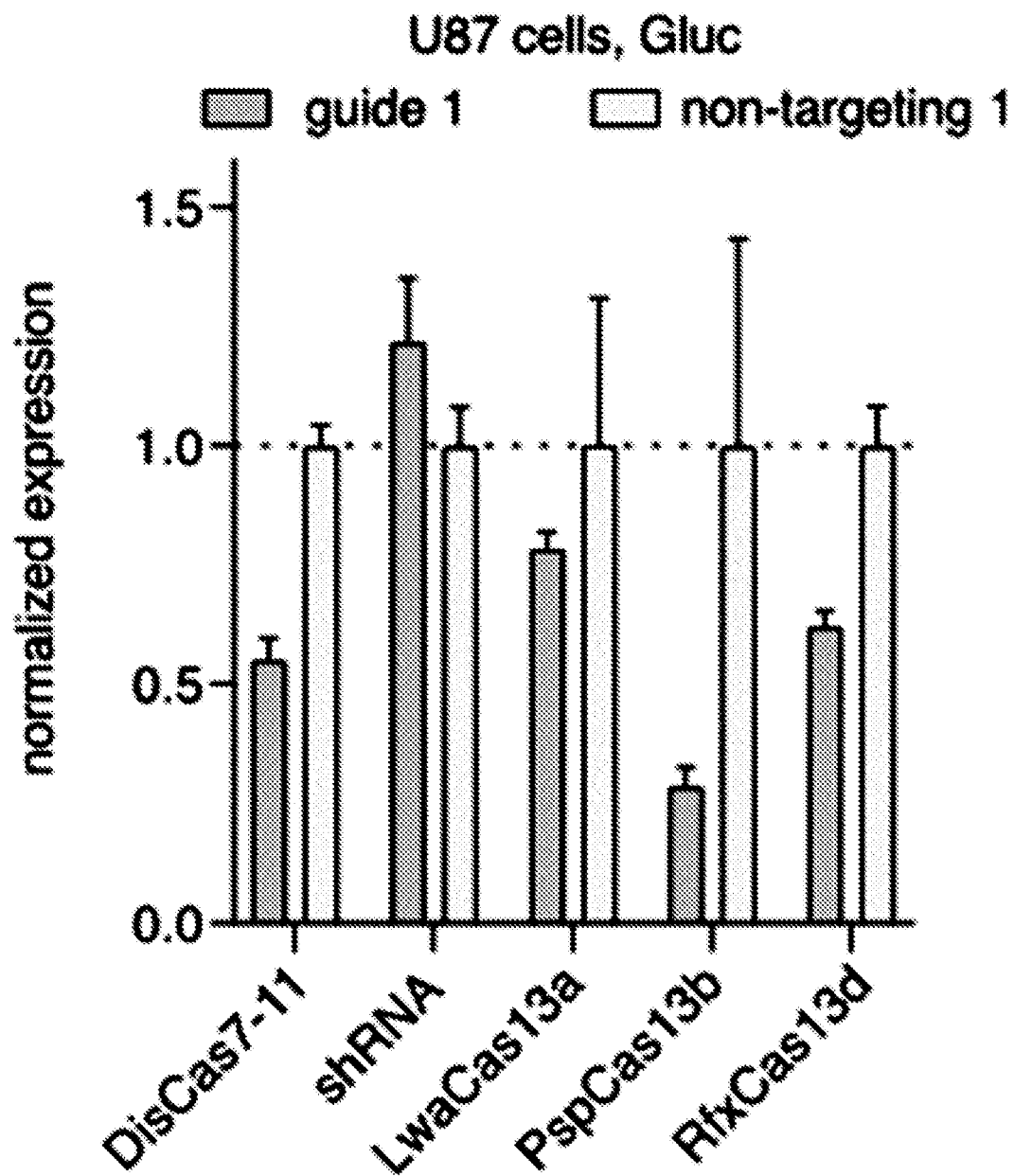
50/142

**FIG. 8M**

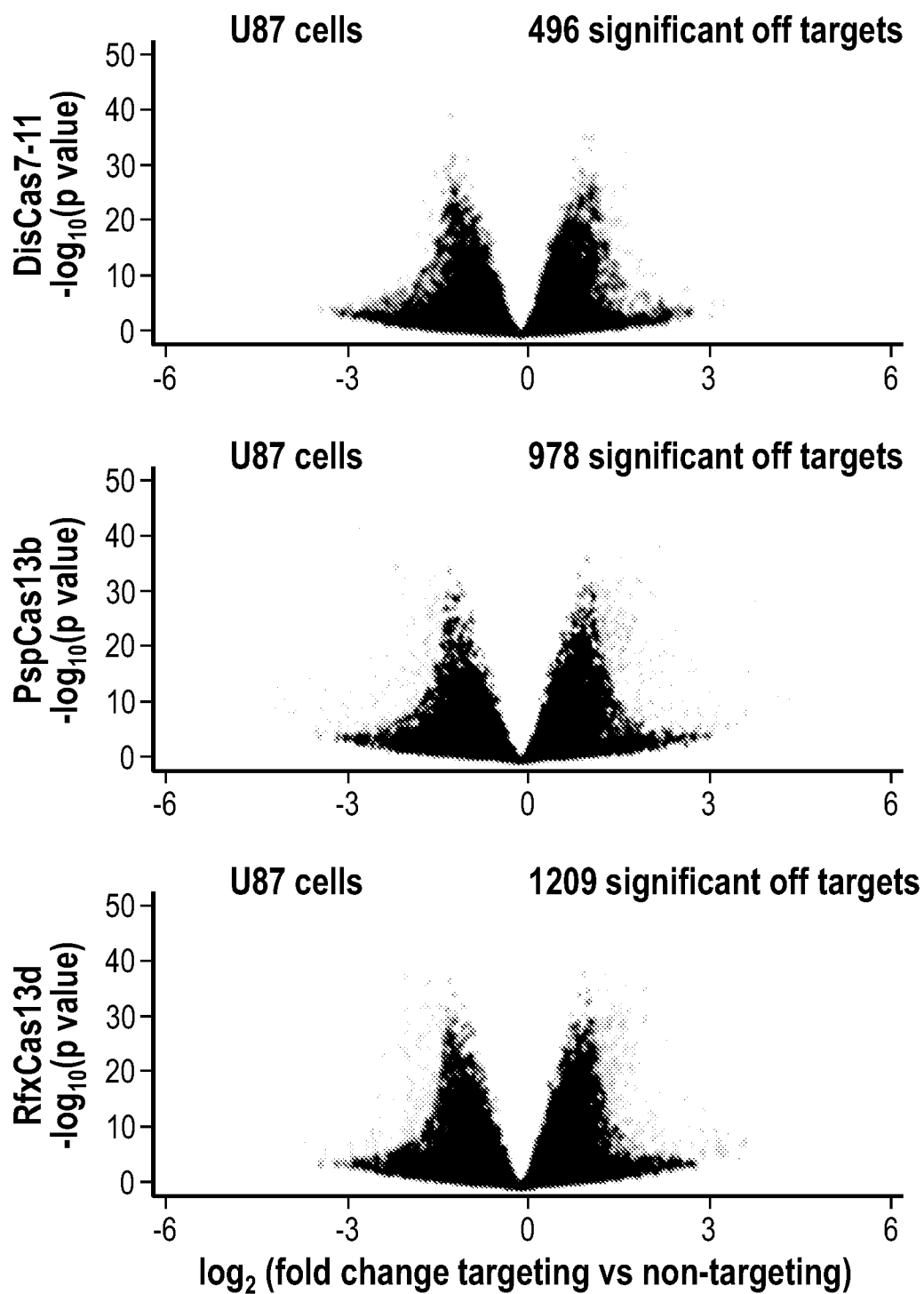
51/142

**FIG. 8N**

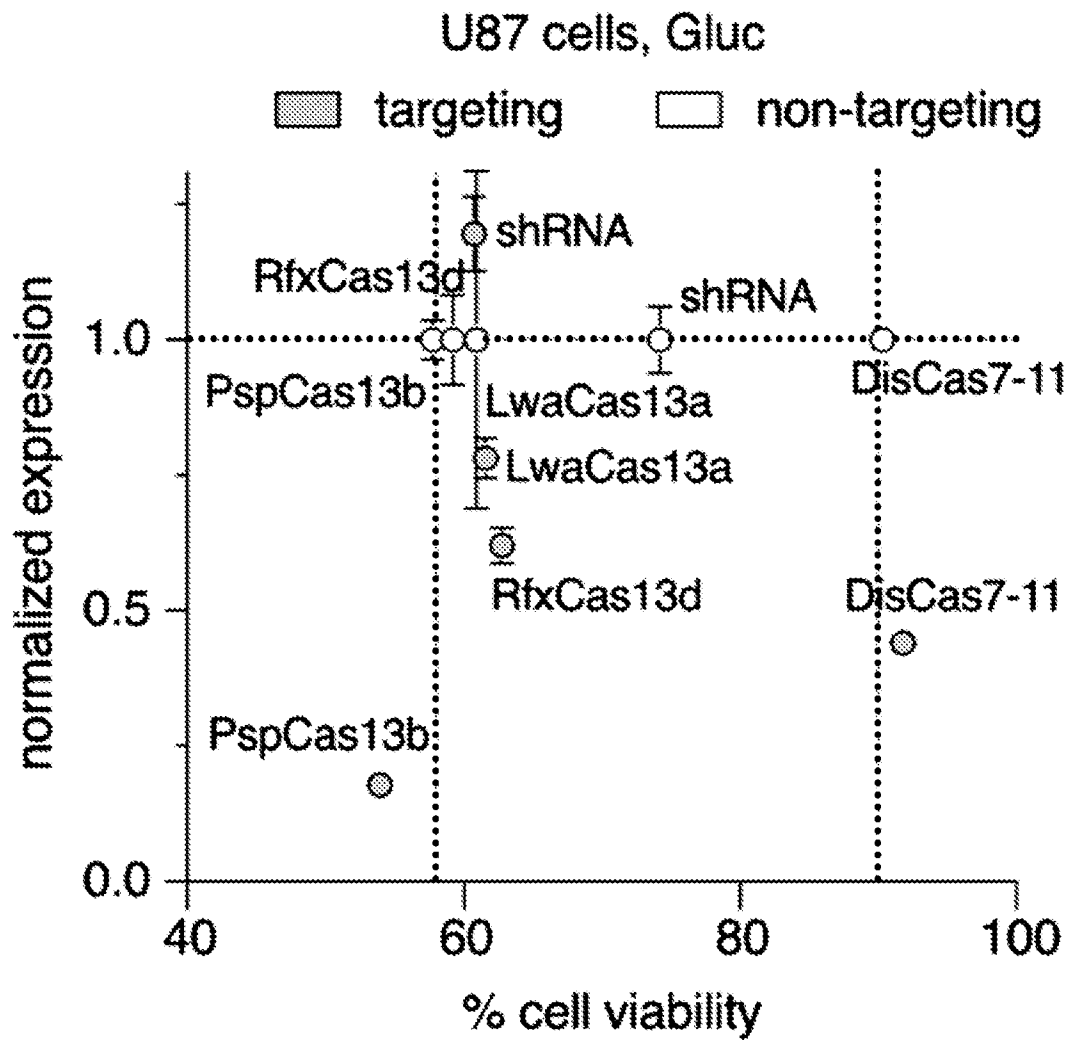
52/142

**FIG. 80**

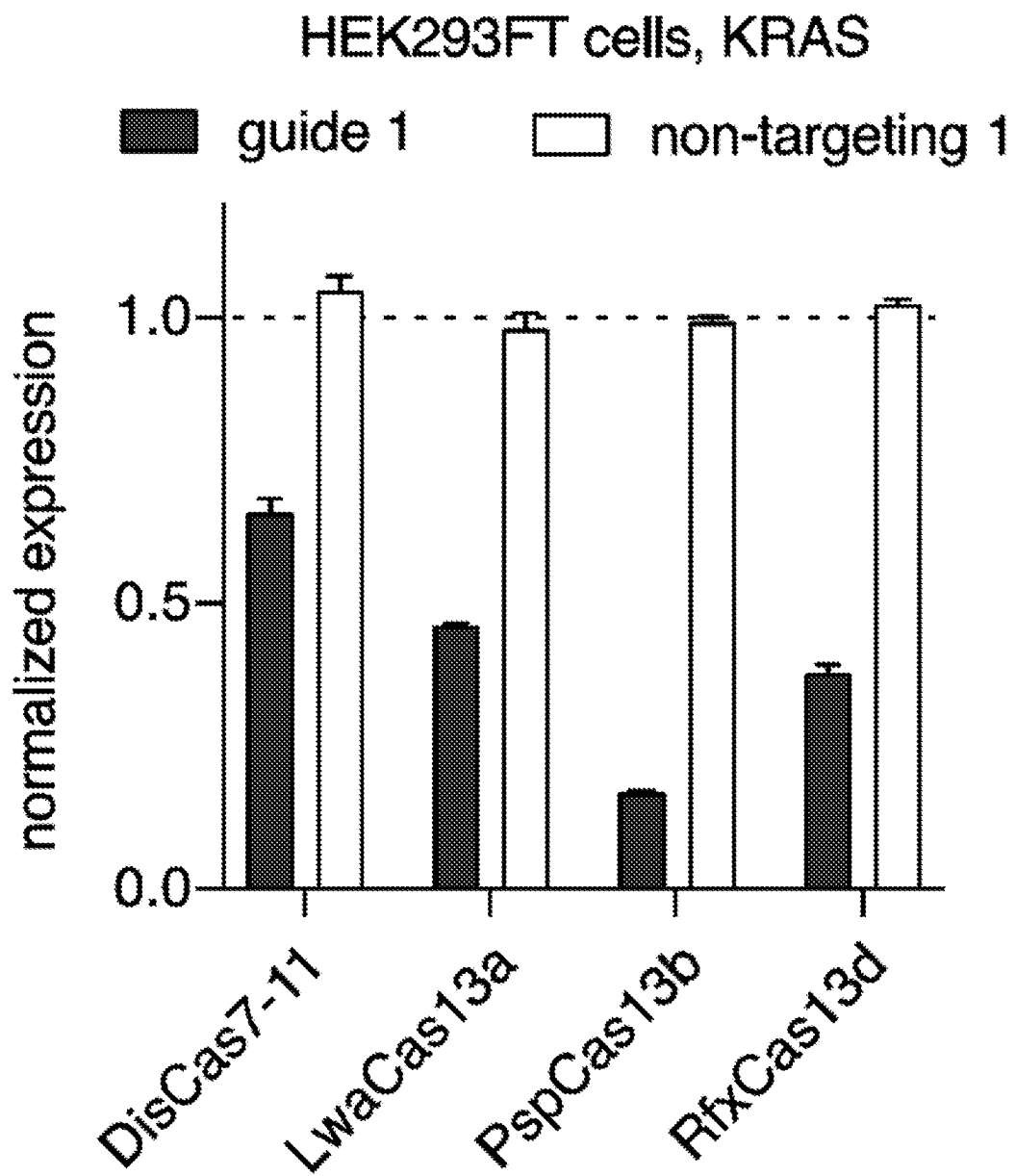
53/142

**FIG. 8P**

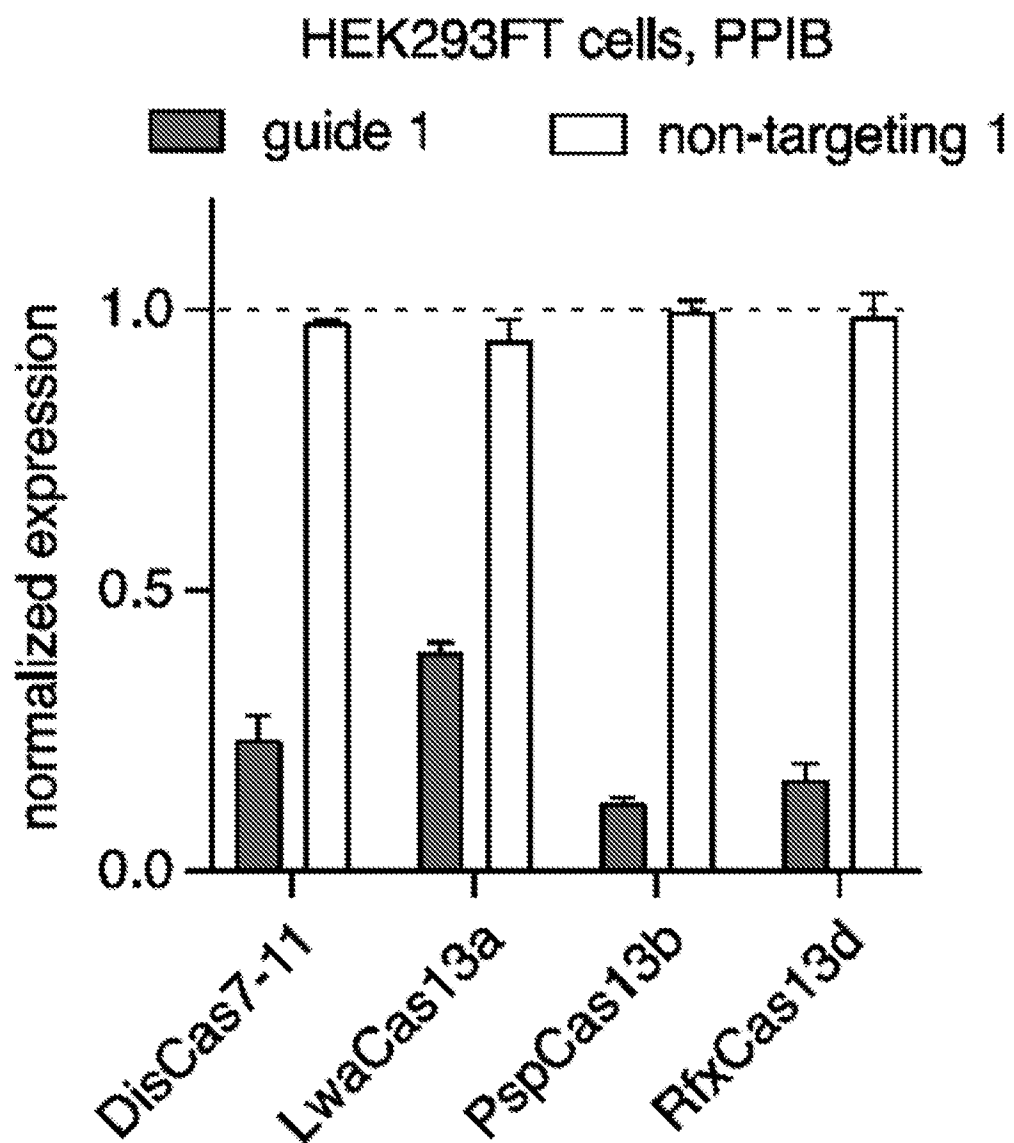
54/142

**FIG. 8Q**

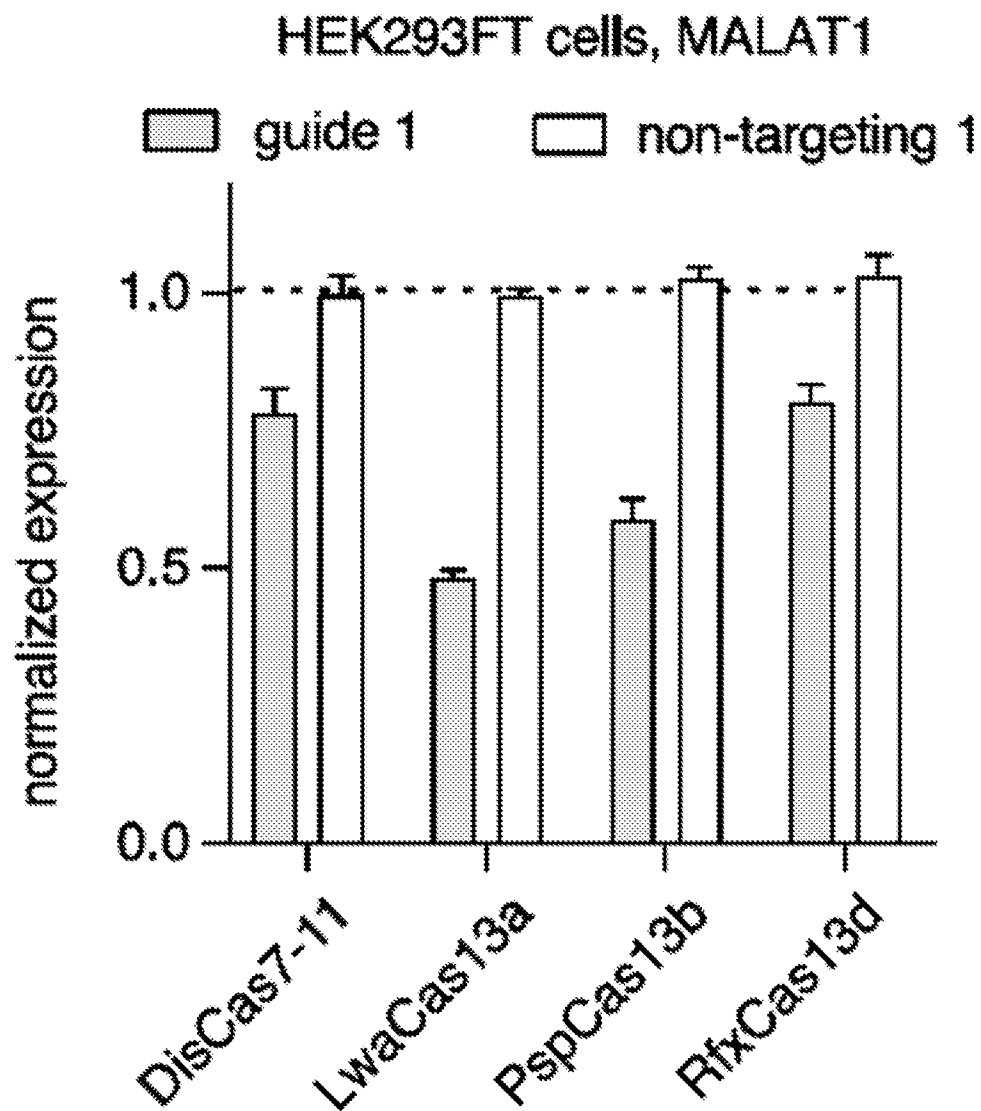
55/142

**FIG. 8R**

56/142

**FIG. 8S**

57/142

**FIG. 8T**

58/142

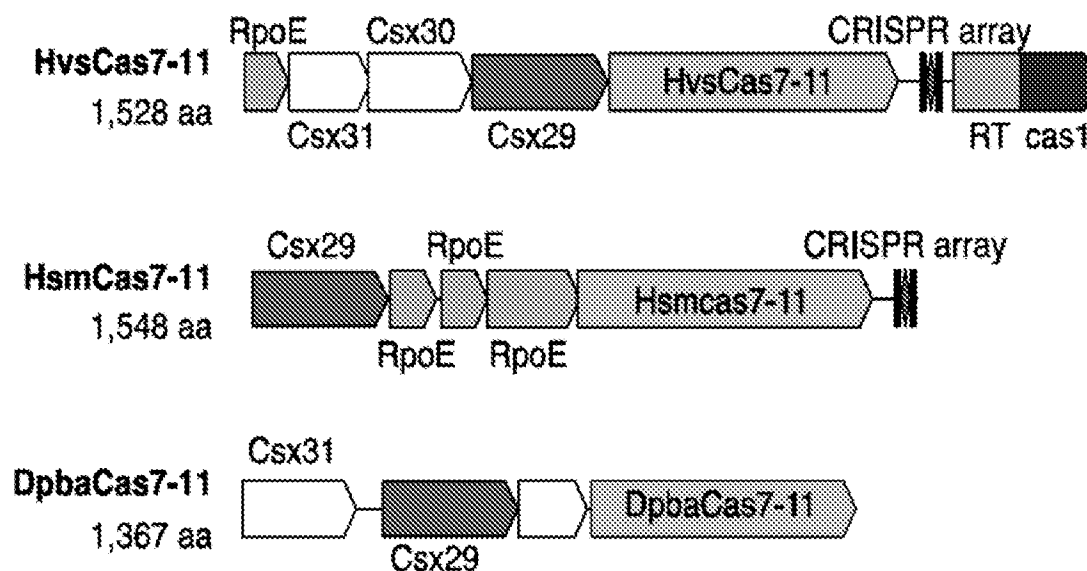


FIG. 8U

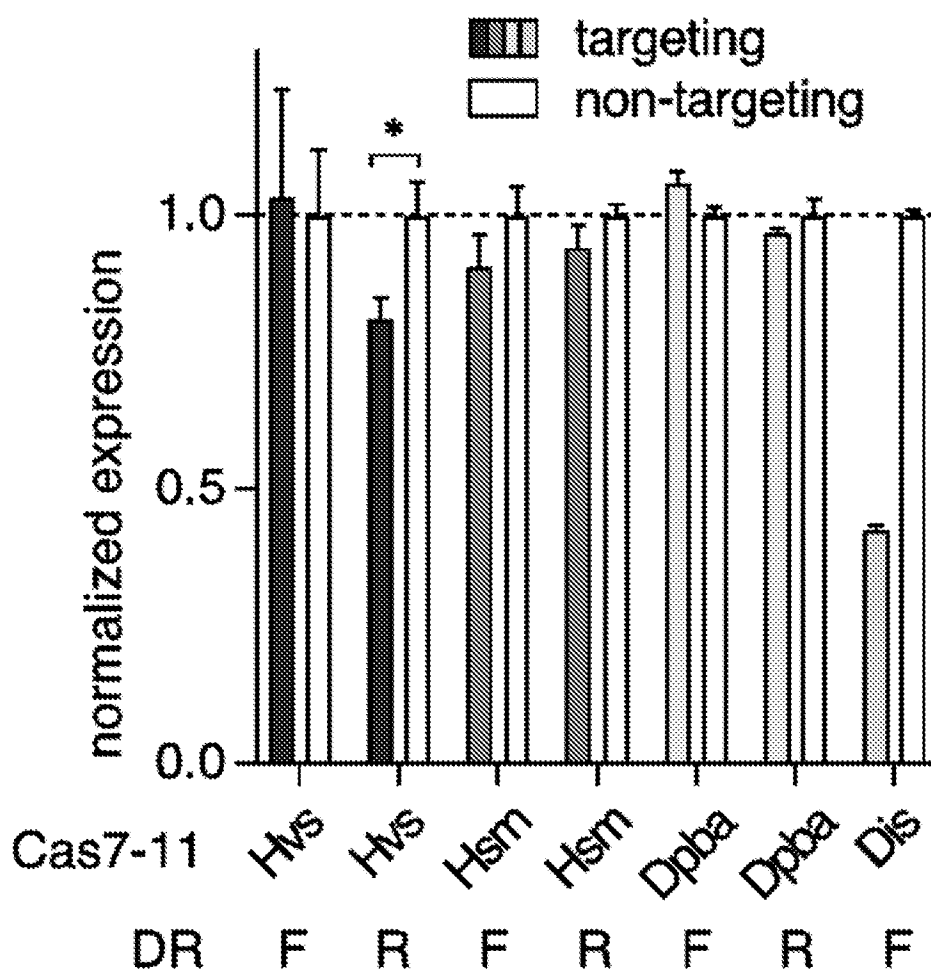
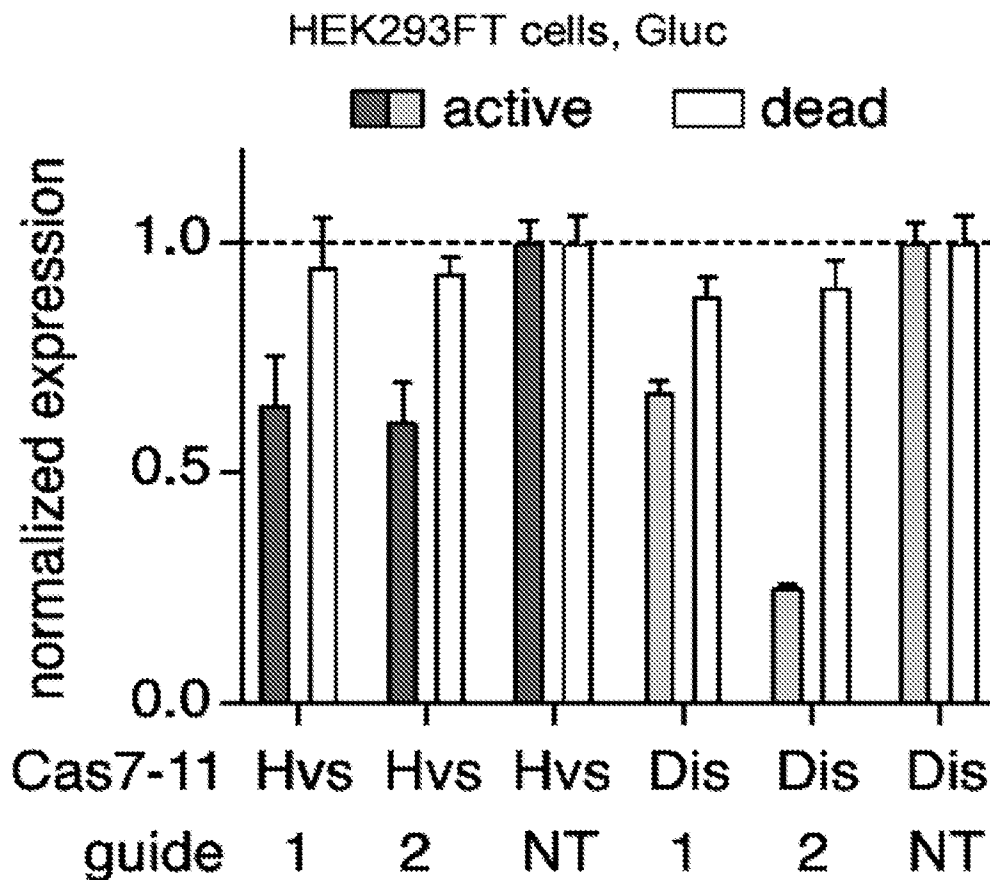
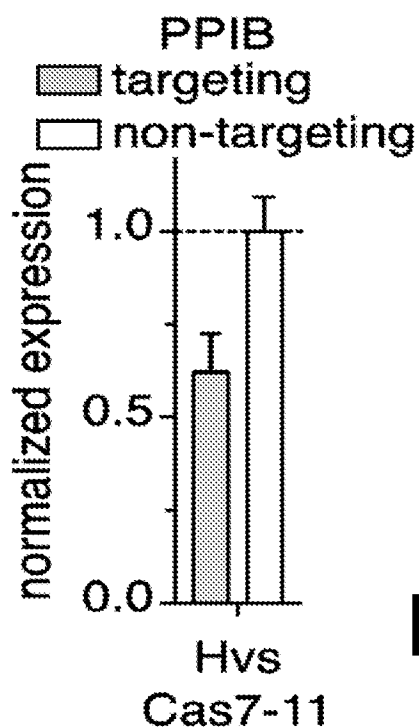
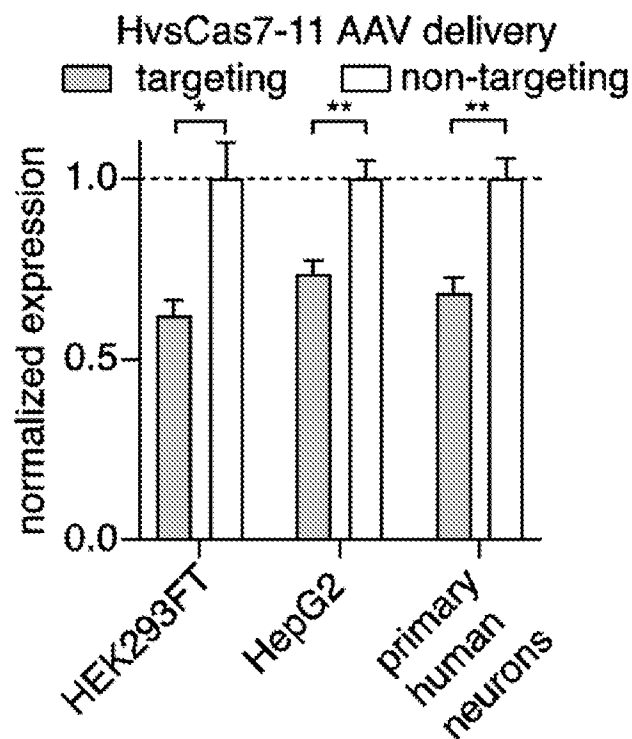
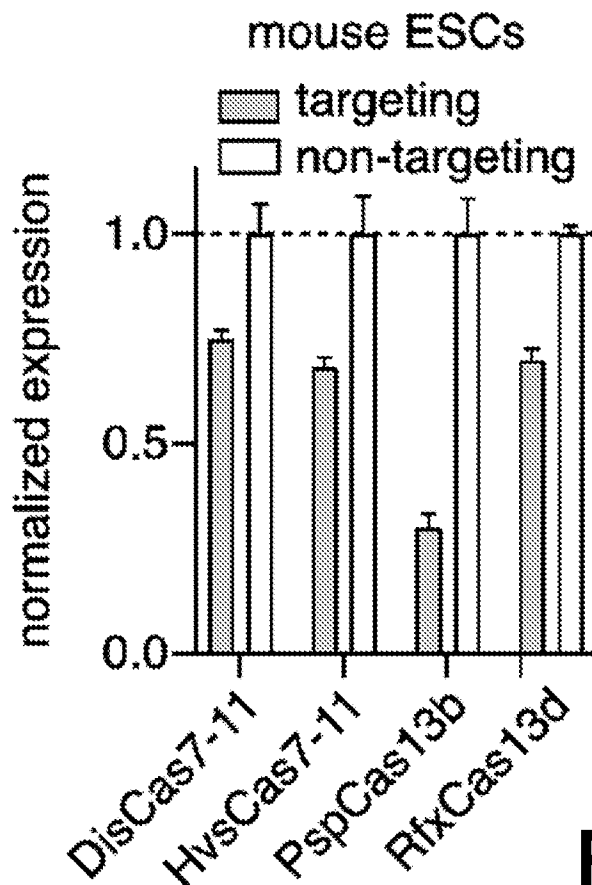


FIG. 8V

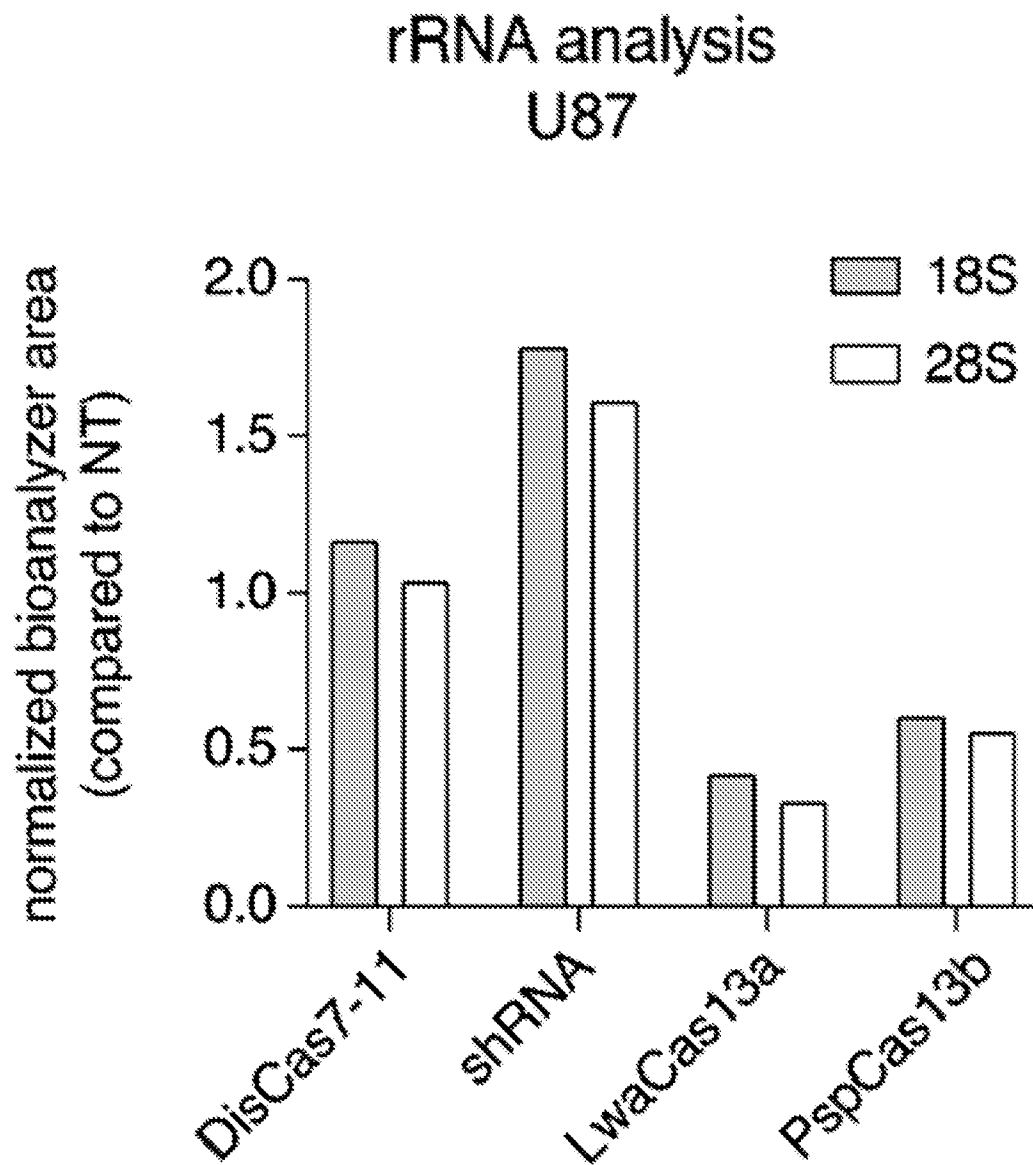
59/142

**FIG. 8W****FIG. 8X**

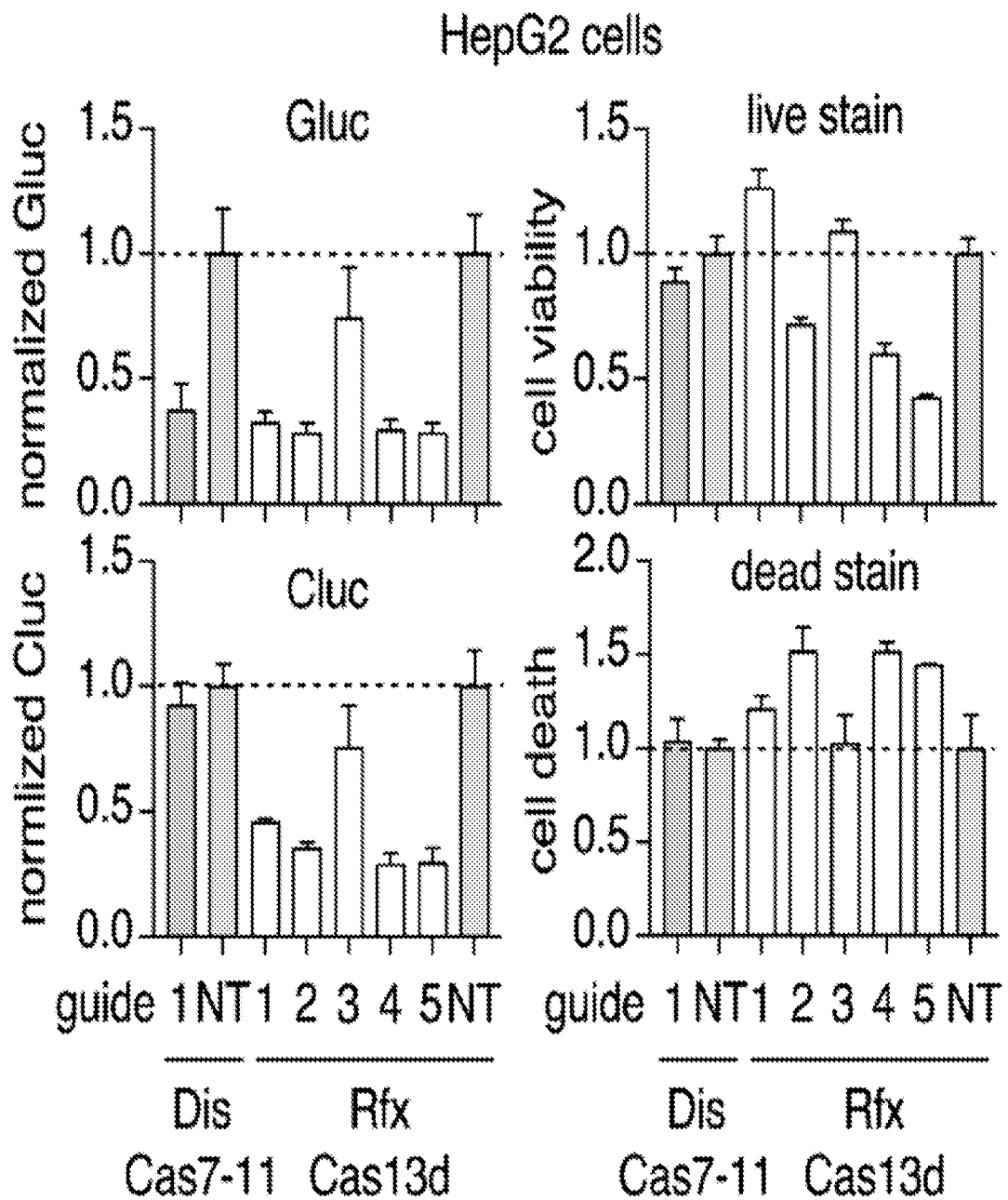
60/142

**FIG. 8Y****FIG. 8Z**

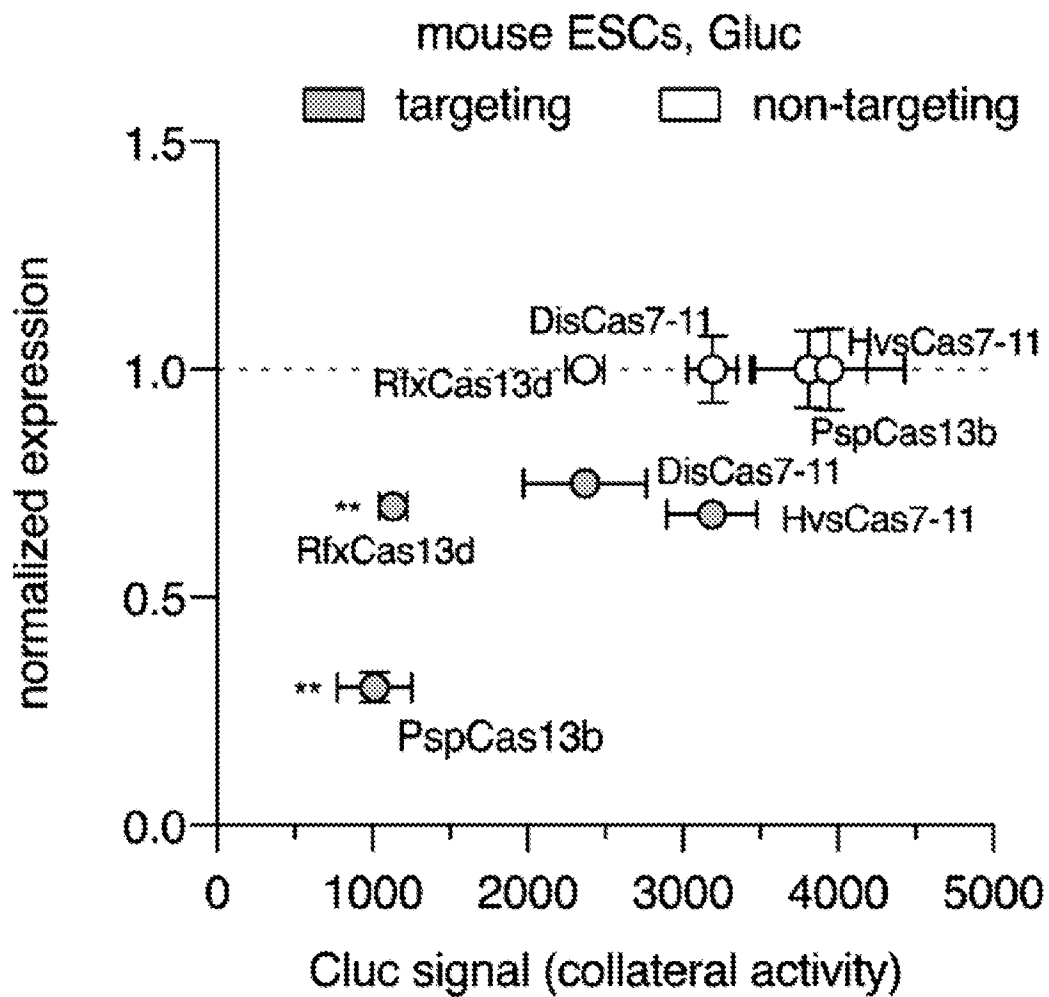
61/142

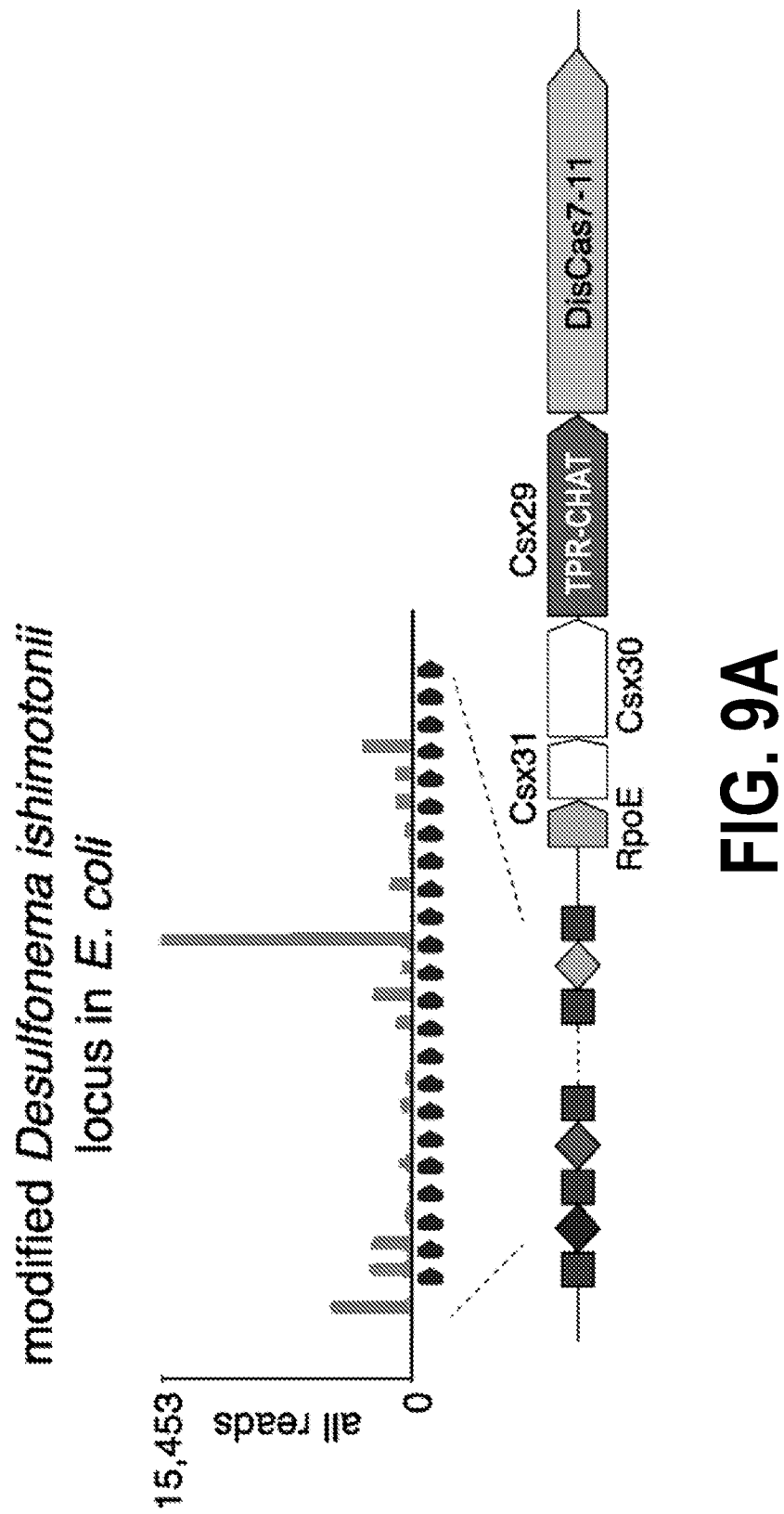
**FIG. 8AA**

62/142

**FIG. 8AB**

63/142

**FIG. 8AC**



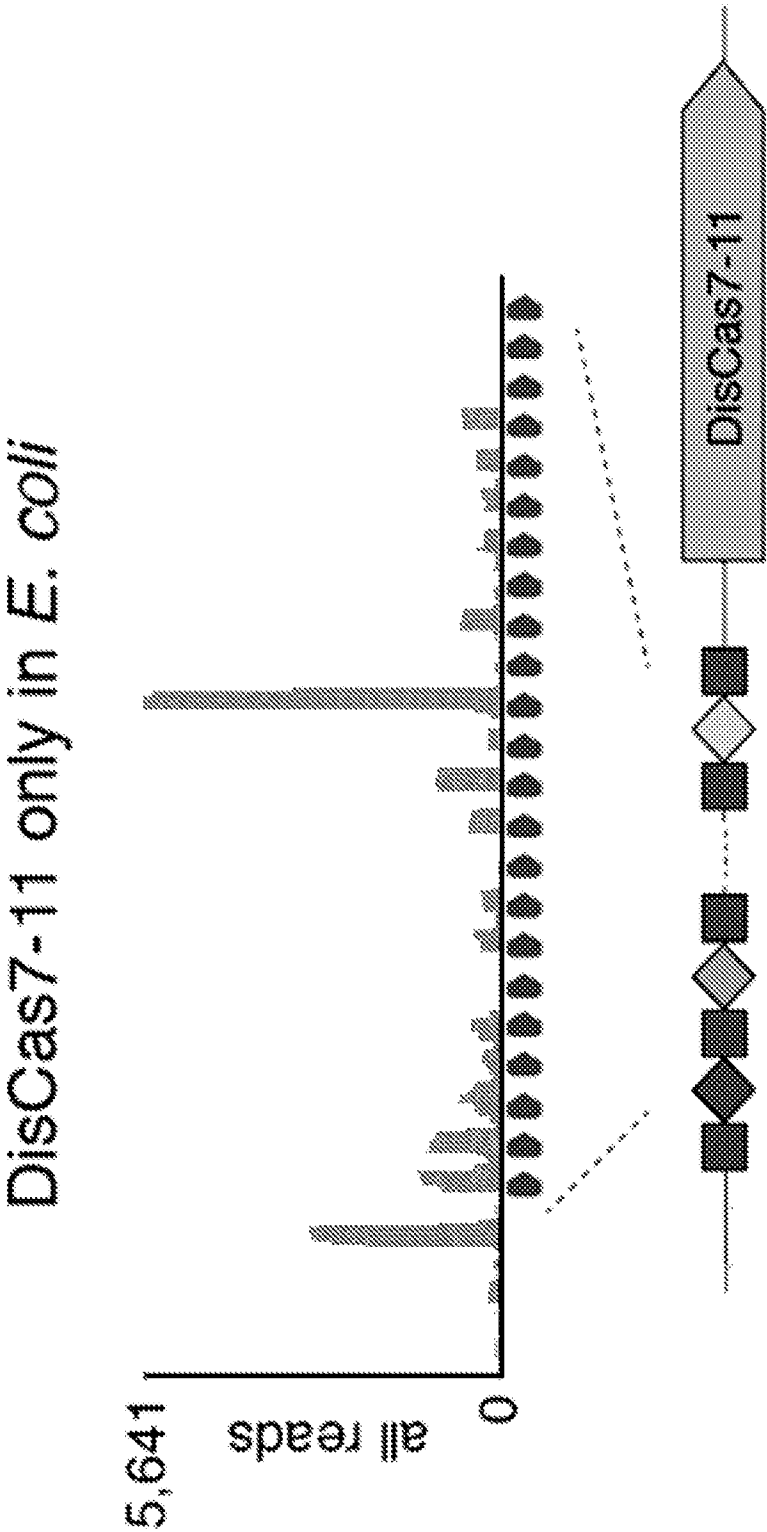
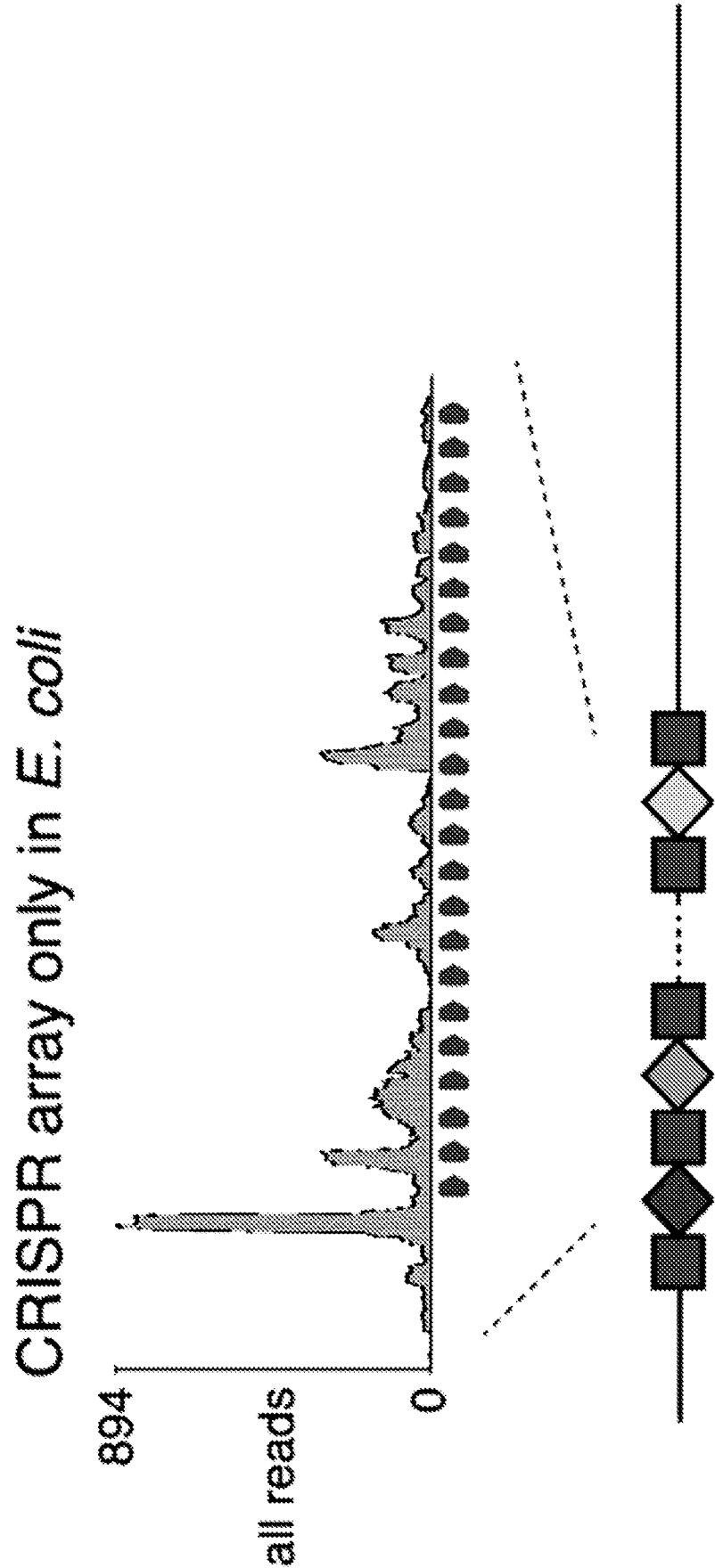


FIG. 9B



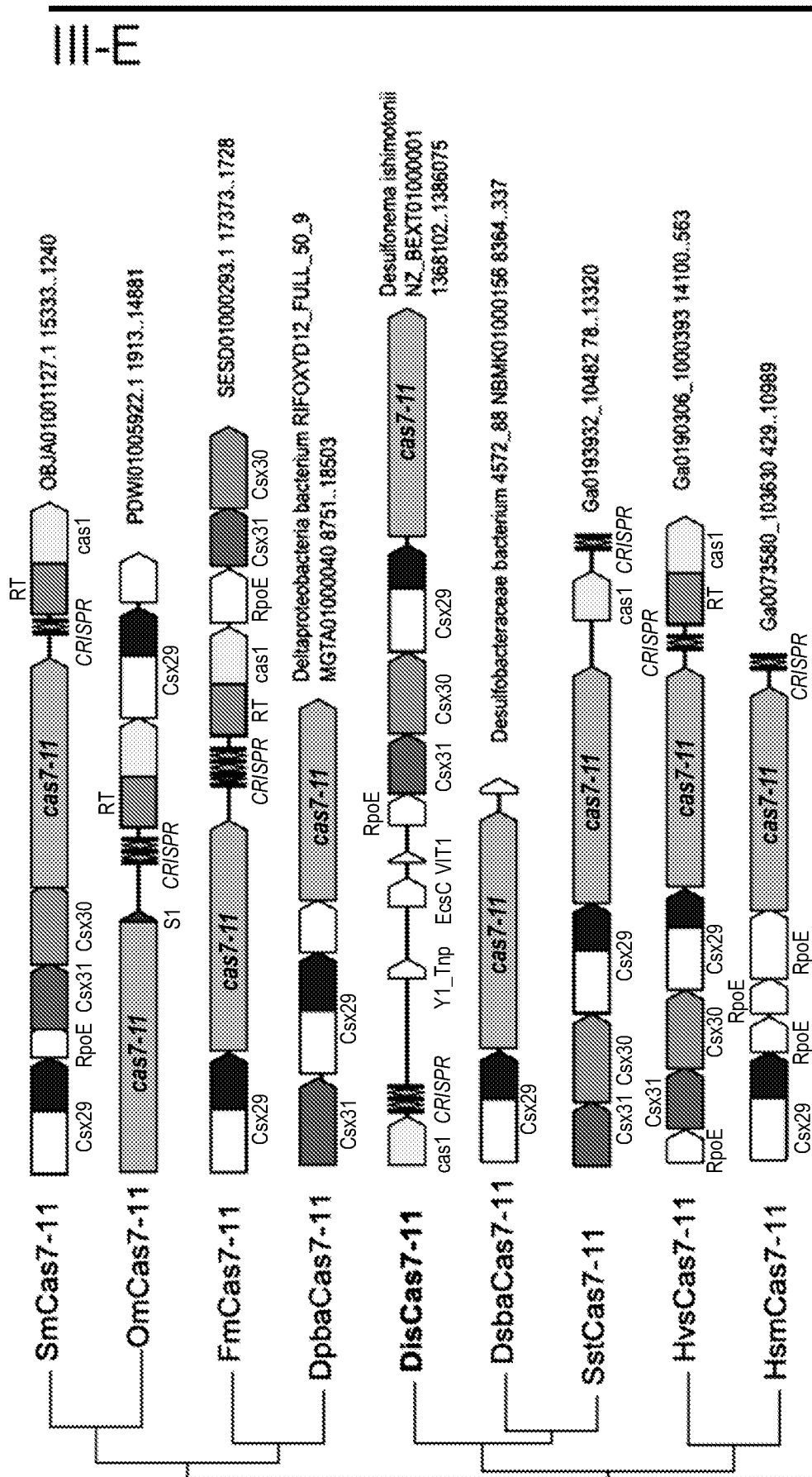


FIG. 9D

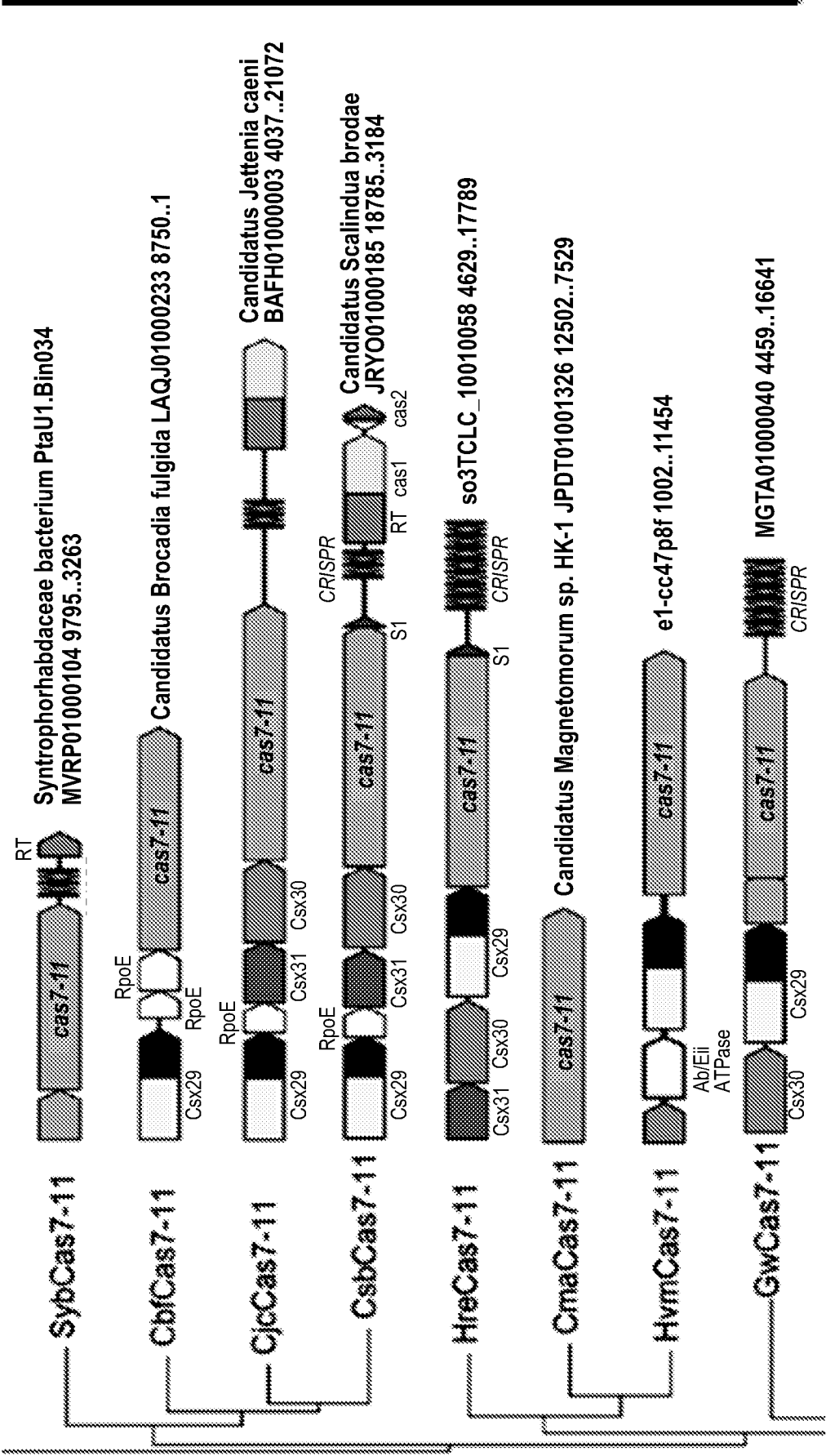


FIG. 9D (cont)

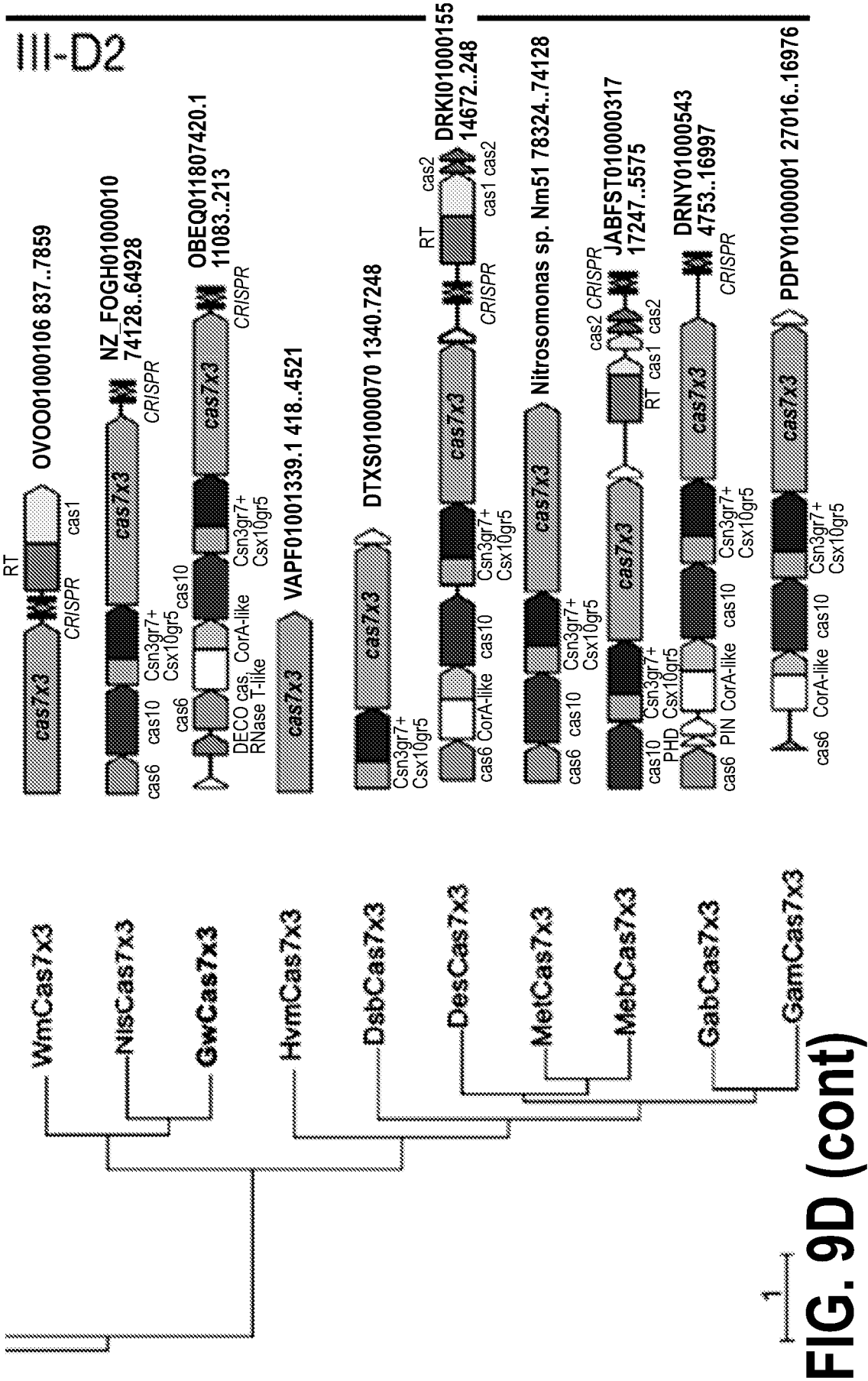
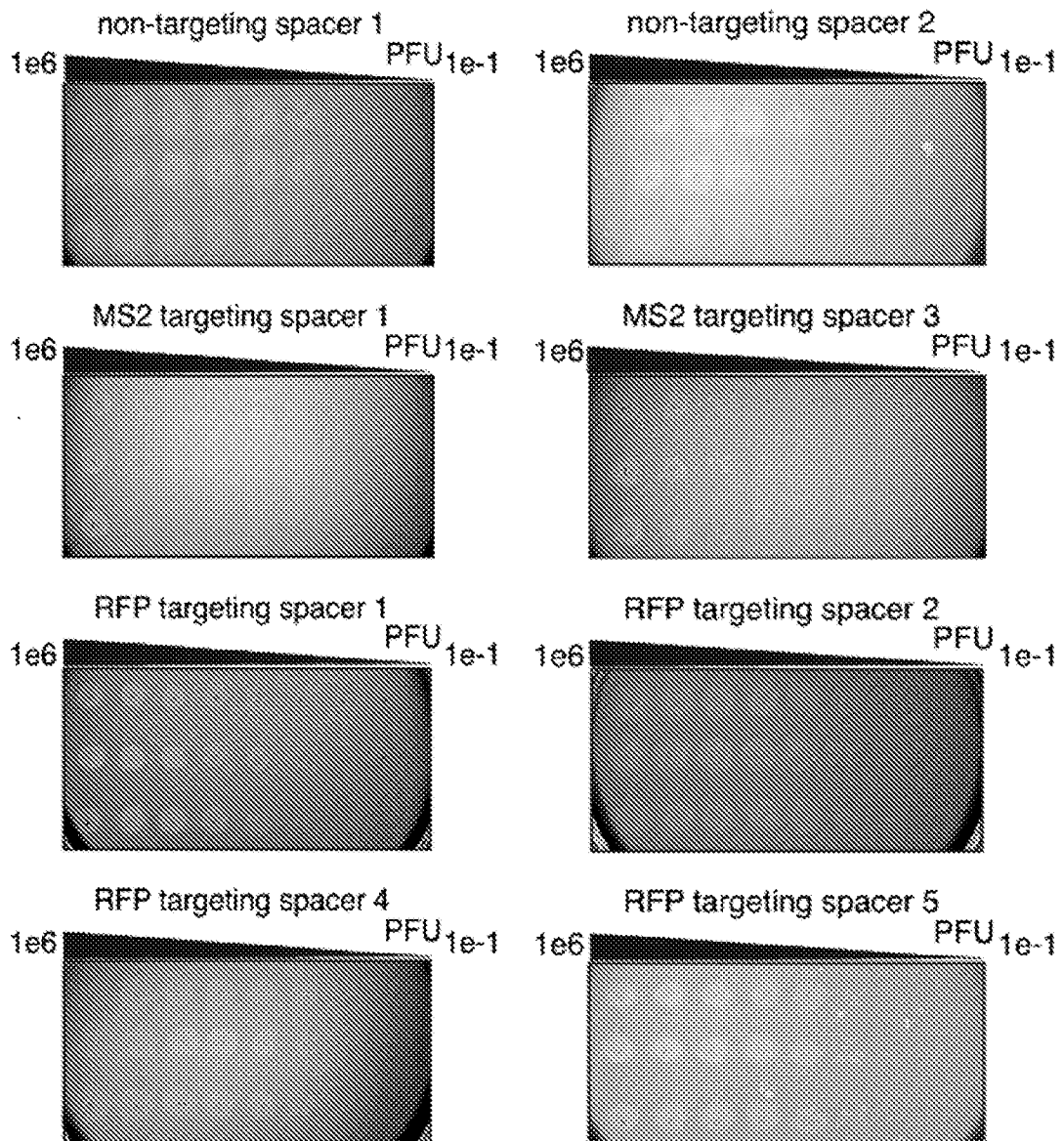


FIG. 9D (cont)

70/142

**FIG. 10**

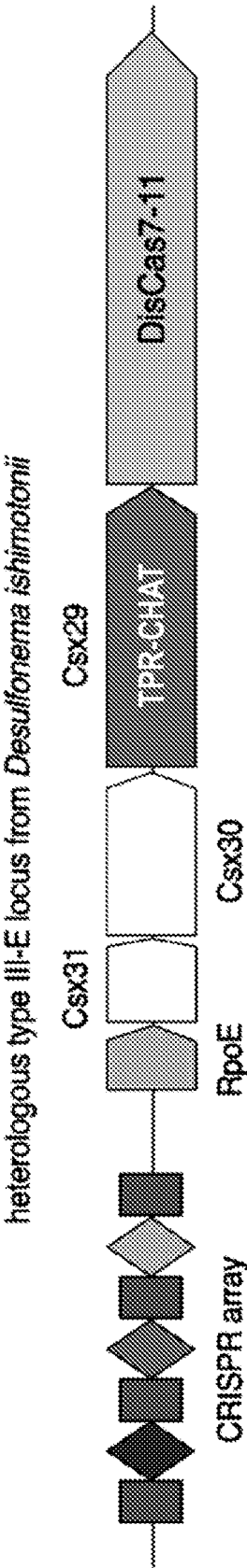


FIG. 11A

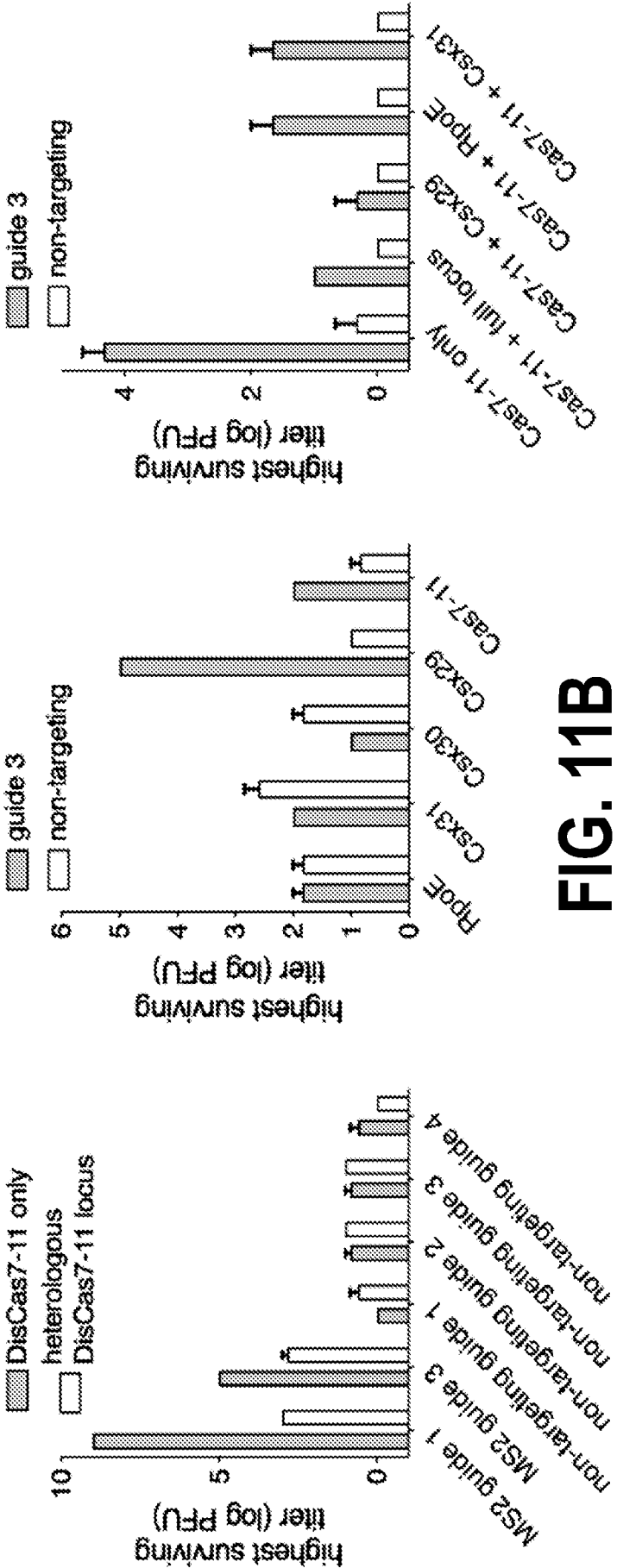
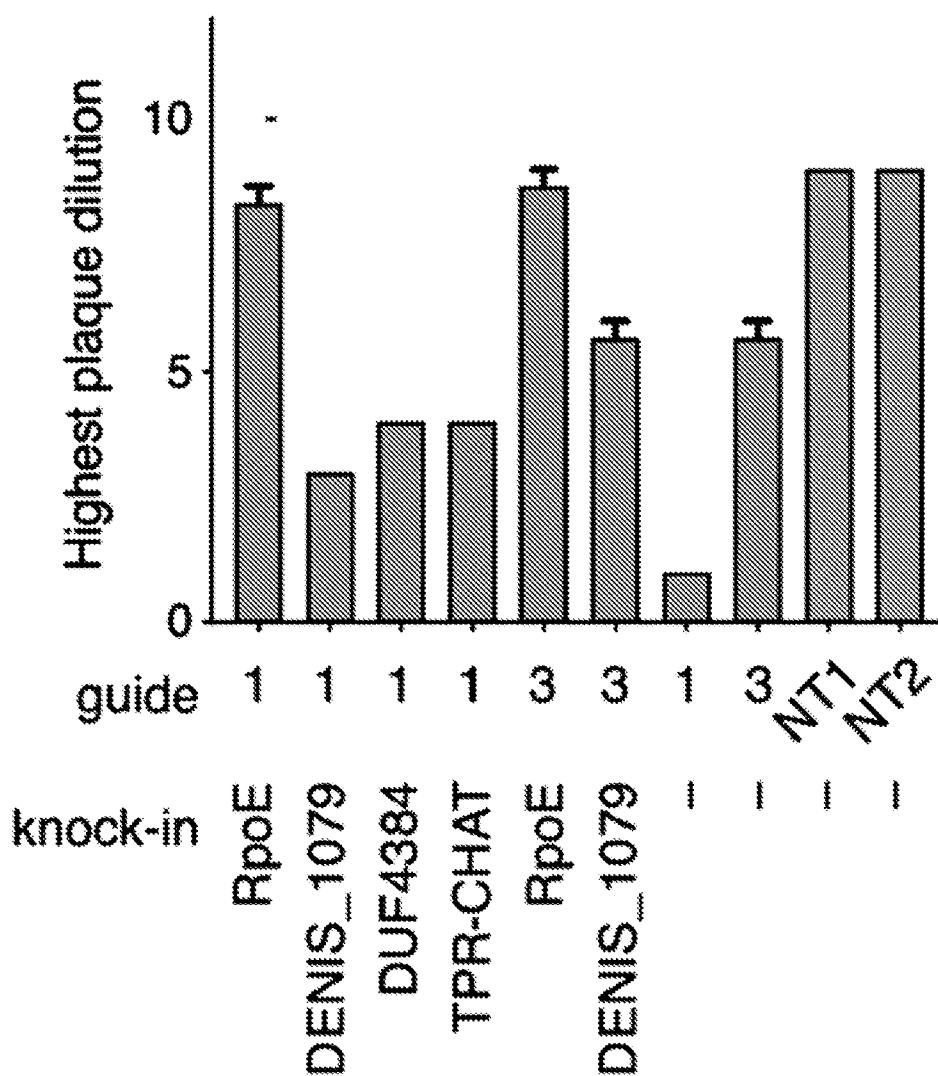


FIG. 11B

72/142

**FIG. 11C**

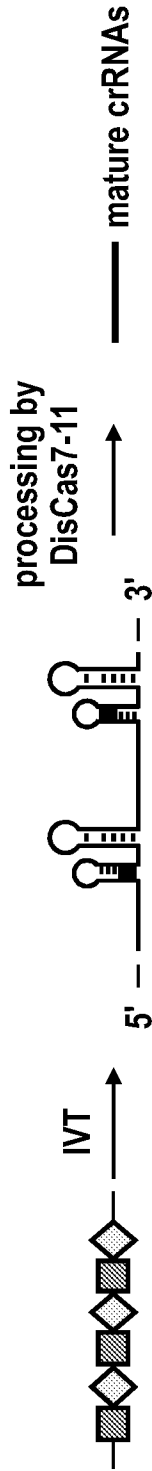


FIG. 12A

73/142

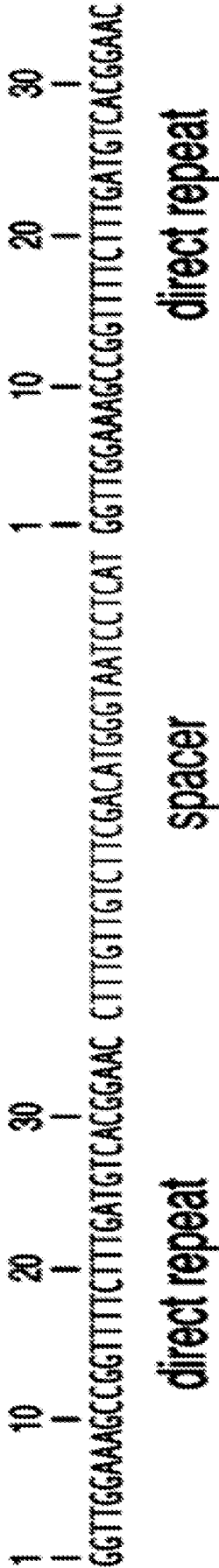


FIG. 12B

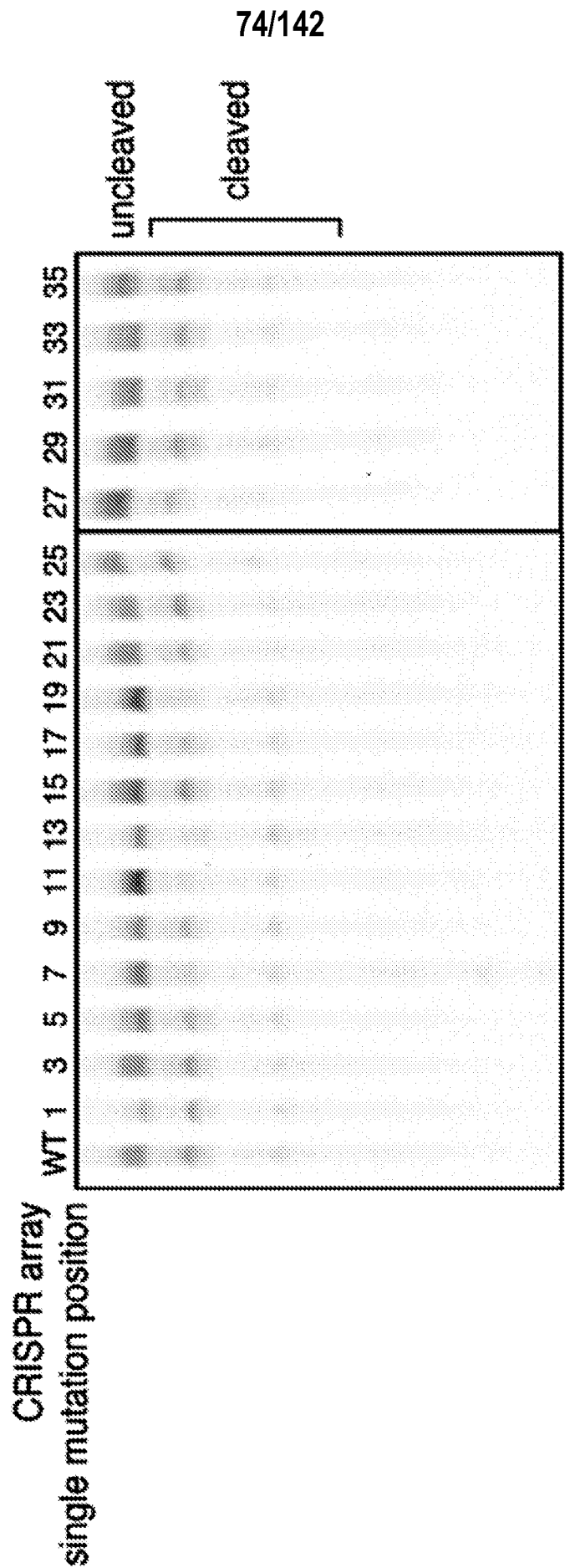


FIG. 12C

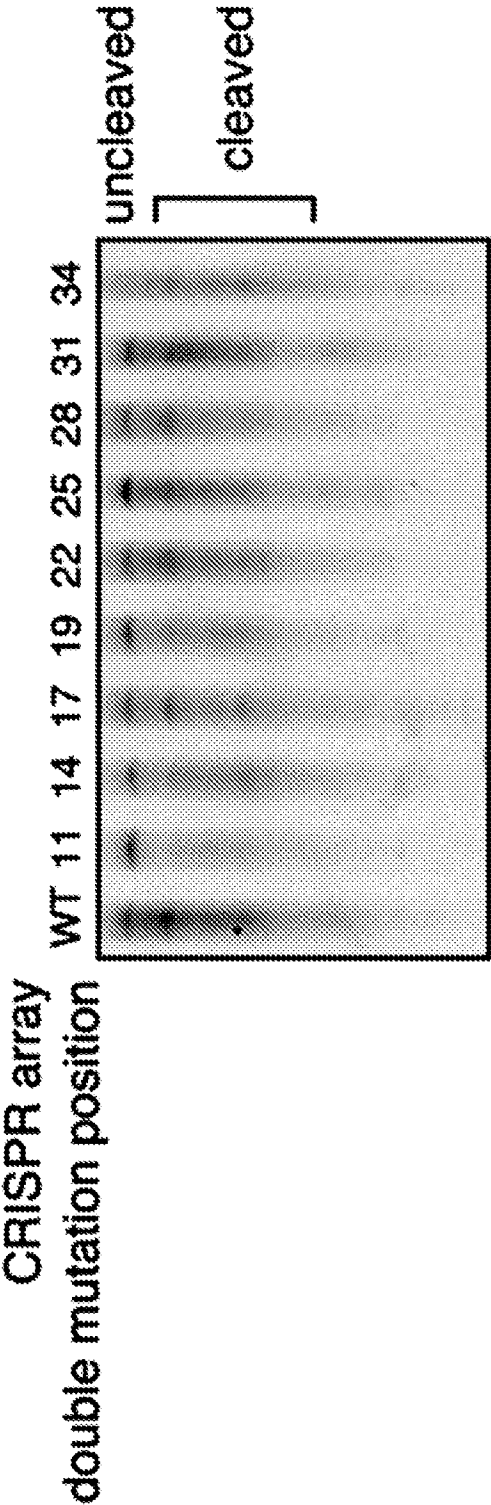


FIG. 12D

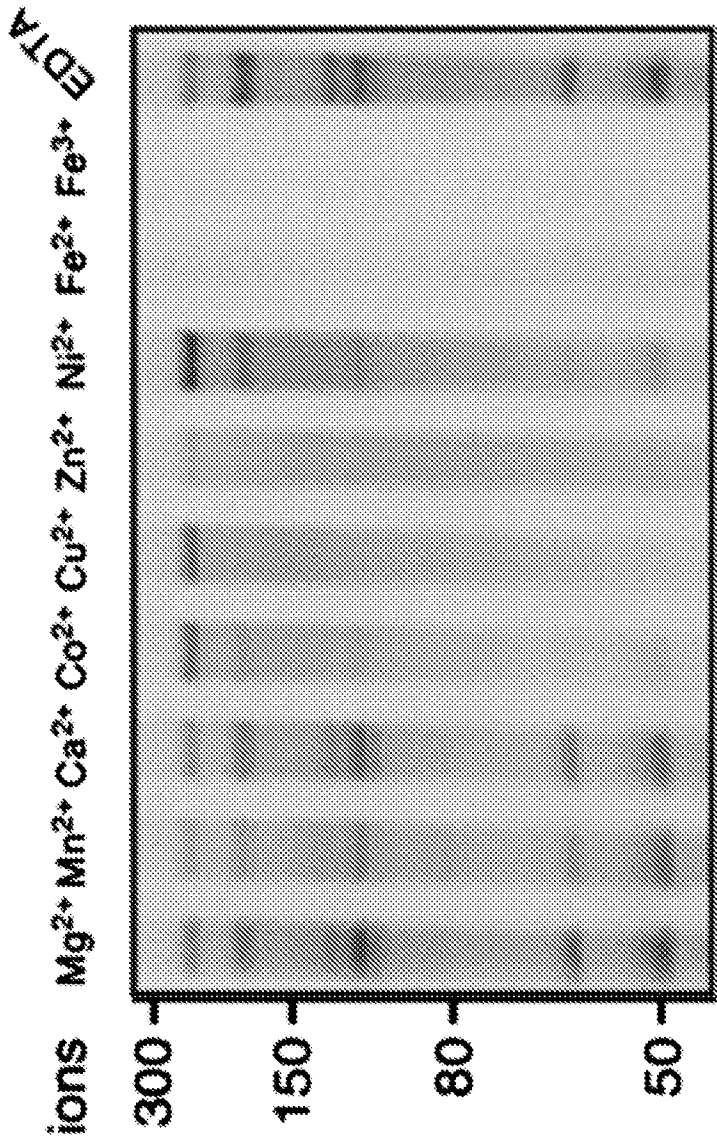
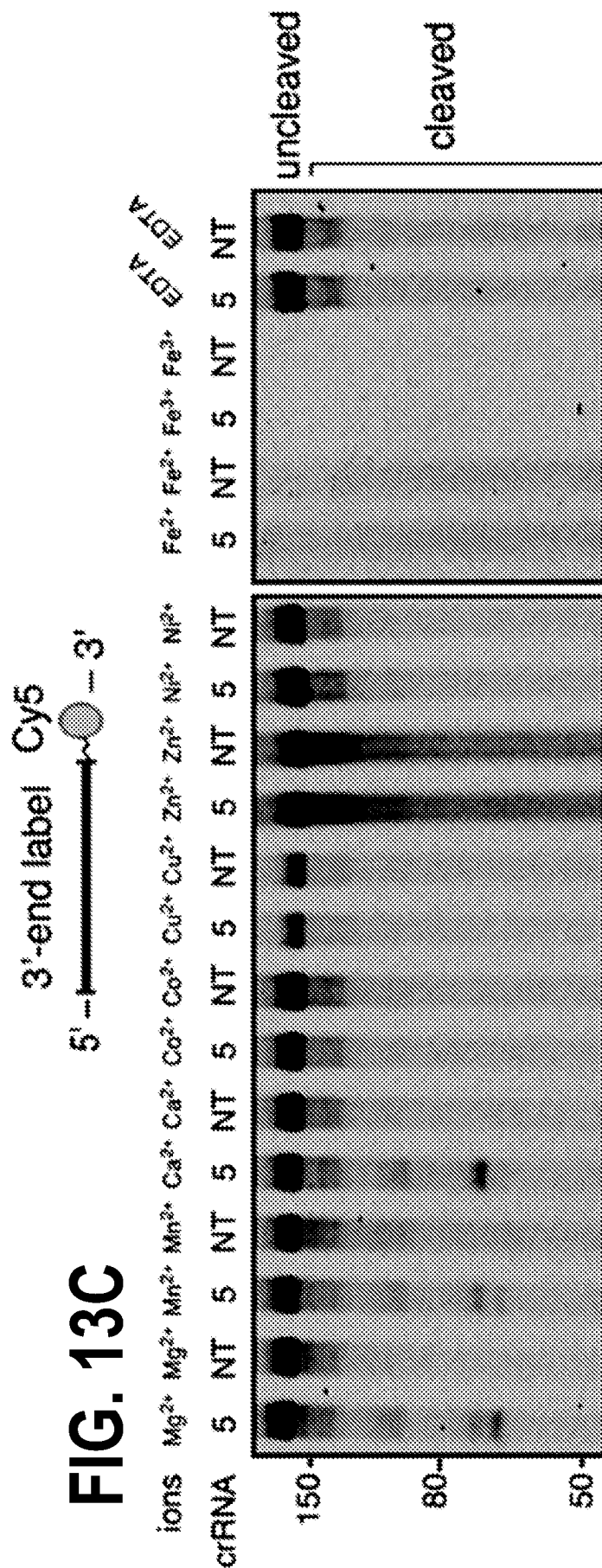


FIG. 13A



78/142

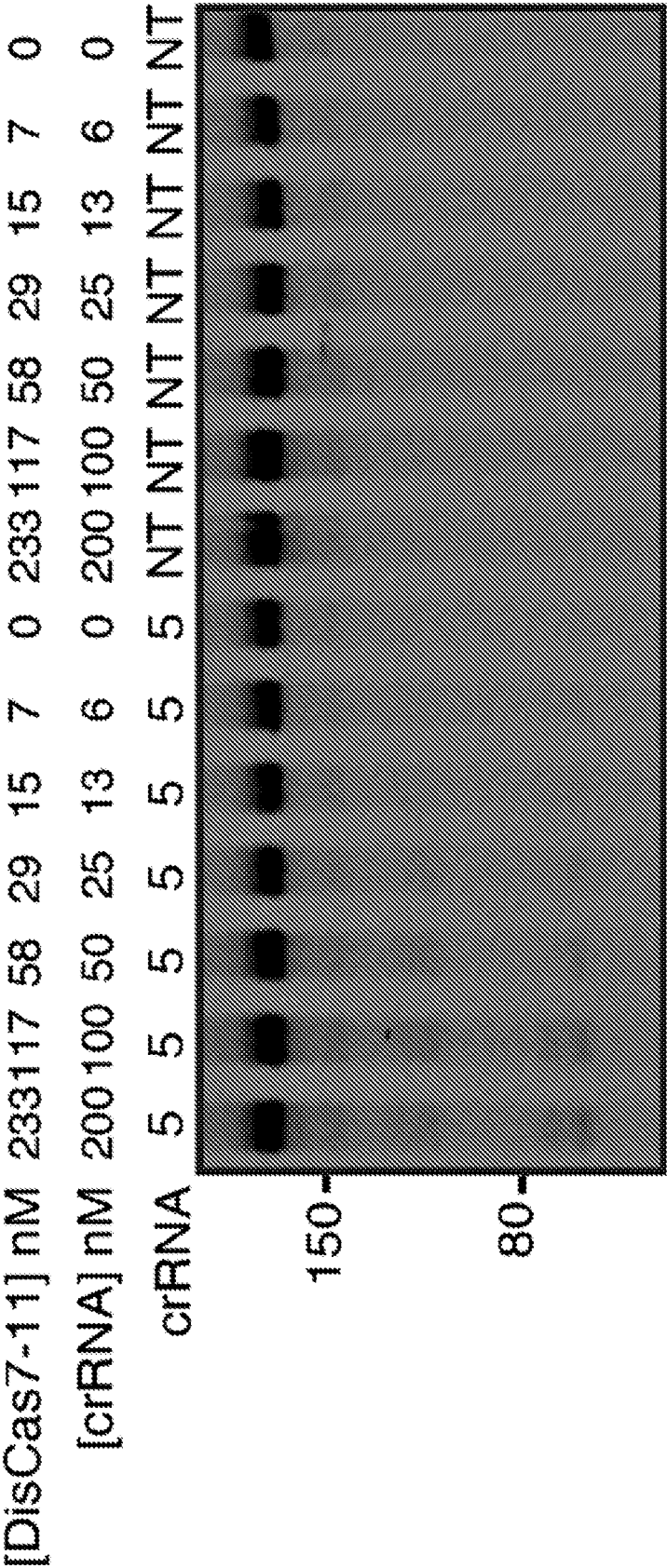


FIG. 13D

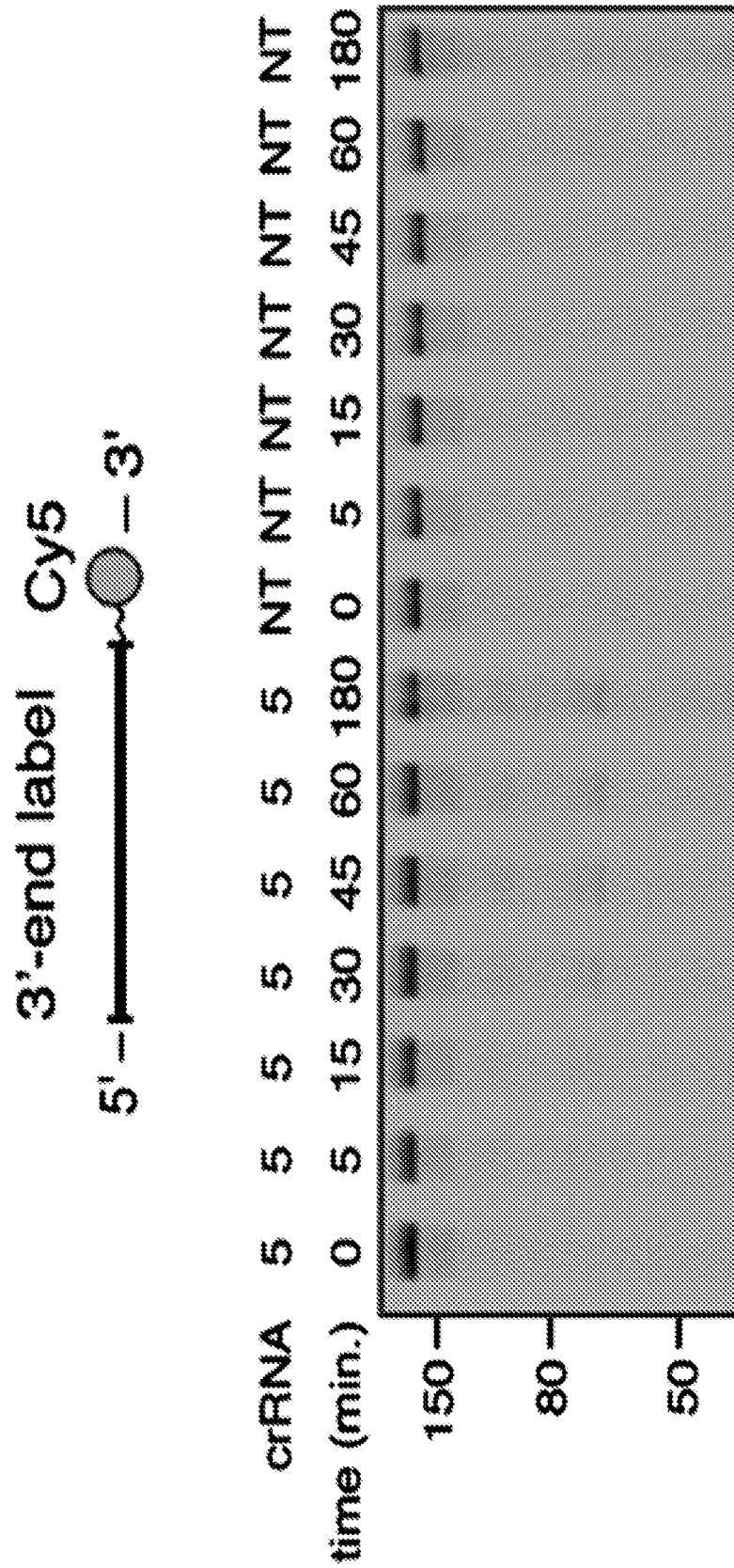


FIG. 13E

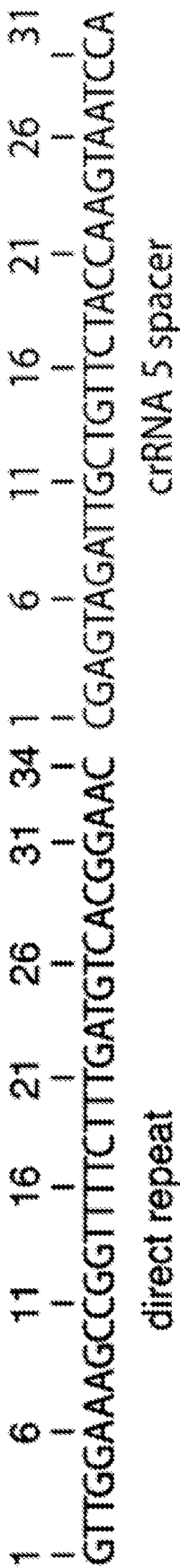


FIG. 14A

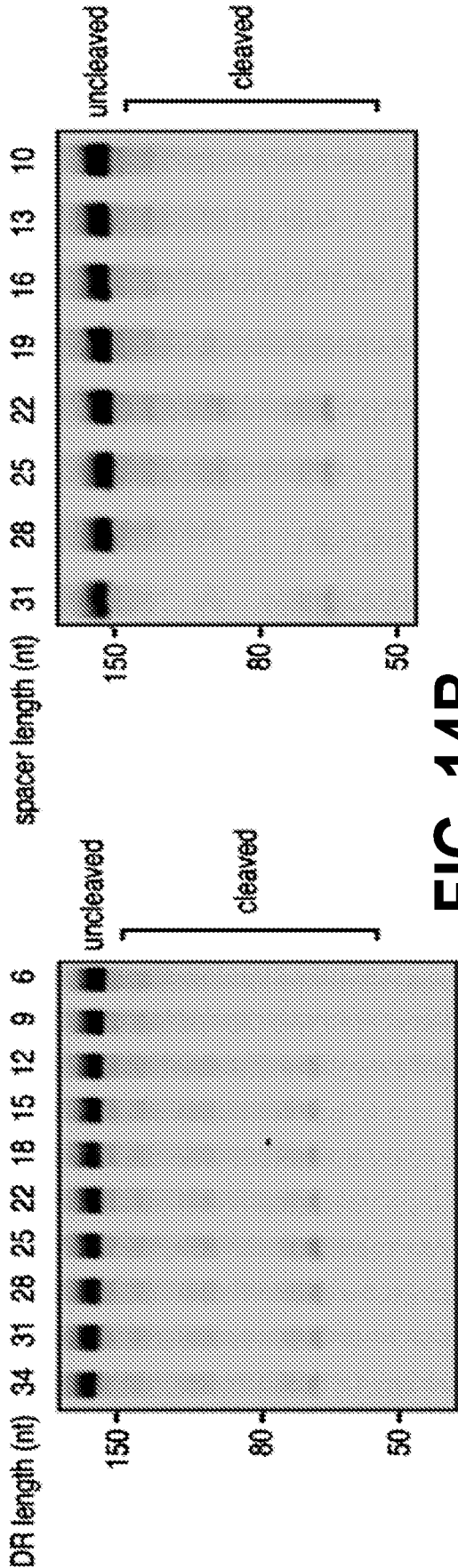
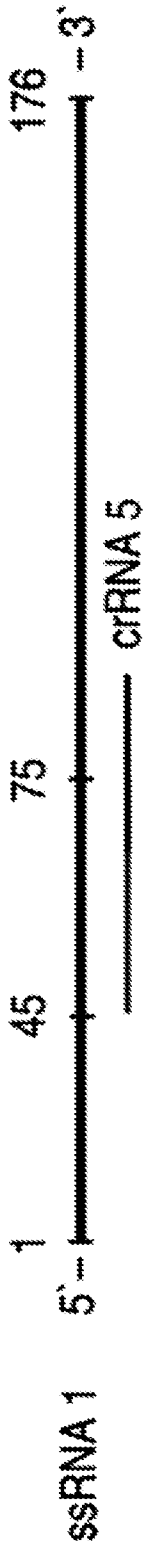


FIG. 14B

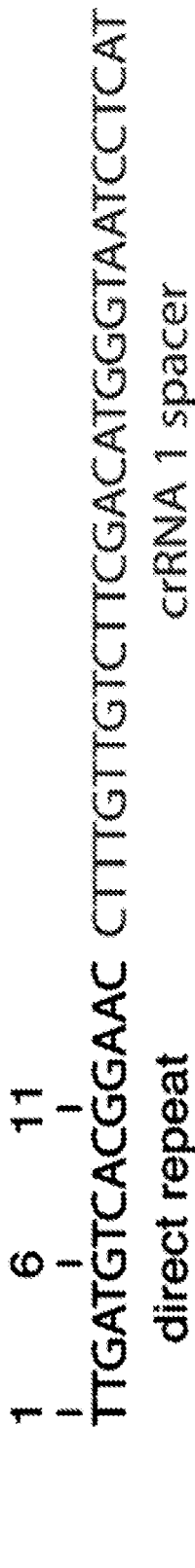


FIG. 14C

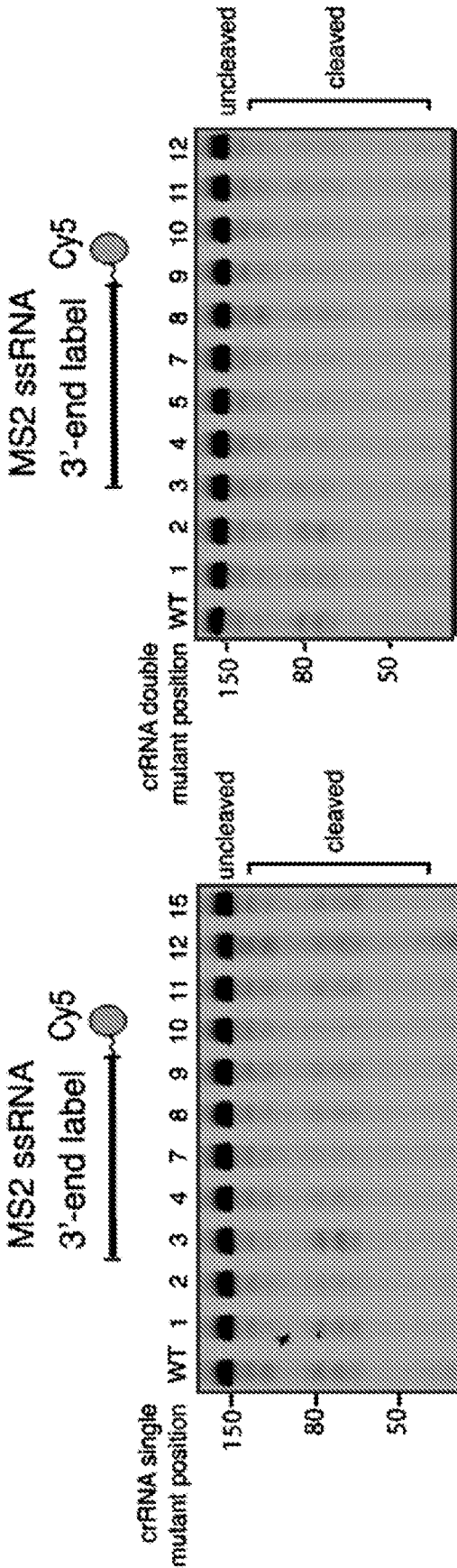
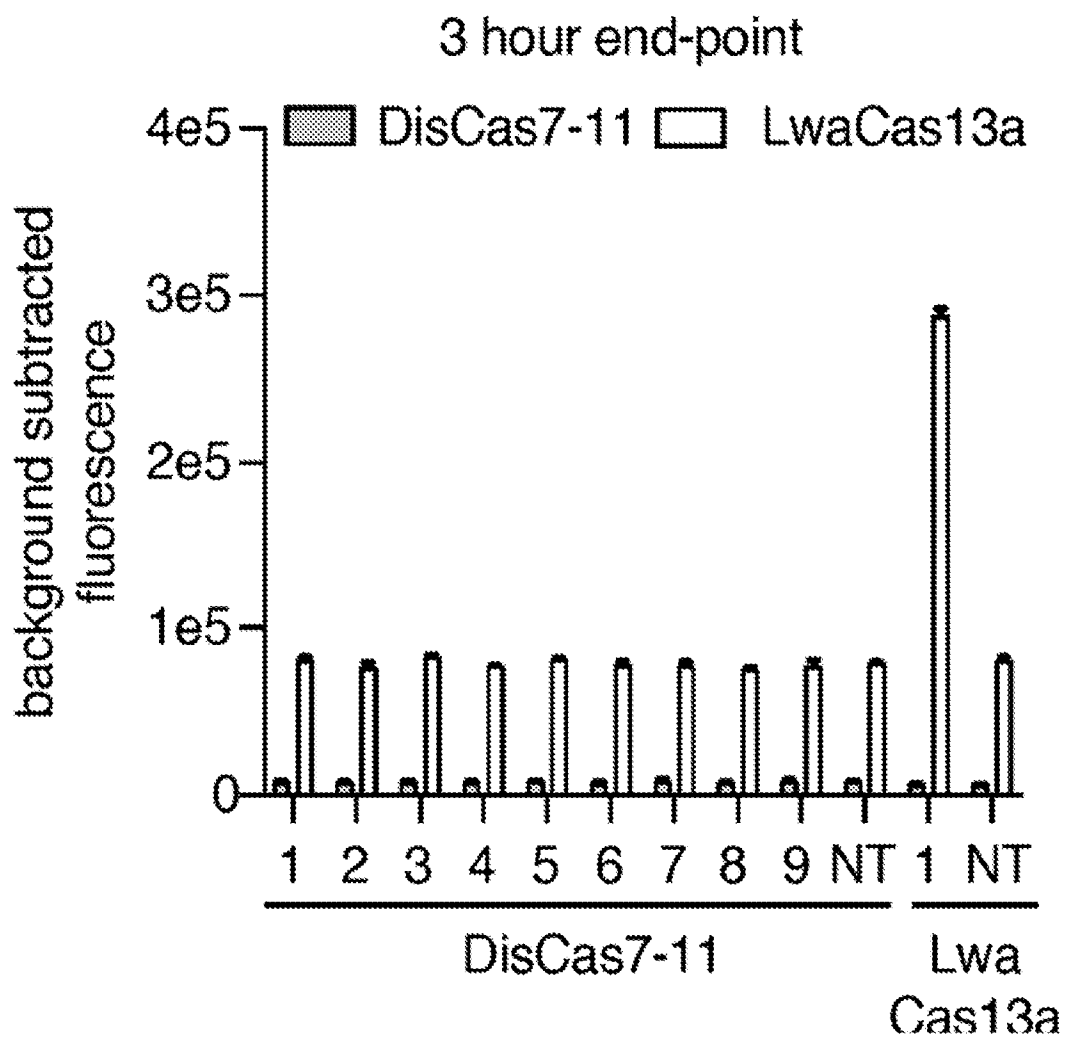
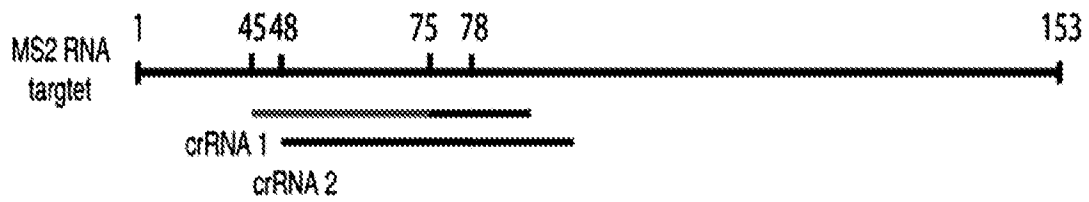
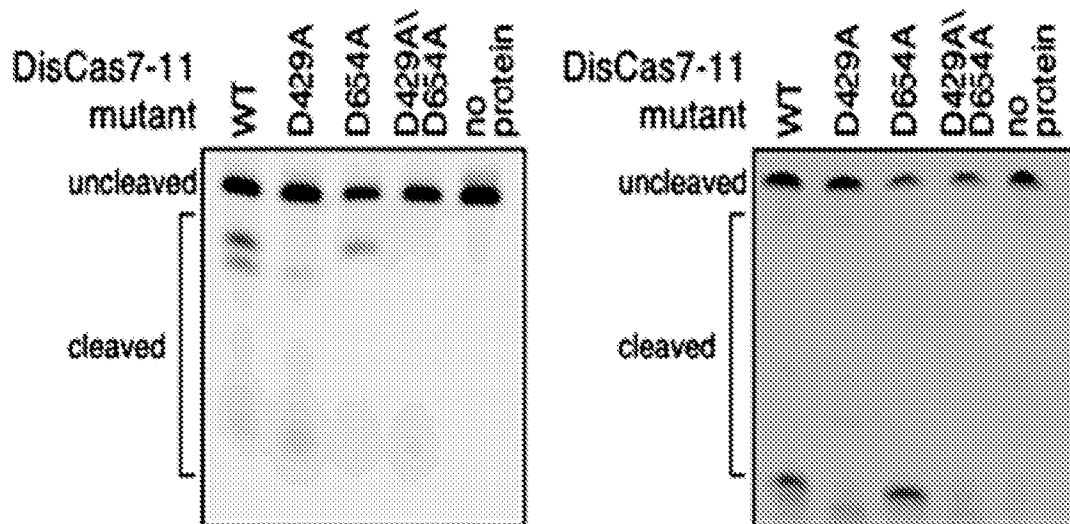
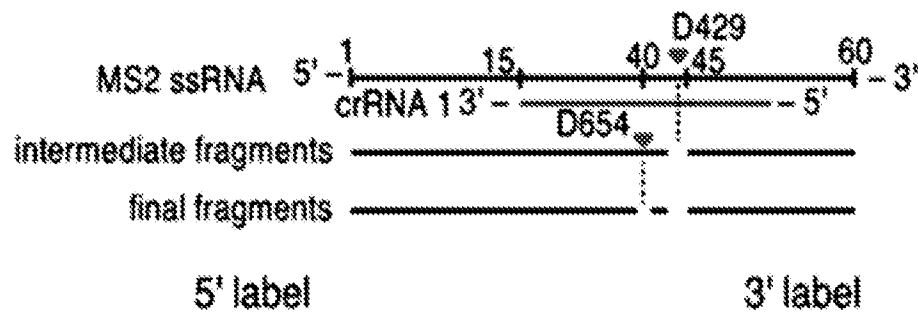


FIG. 14D

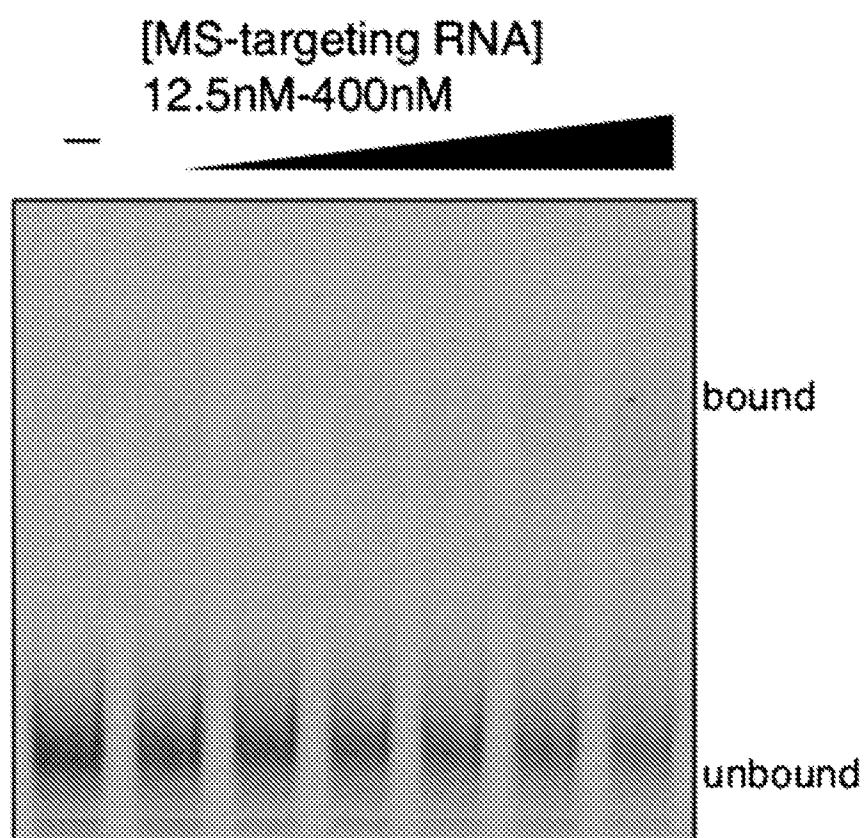
82/142

**FIG. 15**

83/142

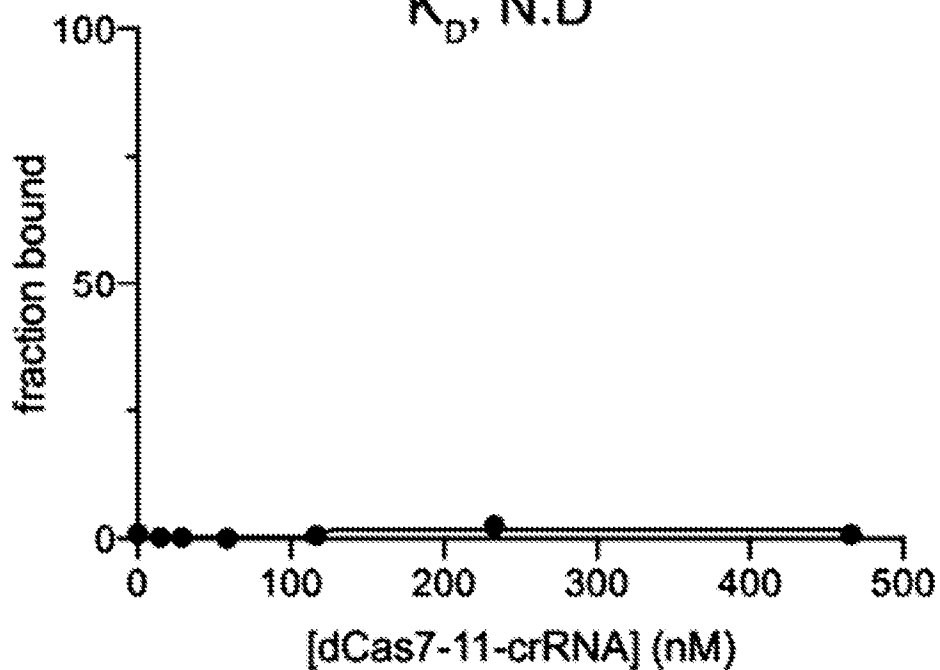
**FIG. 16A****FIG. 16B**

84/142

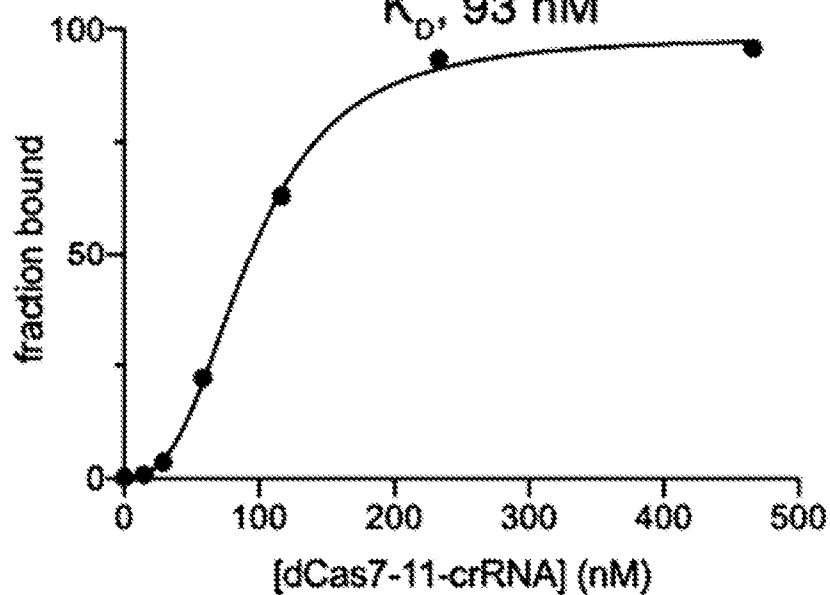
 K_D , N.D**FIG. 17A**

85/142

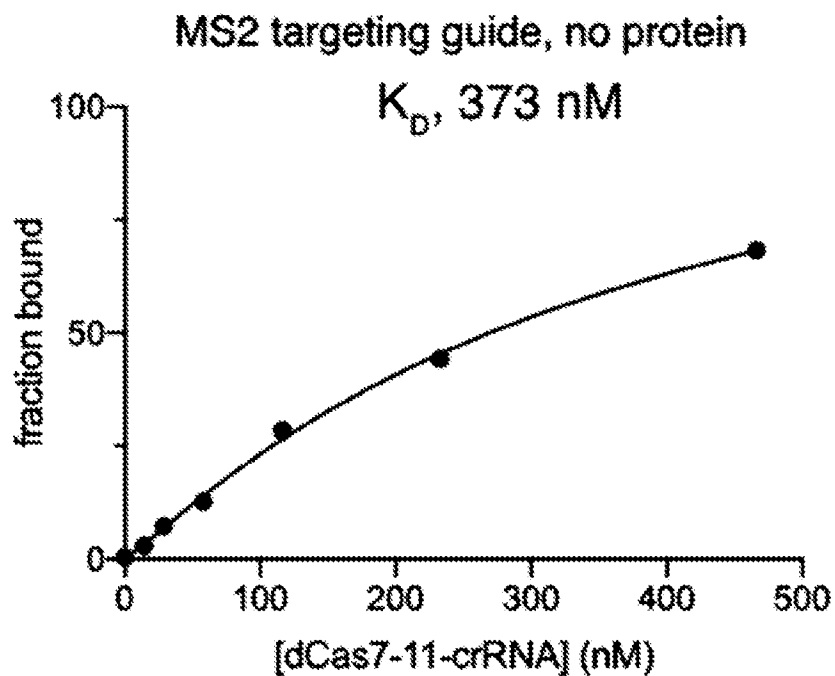
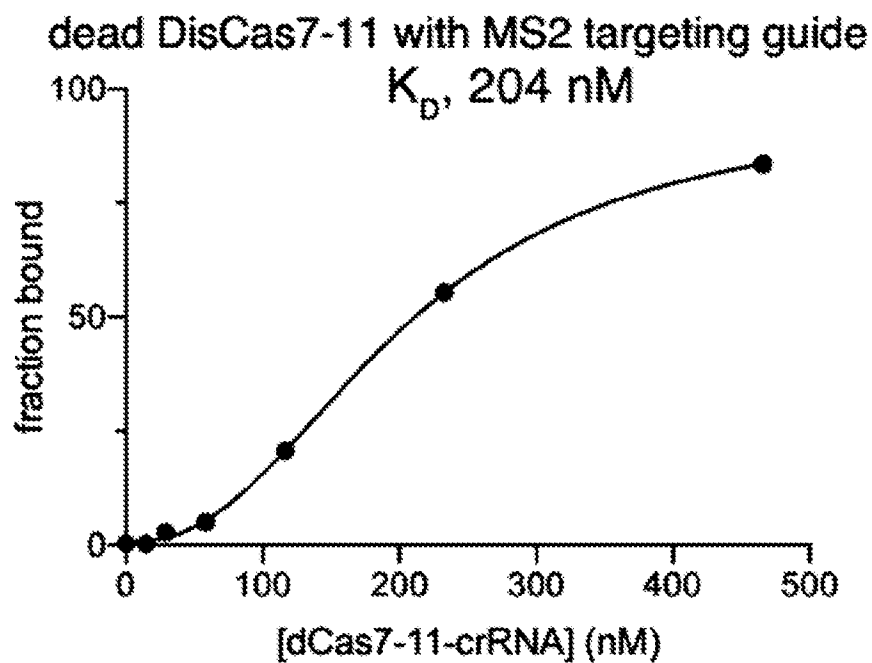
non-targeting guide

 K_D , N.D**FIG. 17B**

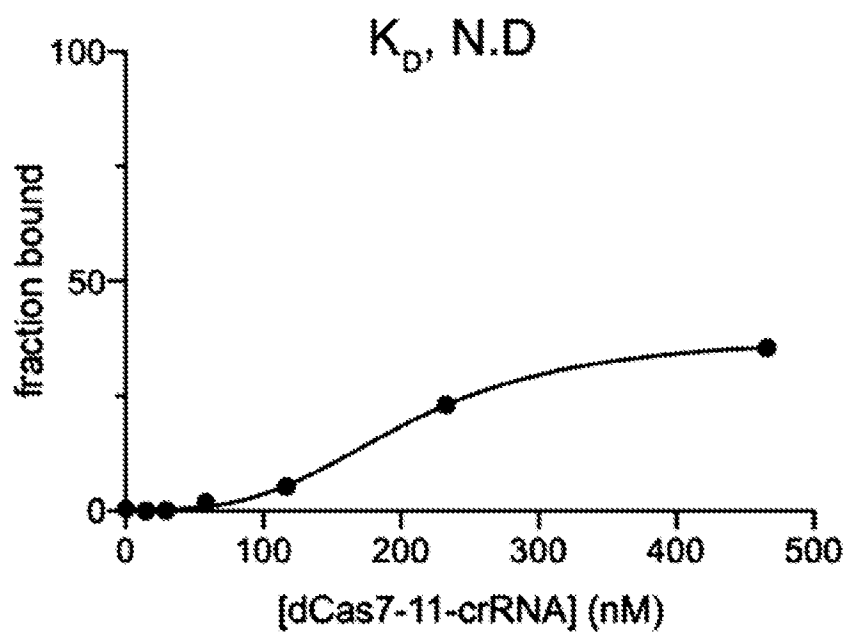
active DisCas7-11 with MS2 targeting guide

 K_D , 93 nM**FIG. 17C**

86/142

**FIG. 17D****FIG. 17E**

87/142

**FIG. 17F**

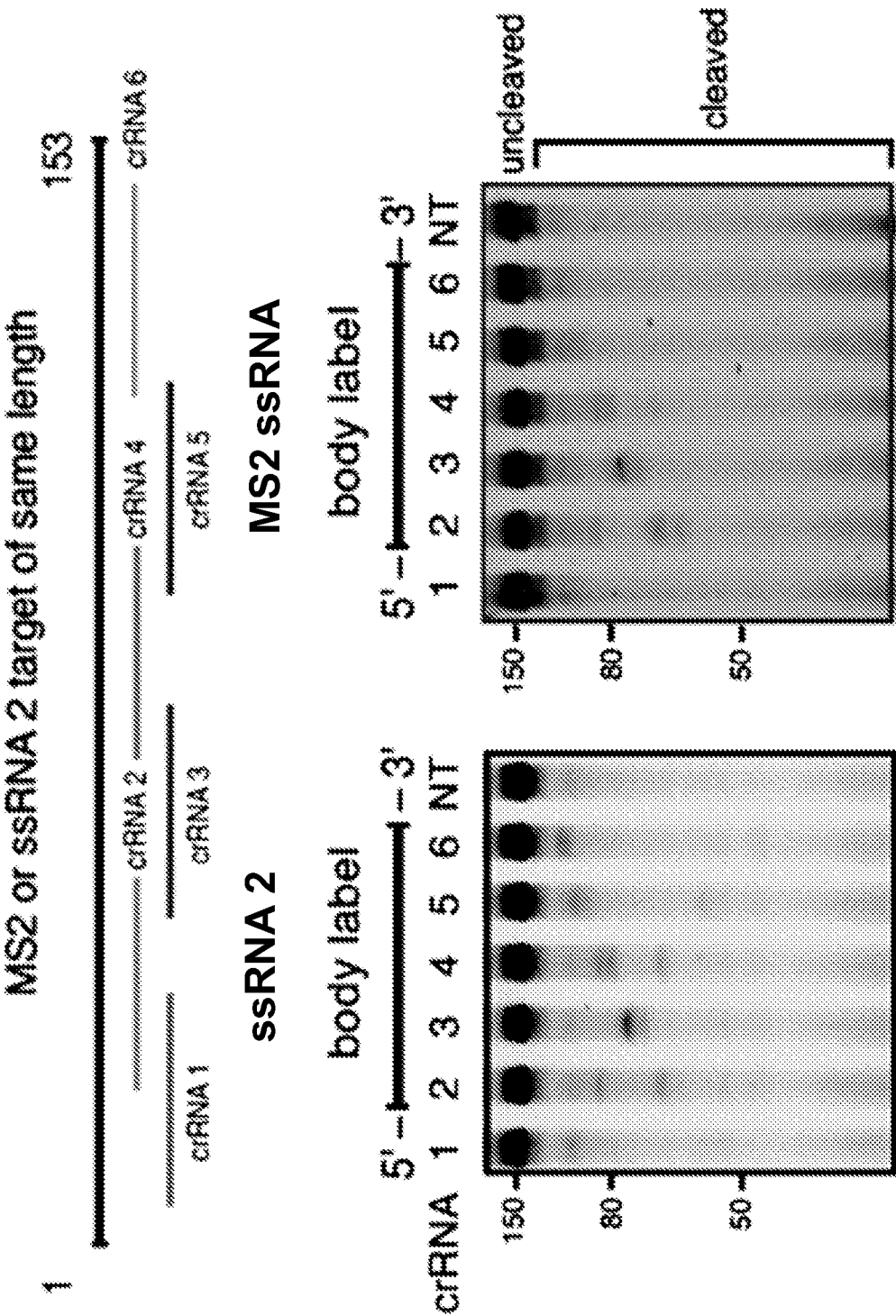
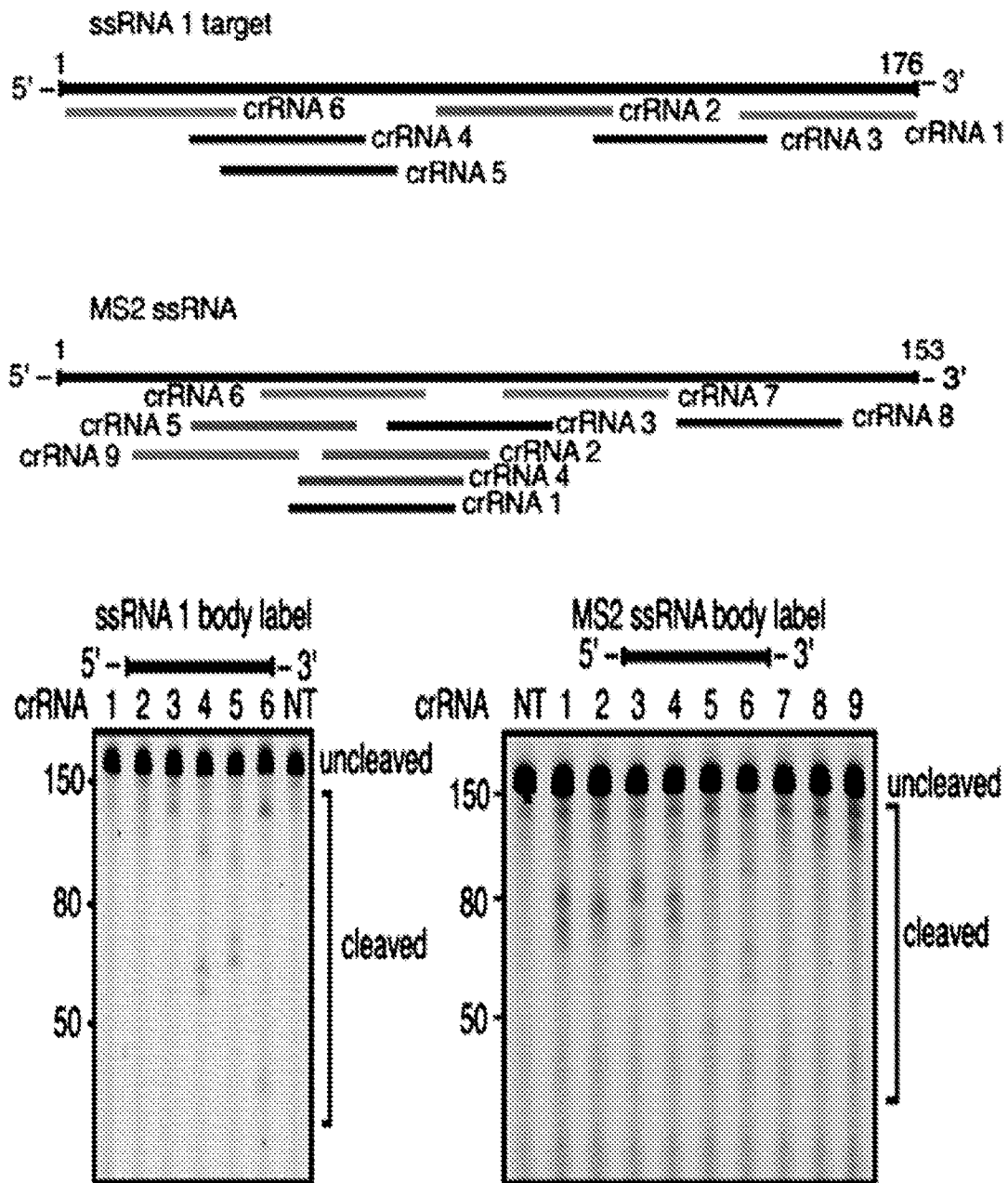


FIG. 18A

89/142

**FIG. 18B**

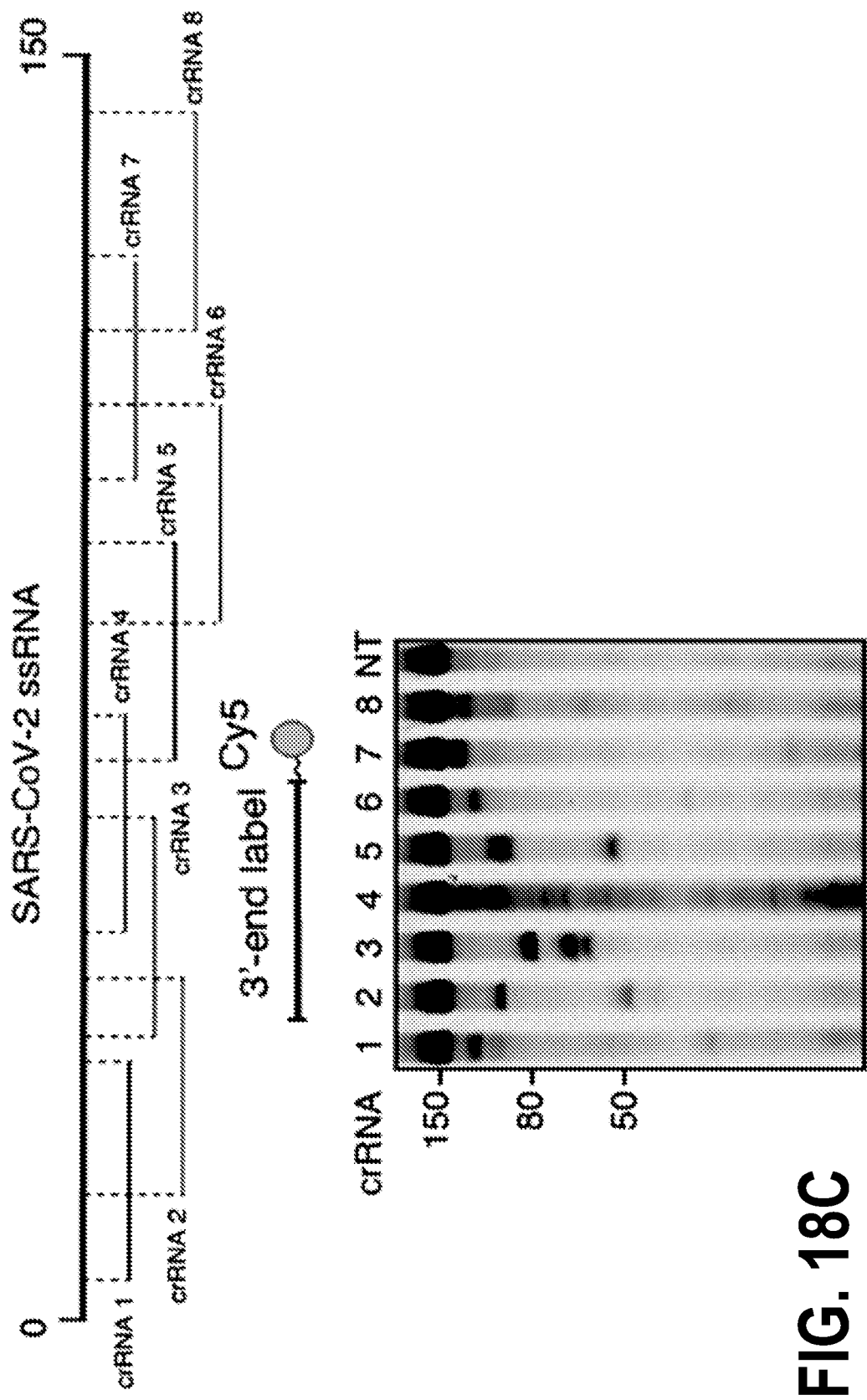


FIG. 18C

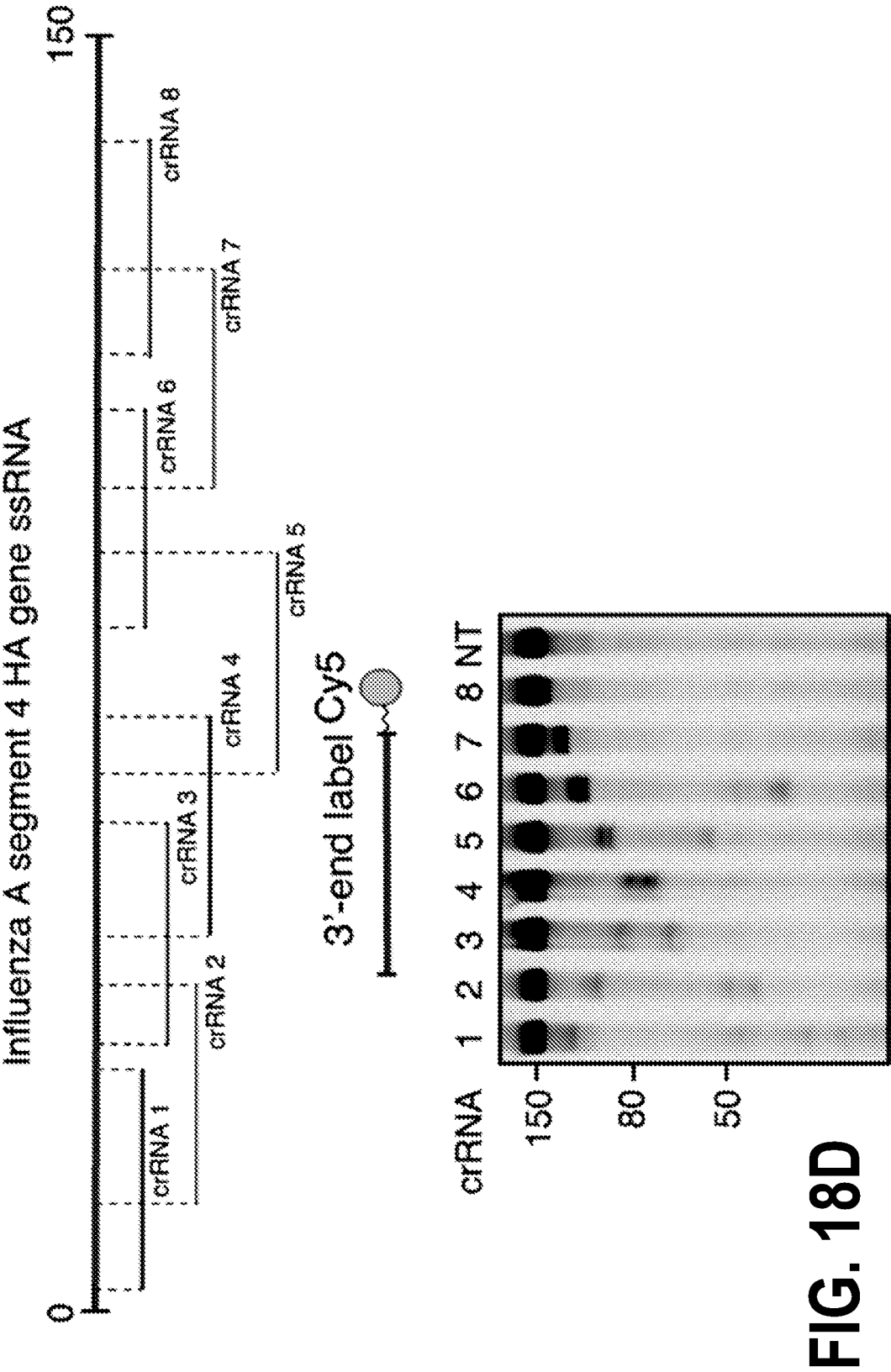
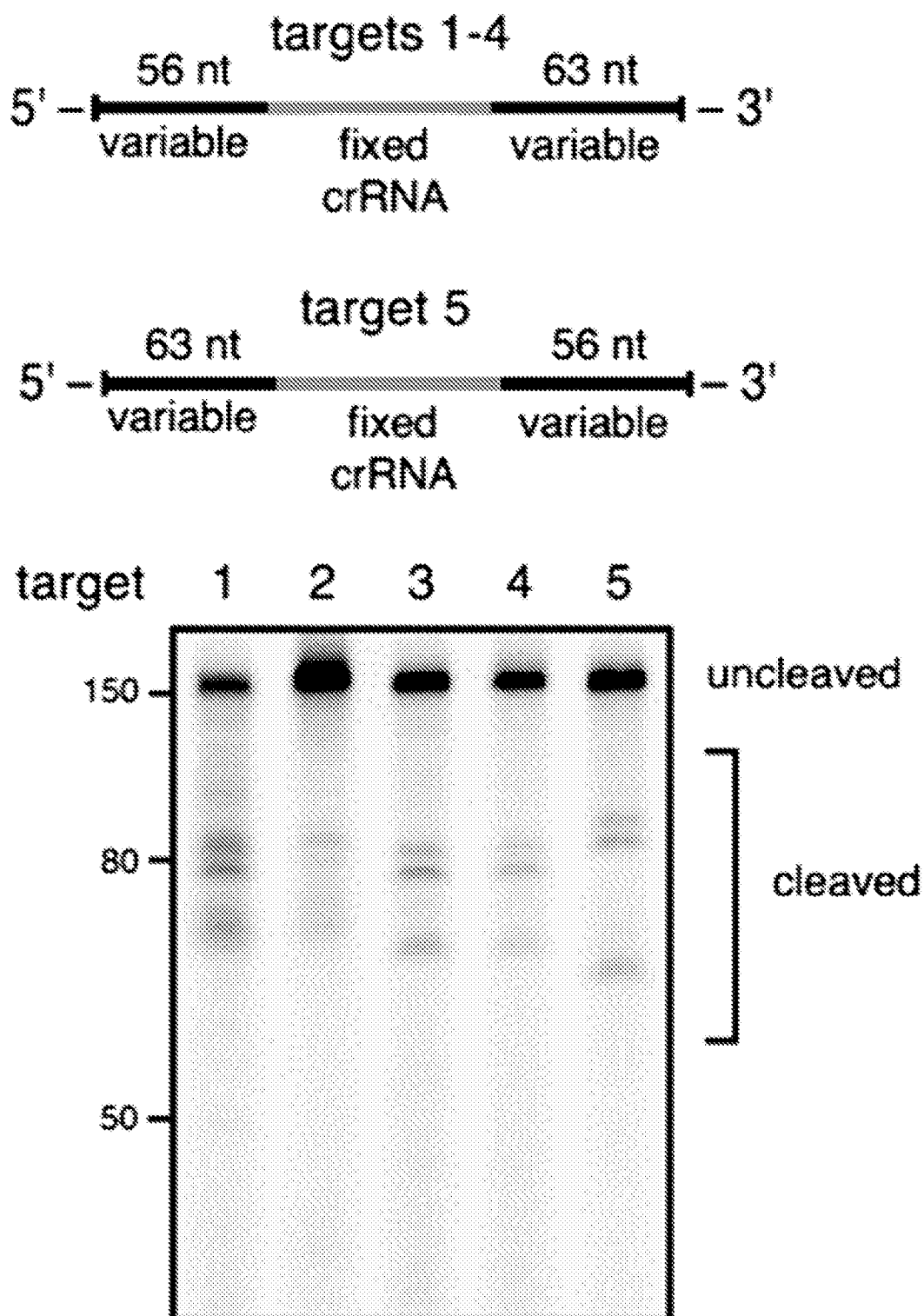


FIG. 18D

92/142

**FIG. 18E**

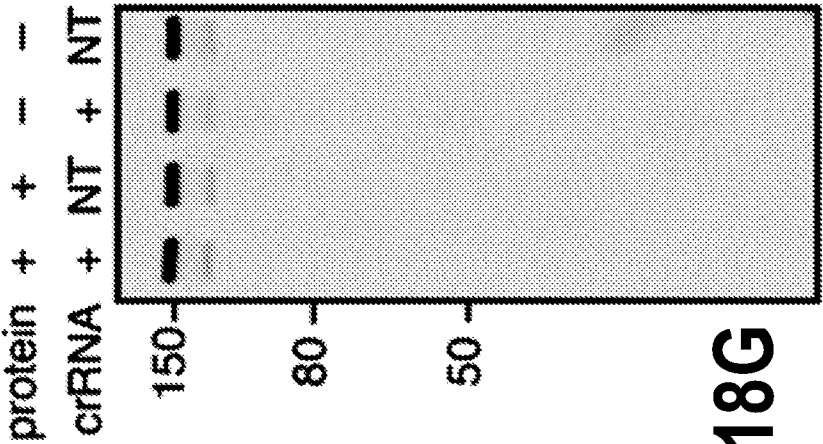
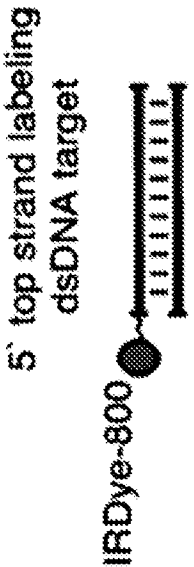


FIG. 18G

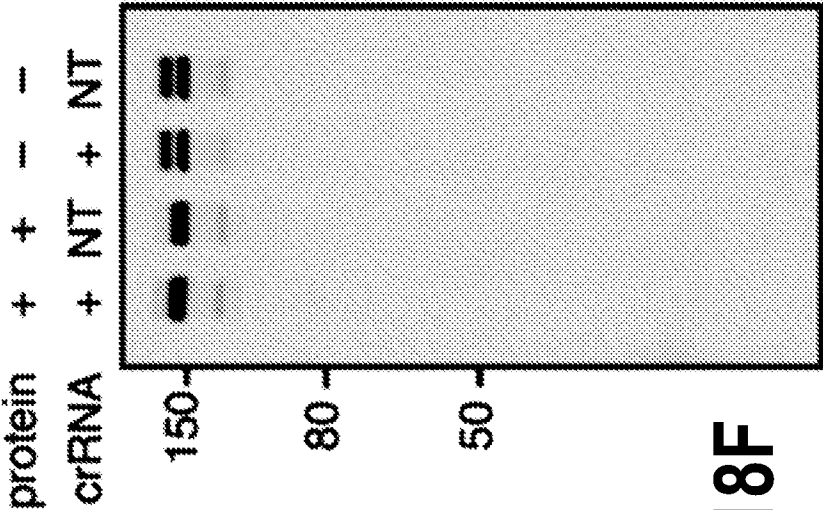
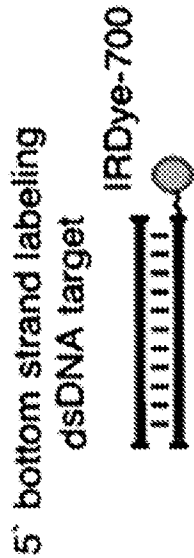
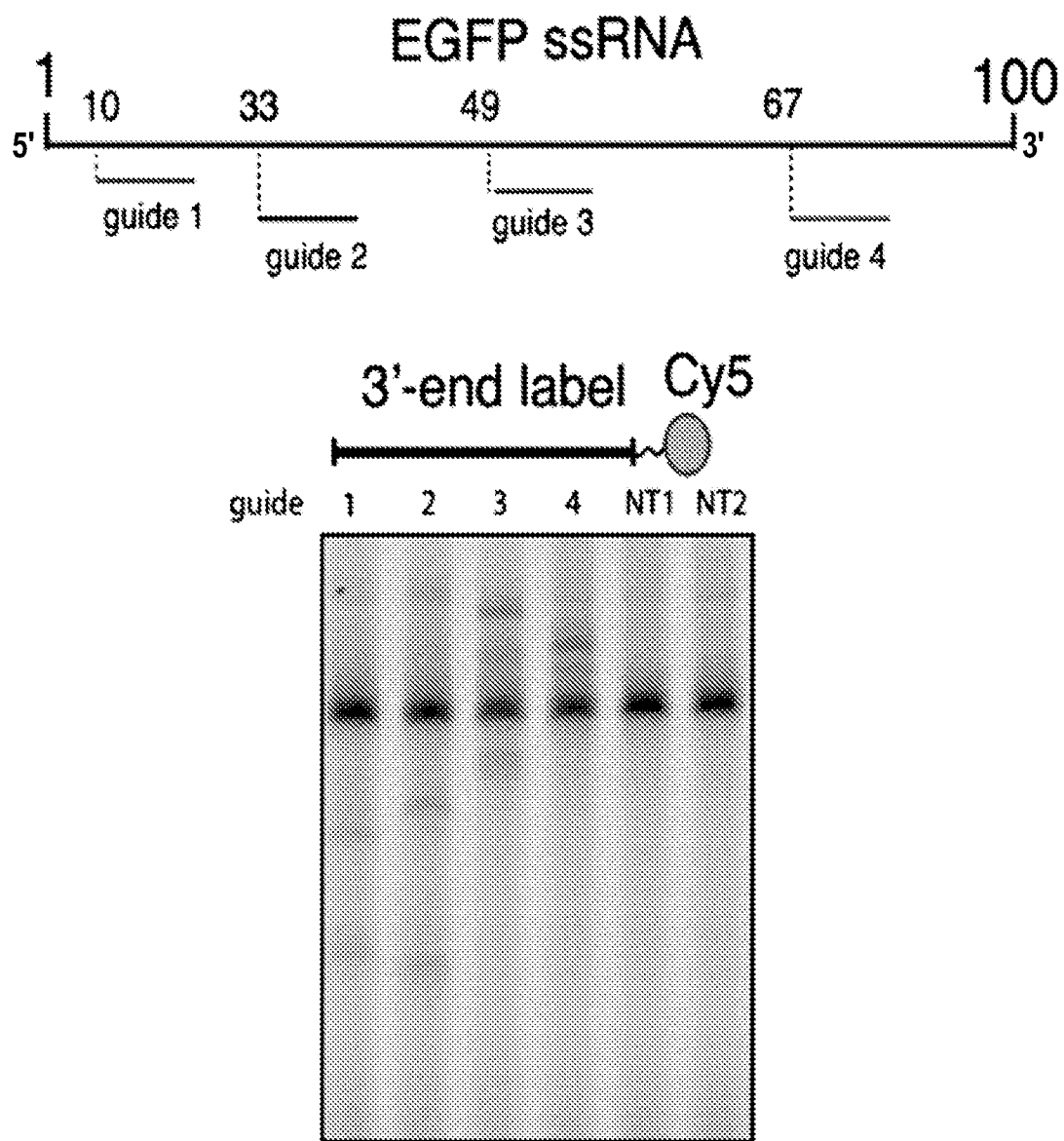


FIG. 18F

94/142

**FIG. 19A**

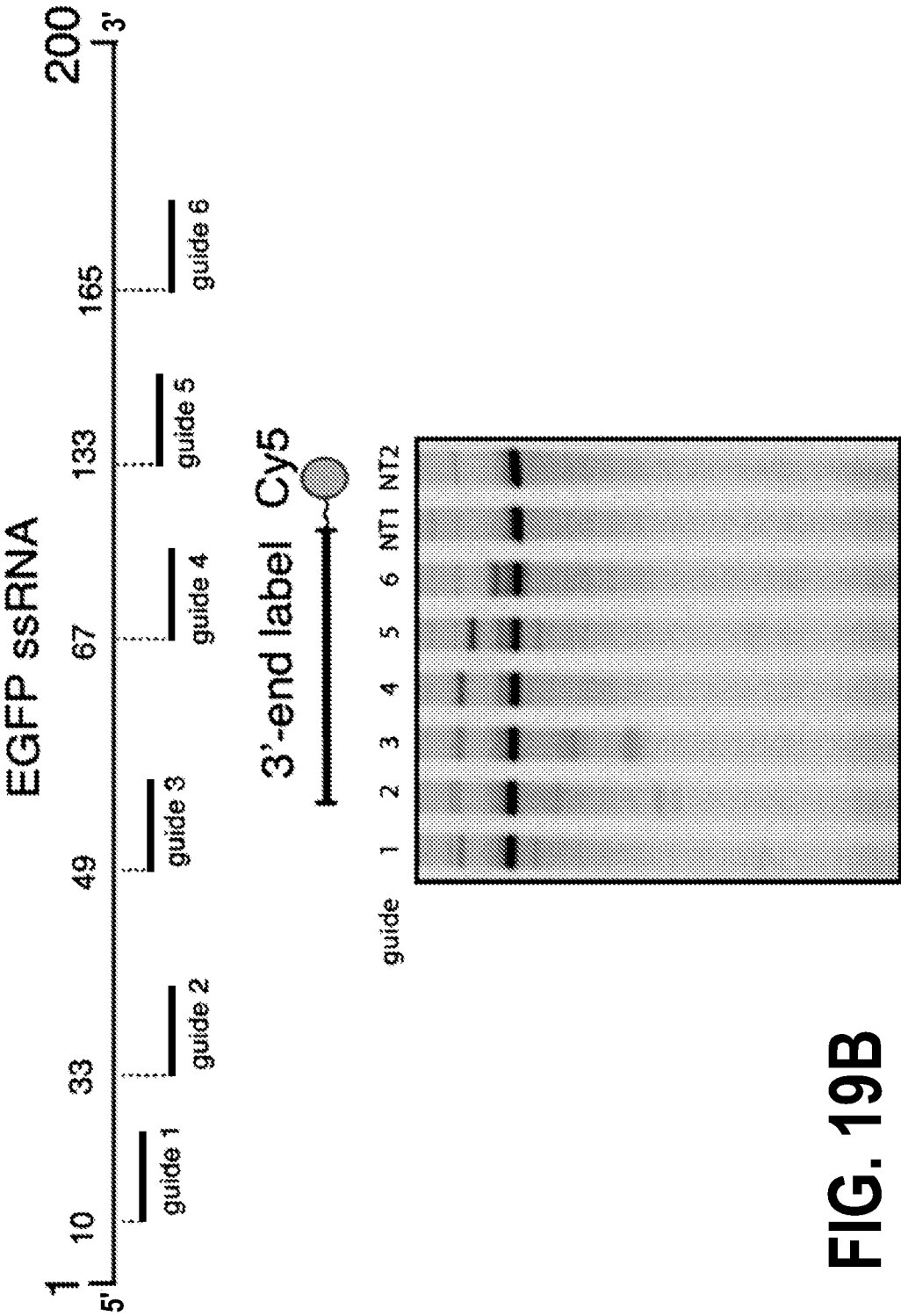


FIG. 19B

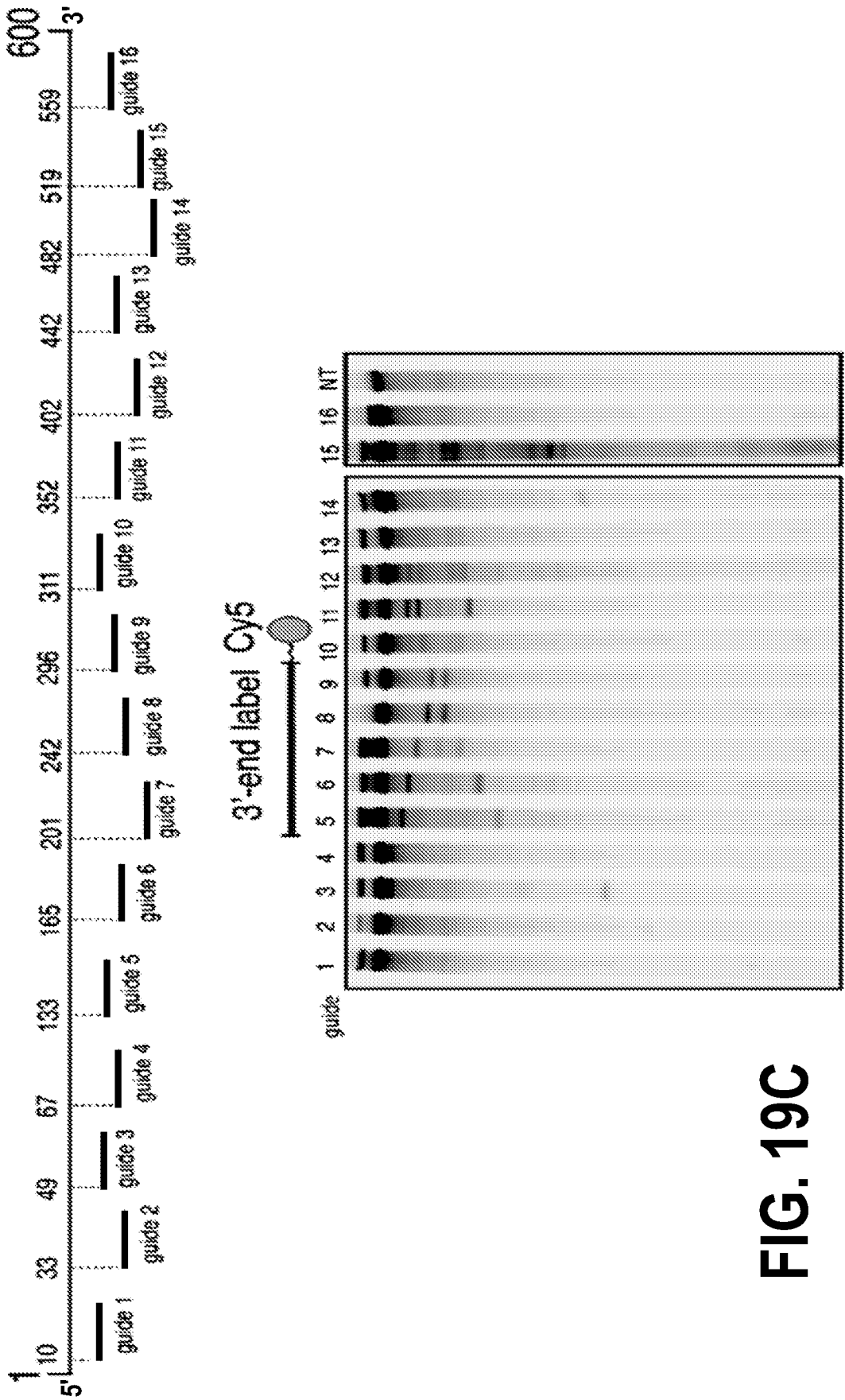
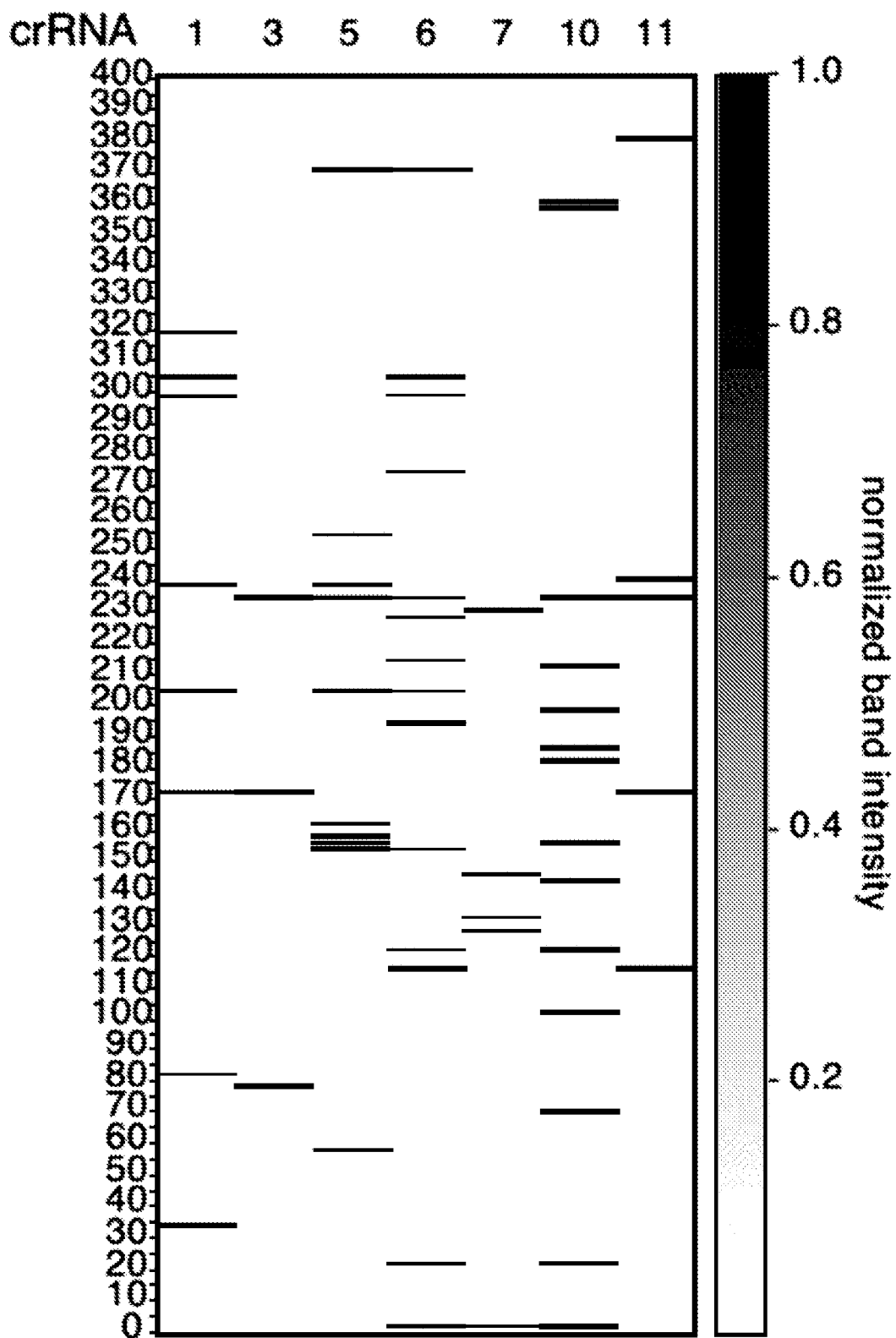


FIG. 19C

97/142

**FIG. 19D**

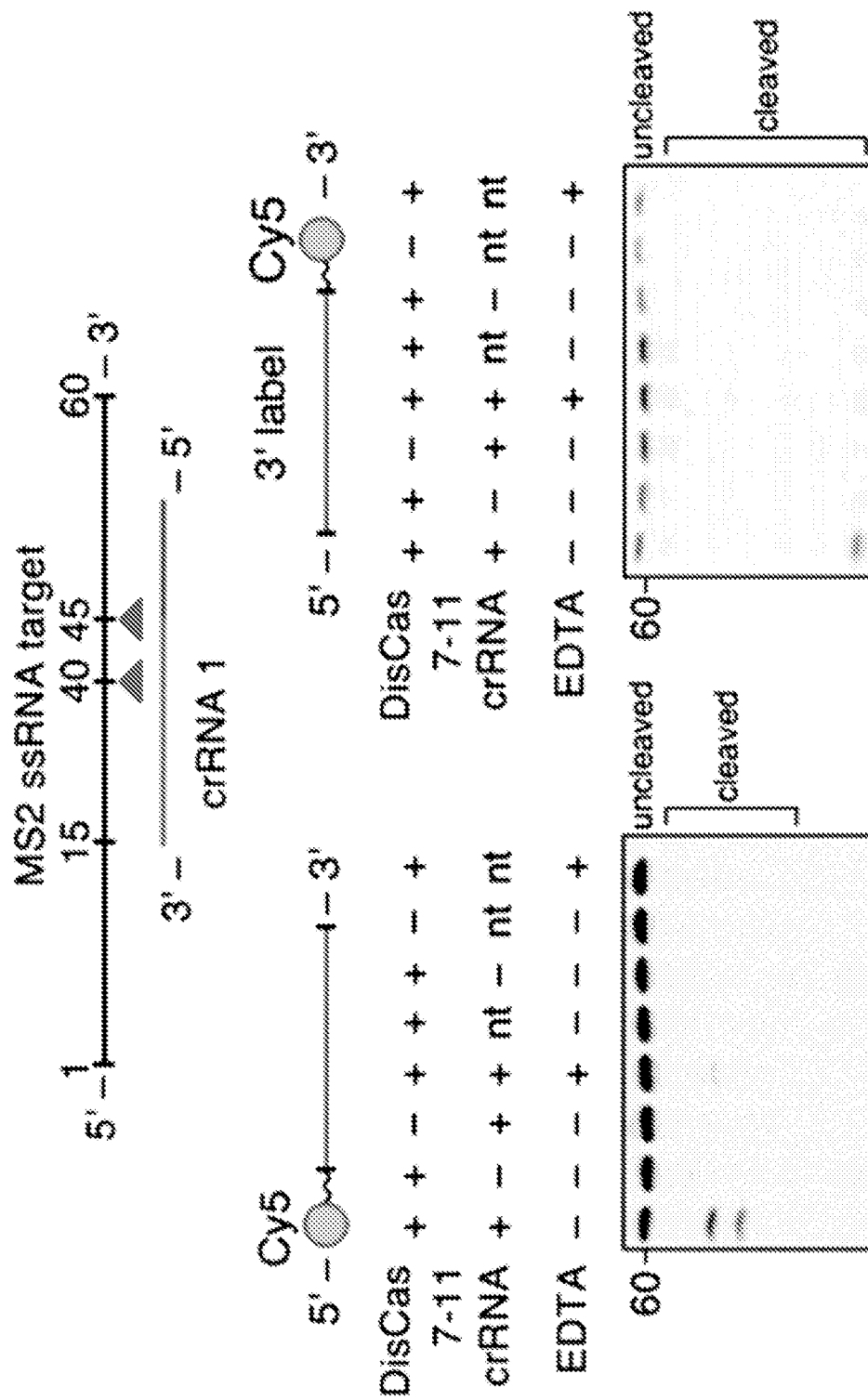
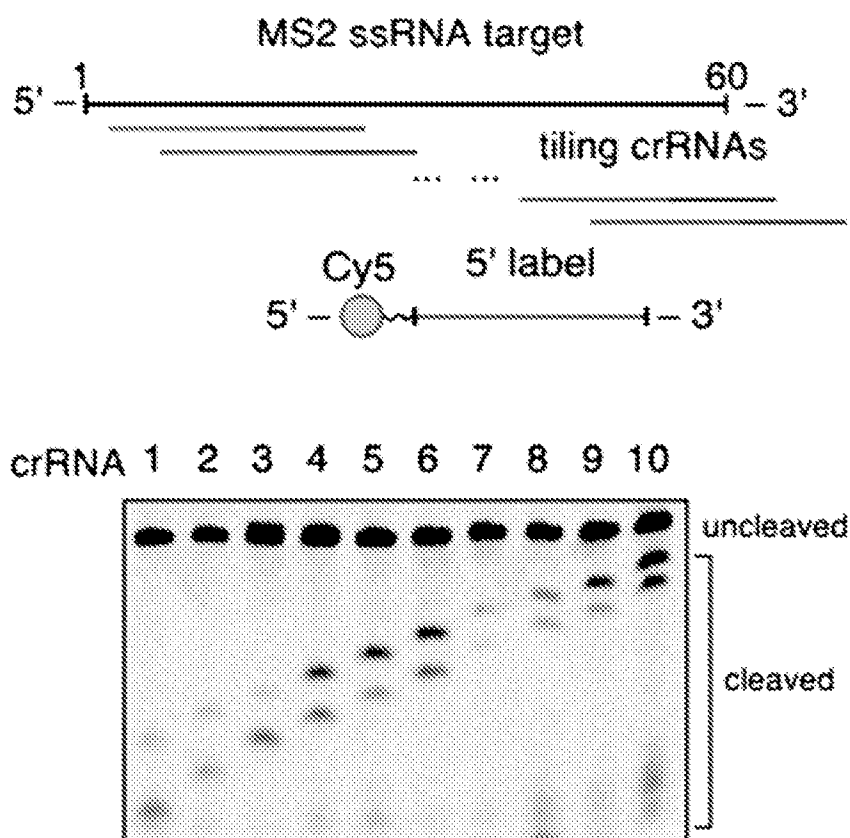


FIG. 19E

99/142

**FIG. 19F**

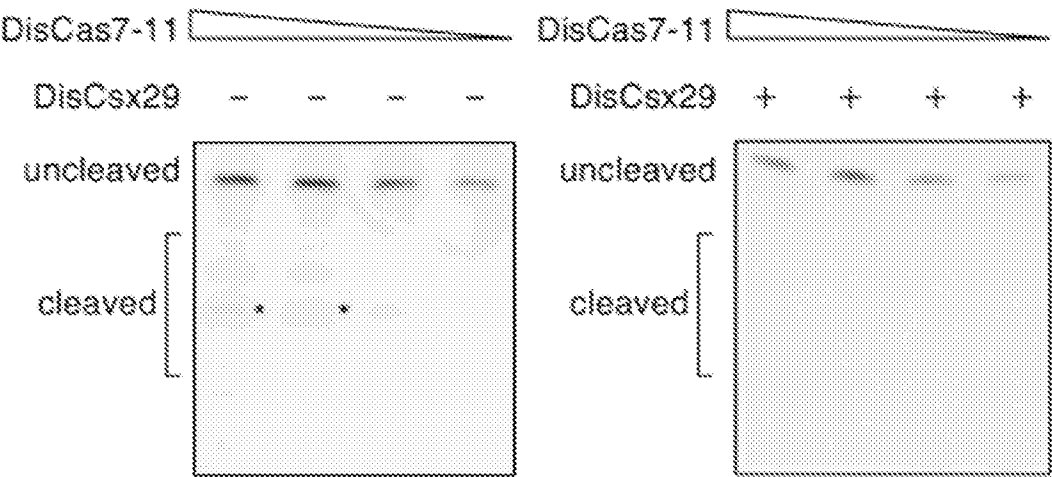


FIG. 19G

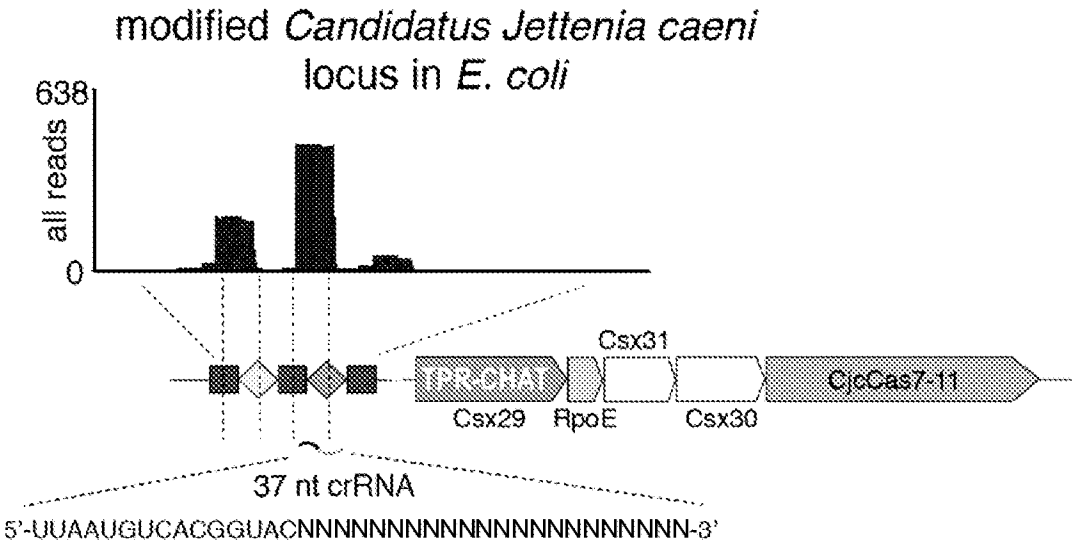


FIG. 20

101/142

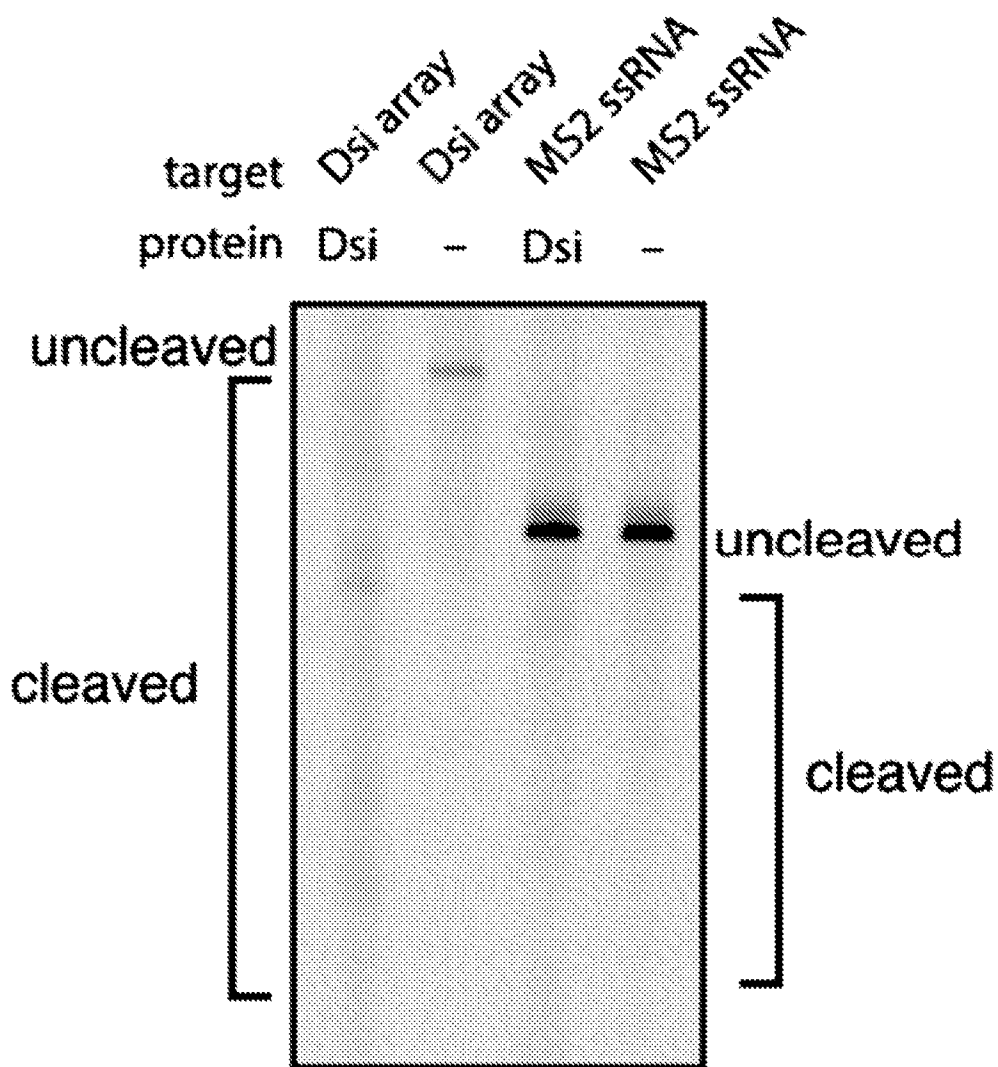


FIG. 21A

102/142

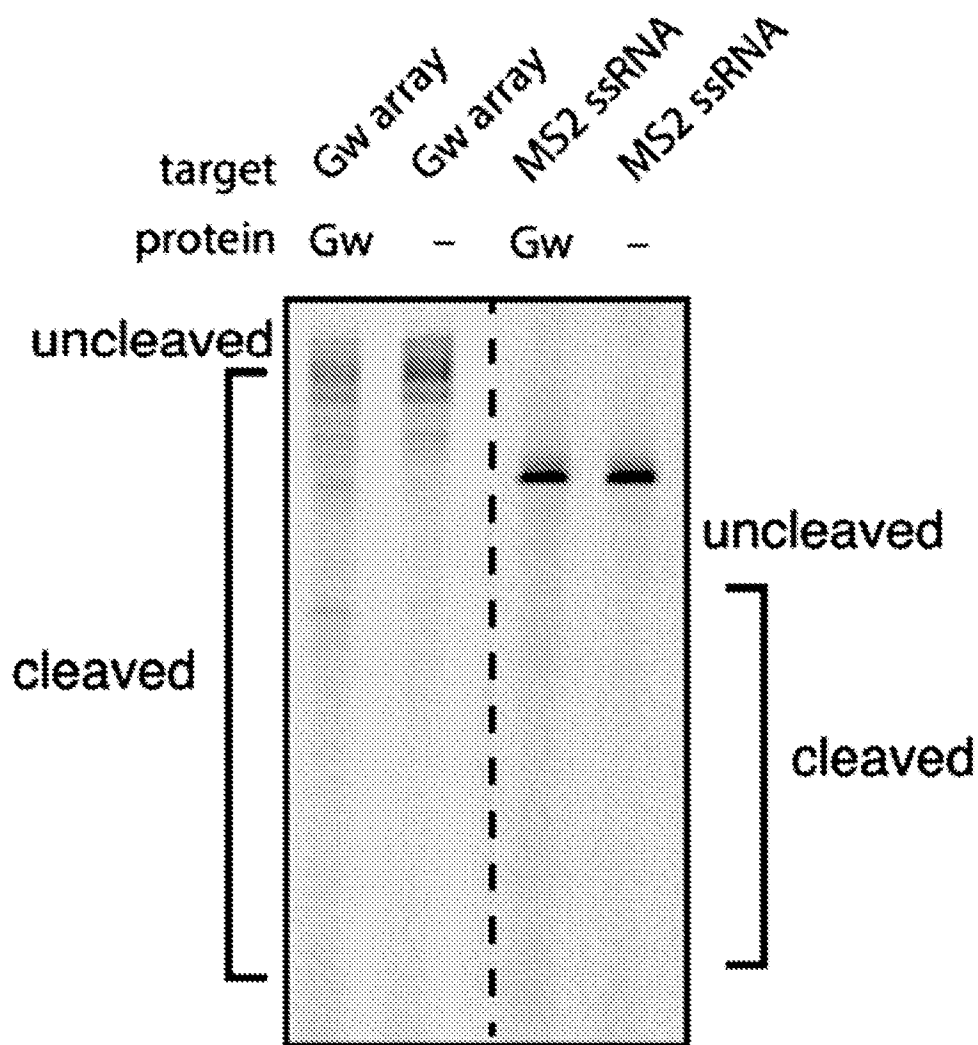


FIG. 21B

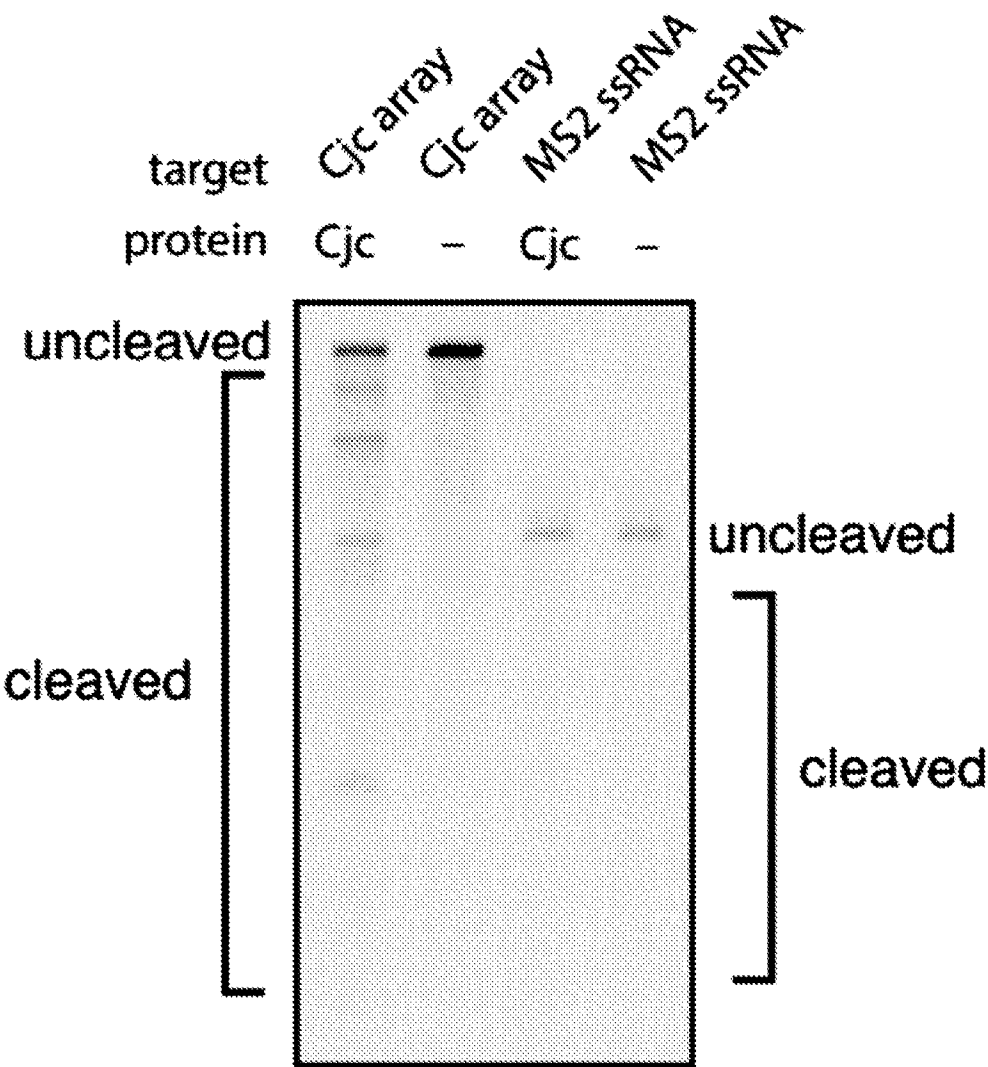
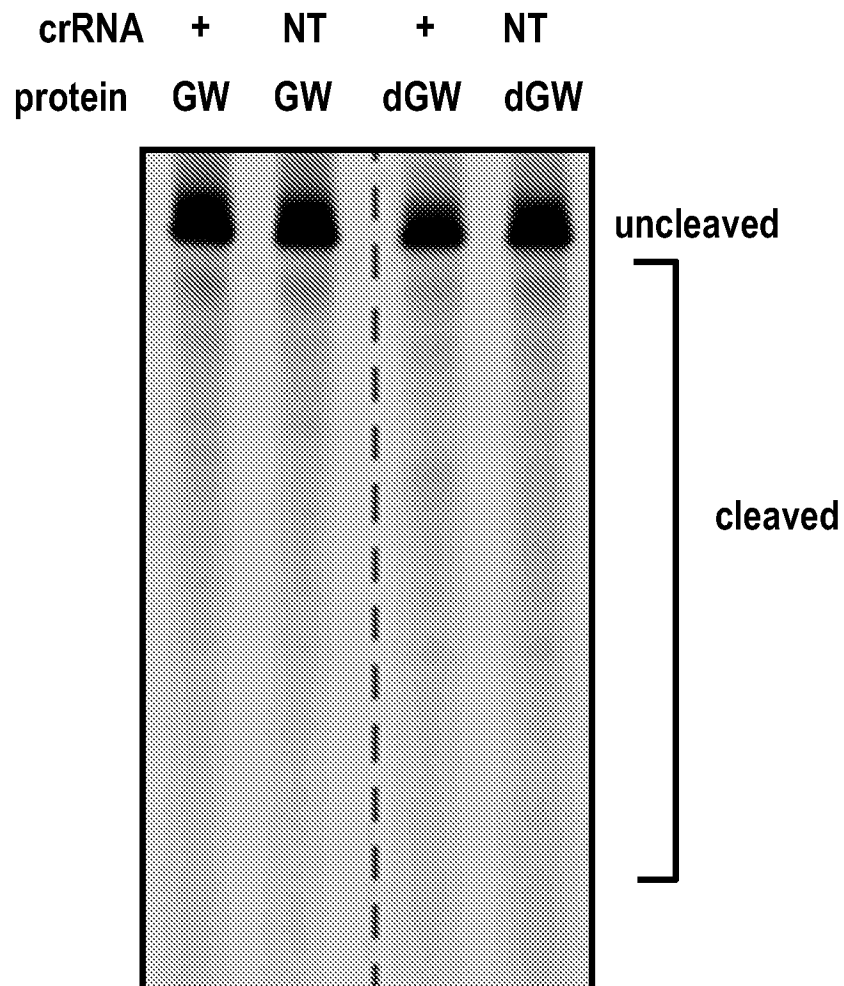


FIG. 21C

104/142

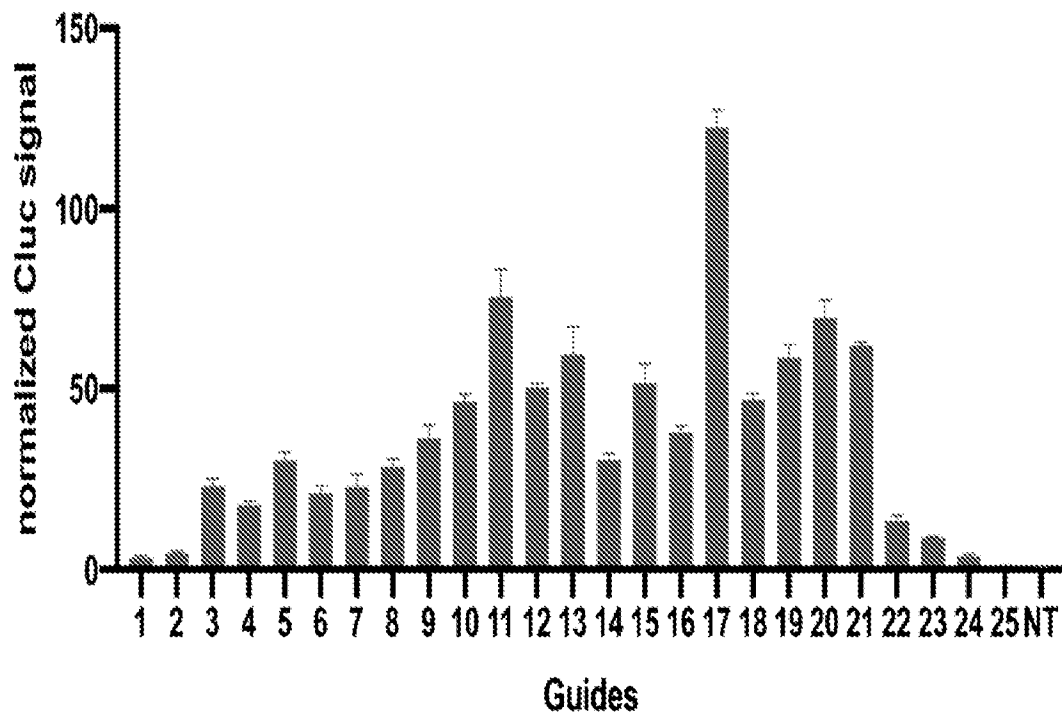
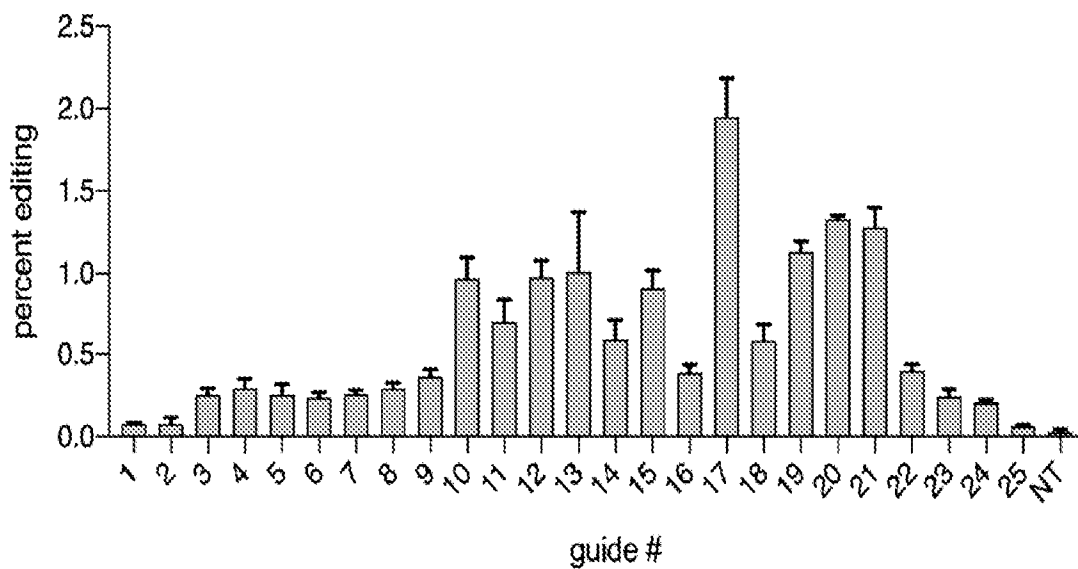
**FIG. 21D**

106/142

sGuide #	Target	MM distance	Guide length	Guide sequence
1	Cluc W85X	50	50	catectgaggcctctactctgcattcaattacatactgacacattggcca
2	Cluc W85X	48	50	accatcctgaggcctctactctgcattcaattacatactgacacattgg
3	Cluc W85X	46	50	aaaccatcctgaggcctctactctgcattcaattacatactgacacattc
4	Cluc W85X	44	50	ctaaaccatcctgaggcctctactctgcattcaattacatactgacacat
5	Cluc W85X	42	50	ttctaaccatcctgaggcctctactctgcattcaattacatactgacac
6	Cluc W85X	40	50	tgtctaaaccatcctgaggcctctactctgcattcaattacatactgac
7	Cluc W85X	38	50	aattgtctaaaccatcctgaggcctctactctgcattcaattacatactg
8	Cluc W85X	36	50	agaatgtctaaaccatcctgaggcctctactctgcattcaattacatac
9	Cluc W85X	34	50	atagaatgtctaaaccatcctgaggcctctactctgcattcaattacat
10	Cluc W85X	32	50	ccatagaatgtctaaaccatcctgaggcctctactctgcattcaattac
11	Cluc W85X	30	50	ttccatagaatgtctaaaccatcctgaggcctctactctgcattcaatt
12	Cluc W85X	28	50	ctttccatagaatgtctaaaccatcctgaggcctctactctgcattcaa
13	Cluc W85X	26	50	cttttccatagaatgtctaaaccatcctgaggcctctactctgcattc
14	Cluc W85X	24	50	atcttttccatagaatgtctaaaccatcctgaggcctctactctgcatt
15	Cluc W85X	22	50	ggaatcttttccatagaatgtctaaaccatcctgaggcctctactctgc
16	Cluc W85X	20	50	tggaaatcttttccatagaatgtctaaaccatcctgaggcctctactct
17	Cluc W85X	18	50	actggaatcttttccatagaatgtctaaaccatcctgaggcctctact
18	Cluc W85X	16	50	gaactggaatcttttccatagaatgtctaaaccatcctgaggcctcta
19	Cluc W85X	14	50	tggaaatggaatcttttccatagaatgtctaaaccatcctgaggccte
20	Cluc W85X	12	50	cctggaactggaatcttttccatagaatgtctaaaccatcctgaggcc
21	Cluc W85X	10	50	ttctgggaactggaatcttttccatagaatgtctaaaccatcctgagg
22	Cluc W85X	8	50	ggttcctggaactggaatcttttccatagaatgtctaaaccatcctgag
23	Cluc W85X	6	50	caggttctggaactggaatcttttccatagaatgtctaaaccatcct
24	Cluc W85X	4	50	accaggttctggaactggaatcttttccatagaatgtctaaaccatc
25	Cluc W85X	2	50	gtaccaggttctggaactggaatcttttccatagaatgtctaaacca
NT	Non-targeting	NA	31	GGTAATGCCTGGCTTGTGACGCATAGTCTG

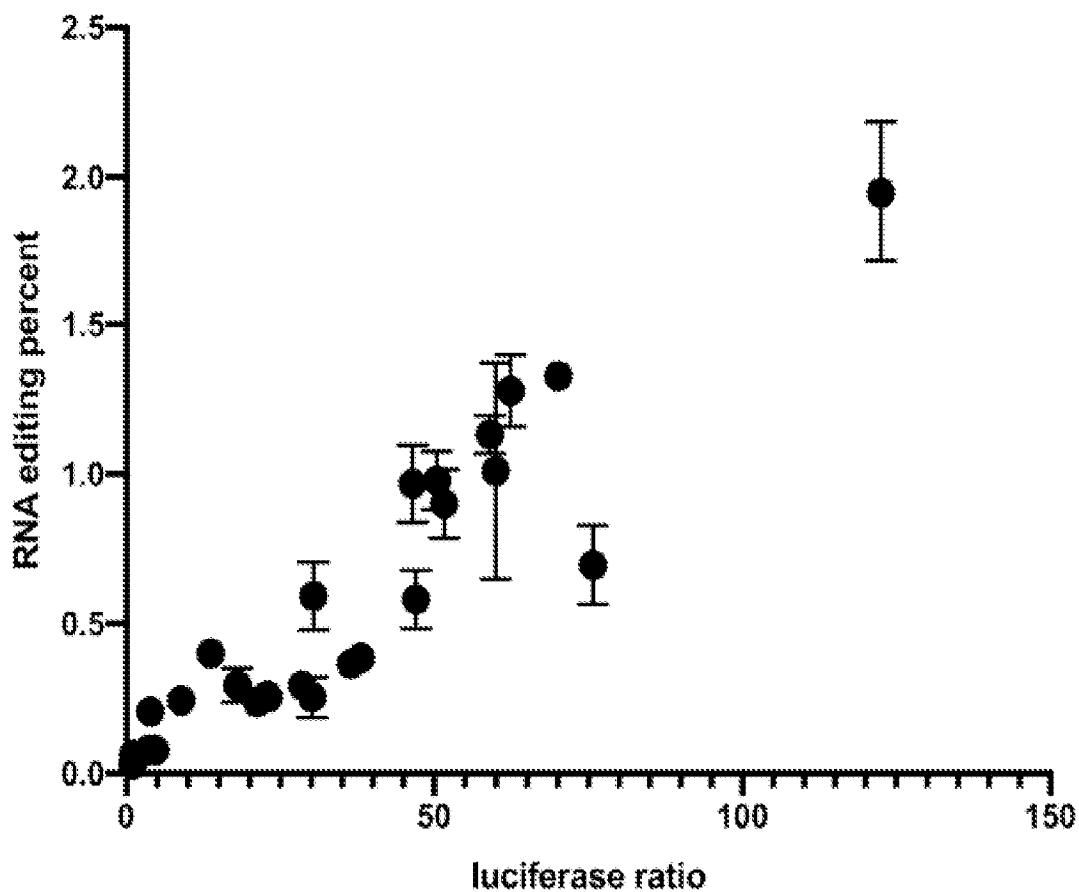
FIG. 22B

107/142

**FIG. 22C****FIG. 22D**

108/142

luc vs editing

**FIG. 22E**

modulate protein function

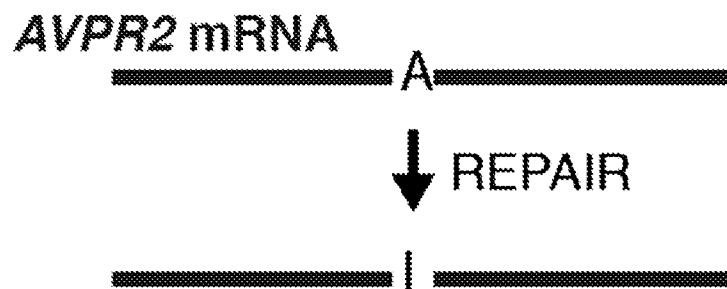
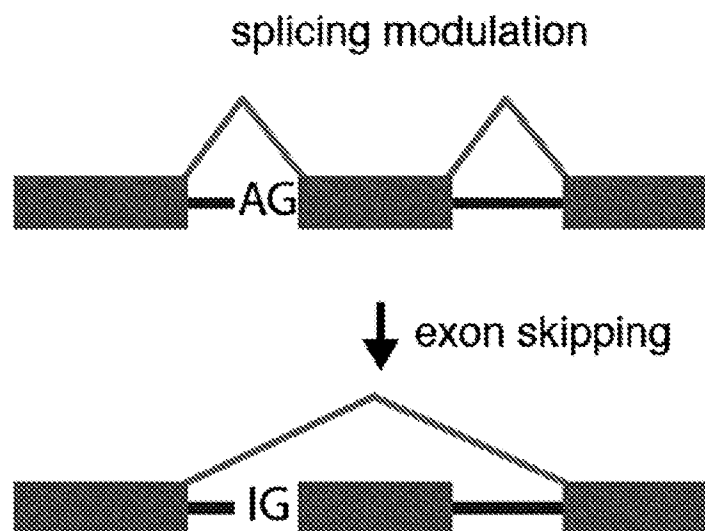
KRAS S12G (A>I)

Constitutively
active**FIG. 23A**

109/142

correction of mendelian disease

(AVPR2 878G>A, Trp293Ter)
X-linked Nephrogenic diabetes insipidus

**FIG. 23B****FIG. 23C**

110/142

<u>initial codon</u>		<u>edited codon</u>
Tyr	→	Cys
stop	→	Trp
His	→	Arg
Gln	→	Arg
Asn	↘ ↗	Ser Asp
Lys	↘ ↗	Arg Glu
Ser	→	Gly
Arg	→	Gly
Thr	→	Ala
Met	→	Val
Ile	↘ ↗	Met Val
Asp	→	Gly
Glu	→	Gly

FIG. 24A

111/142

		second position				
		U	C	A → G		
first position	U	UUU Phe UUC UUA Leu UUG	UCU Ser UCC UCA UCG	UAU Tyr UAC UAA stop UAG stop	UGU Cys UGC UGA stop UGG Trp	U C A G
	C	CUU Leu CUC CUA CUG	CCU Pro CCC CCA CCG	CAU His CAC CAA Gln CAG	CGU Arg CGC CGA CCG	U C A G
	A	AUU Ile AUC AUA AUG Met	ACU Thr ACC ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U C A G
	G	GUU Val GUC GUA GUG	GCU Ala GCC GCA GCG	GAU Asp GAC GAA Glu GAG	GGU Gly GGC GGA GGG	U C A G

FIG. 24B

112/142

<i>initial codon</i>	<i>final codon</i>	<i>potential function</i>
Leu	→ Phe	
Gln	→ stop	
His	→ Tyr	<i>phosphorylation catalysis</i>
Ala	→ Val	
Pro	→ Ser → Leu	
Ser	→ Leu → Phe	<i>phosphorylation catalysis glycosylation</i>
Thr	→ Ile → Met	<i>phosphorylation catalysis glycosylation</i>
Arg	→ stop → Cys → Trp	<i>methylation catalysis</i>

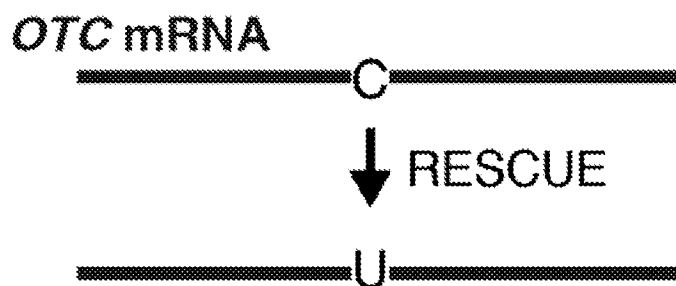
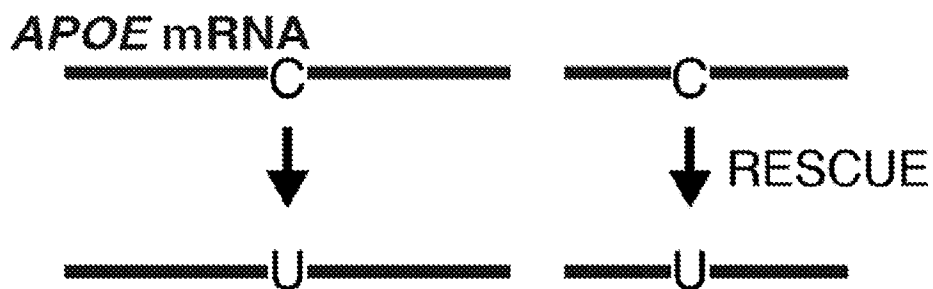
FIG. 25

113/142

correction of mendelian disease

(OTC 134T>C, Leu45Pro)

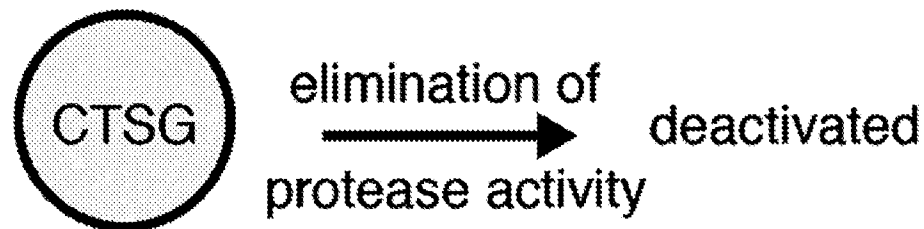
Ornithine carbamoyltransferase deficiency

**FIG. 26A**multiplexed creation of
disease-protective alleles**FIG. 26B**

114/142

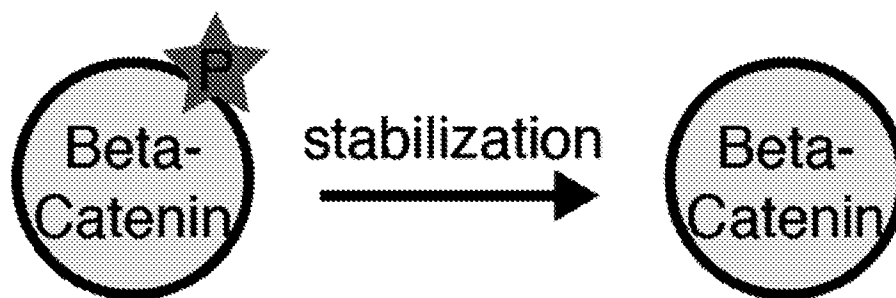
modulate catalytic activity

CTSG H12G (C>Y)

**FIG. 26C**

alter post-translation modifications

CTNNB1 S33F (C>U)

**FIG. 26D**

115/142

SEQ ID NO	Database	Genome/ Metagenome ID	Protein ID/Contig	Cas7-11 name
1	NCBI	NZ_BAFH01000003	WP_007220849	CjcCas7-11b
2	NCBI	JRYO01000185	KHE91659	CsbCas7-11b
3	NCBI	NBMK01000156	OQY58162	DsbaCas7-11a
4	NCBI	JPDT01001326	KPA14974	CmaCas7-11d
5	NCBI	NZ_BEXT01000001	WP_124327589	DisCas7-11a
6	NCBI	LAQJ01000233	KKO18793	CbfCas7-11b
7	NCBI	QMMU01000439	RLC14096	DpbaCas7-11a
8	NCBI	MGTA01000040	OGR07205	Dpba1Cas7-11d
9	NCBI	QMMU01000323	RLC15988	Dpba2Cas7-11d
10	NCBI	MVRP01000104	OPY65763	Sba1Cas7-11b
11	NCBI	QMMU01001221	RLC02083	Dpba3Cas7-11d
12	NCBI	QMMU01001221	RLC02082	Dpba4Cas7-11d
13	NCBI	MVRP01000104	OPY65764	Sba2Cas7-11c
14	NCBI	RFJP01000735	RME63343	NbaCas7-11b
15	JGI	3300021493	Ga0190306_10003932	HvsCas7-11b
16	JGI	3300019457	Ga0193932_104825	SstCas7-11a
17	JGI	3300021508	Ga0190283_10011062	HvmCas7-11d
18	JGI	3300005265	Ga0073580_1036305	HsmCas7-11b
19	JGI	3300002821	Iso3TCLC_1001005823	HreCas7-11d
20	NCBI Metagenome	ASM579580v1	SESD01000293.1	FmCas7-11a
21	NCBI Metagenome	OBJA01000000	OBJA01001127	SmCas7-11c
22	NCBI Metagenome	OBEQ01000000	OBEQ011807420	GwCas7-11c
23	ENA	OVOO01000000	OVOO01000106	WmCas7-11c
24	ENA	PDWI01000000	PDWI01005922.1	OmCas7-11c
25	NCBI Metagenome	VAPF01001339.1	-	HvmCas7x3
26	NCBI	DRKI01000155.1	HEB50754.1	DesCas7x3
27	NCBI	DRNY01000543.1	HHJ20681.1	GabCas7x3
28	NCBI	DTXS01000070.1	HID69069.1	DsbCas7x3
29	NCBI	JABFST010000317.1	NOS89056.1	MebCas7x3
30	NCBI	PDPY01000001.1	PID64649.1	GamCas7x3
31	NCBI	NZ_JMLA01000001.1	WP_027150709.1	MetCas7x3
32	NCBI	NZ_FOGH01000010.1	WP_090332053.1	NisCas7x3
359	NCBI	MVRP01000104.1	-	SybCas7-11

FIG. 27A

116/142

SEQ ID NO	Cas7-11 effector annotation
1	TIGR03986 family CRISPR-associated RAMP protein [Candidatus Jettenia caeni]
2	RAMP superfamily protein [Candidatus Scalindua brodae]
3	hypothetical protein B6245_13290 [Desulfobacteraceae bacterium 4572_88]
4	protein containing DUF324 [Candidatus Magnetomorum sp. HK-1]
5	hypothetical protein [Desulfonema ishimotonii]
6	RAMP superfamily protein, partial [Candidatus Brocadia fulgida]
7	TIGR03986 family CRISPR-associated RAMP protein, partial [Deltaproteobacteria bacterium]
8	hypothetical protein A2511_12465 [Deltaproteobacteria bacterium RIFOXYD12_FULL_50_9]
9	TIGR03986 family CRISPR-associated RAMP protein, partial [Deltaproteobacteria bacterium]
10	RAMP superfamily protein [Syntrophorhabdaceae bacterium PtaU1.Bin034]
11	hypothetical protein DRI57_30415, partial [Deltaproteobacteria bacterium]
12	hypothetical protein DRI57_30410, partial [Deltaproteobacteria bacterium]
13	hypothetical protein A4E57_03052 [Syntrophorhabdaceae bacterium PtaU1.Bin034]
14	hypothetical protein D6778_09755, partial [Nitrospirae bacterium]
15	CRISPR/Cas system CMR subunit Cmr4 (Cas7 group RAMP superfamily)
16	CRISPR/Cas system CMR subunit Cmr4, Cas7 group, RAMP superfamily
17	cold shock CspA family protein/CRISPR/Cas system CMR subunit Cmr4 (Cas7 group RAMP superfamily)
18	Uncharacterized protein predicted to be involved in DNA repair (RAMP superfamily)
19	Cold shock proteins
20	None
21	None
22	None
23	None
24	None
25	None
26	None
27	None
28	None
29	None
30	None
31	None
32	None
359	None

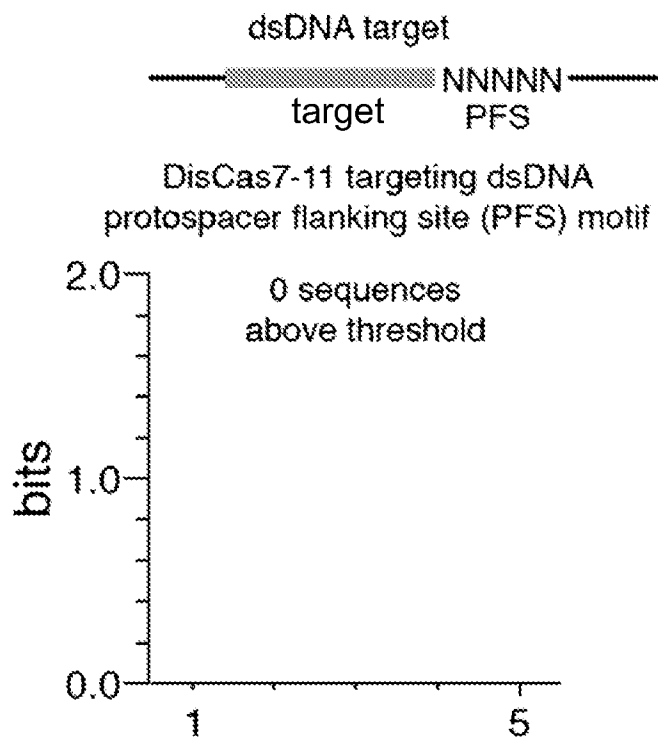
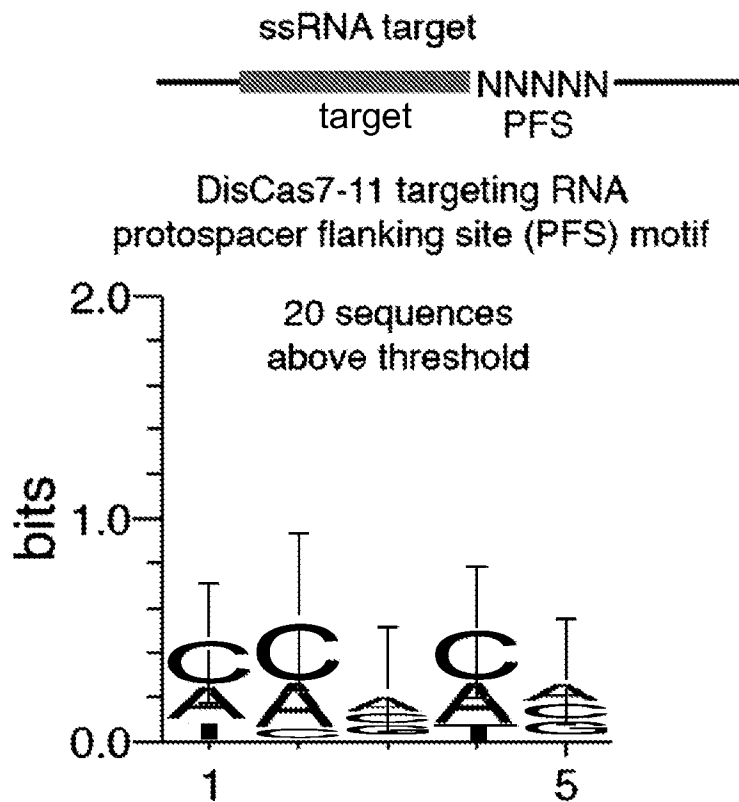
FIG. 27B

117/142

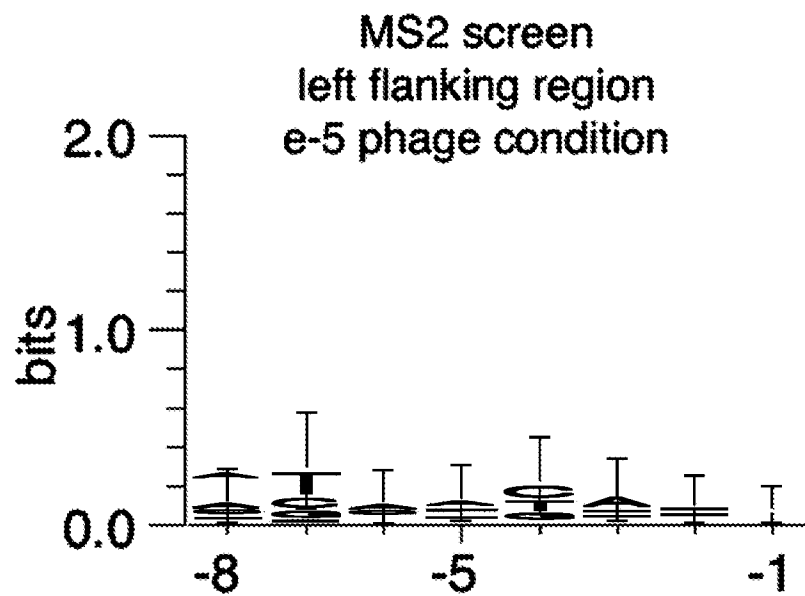
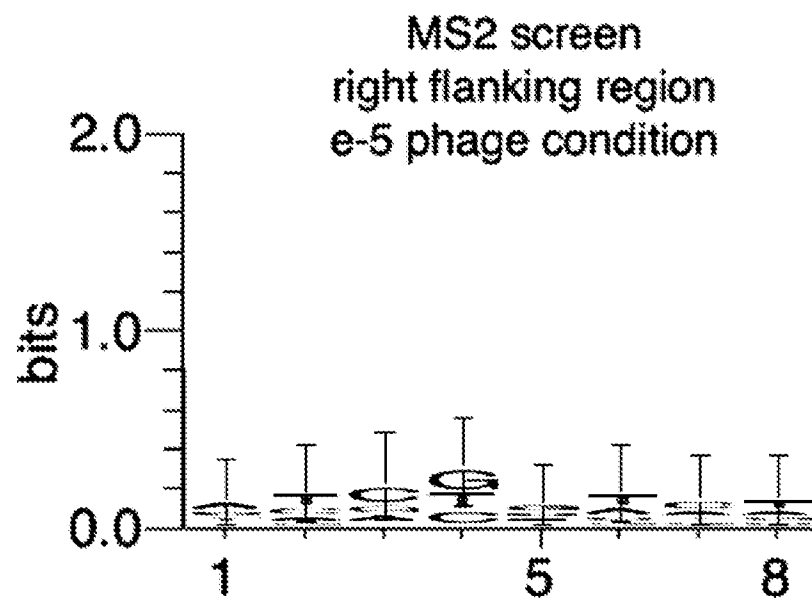
SEQ ID NO	Organism/source	Length
1	Candidatus Jettenia caeni	1812
2	Candidatus Scalindua brodae	1717
3	Desulfobacteraceae bacterium 4572_88	1659
4	Candidatus Magnetomorum sp. HK-1	1657
5	Desulfonema ishimotonii	1601
6	Candidatus Brocadia fulgida	1573
7	Deltaproteobacteria bacterium	1526
8	Deltaproteobacteria bacterium RIFOXYD12_FULL_50_9	1403
9	Deltaproteobacteria bacterium	1367
10	Syntrophorhabdaceae bacterium PtaU1.Bin034	1322
11	Deltaproteobacteria bacterium	476
12	Deltaproteobacteria bacterium	437
13	Syntrophorhabdaceae bacterium PtaU1.Bin034	350
14	Nitrospirae bacterium	348
15	Hydrothermal vent sediment bacterial communities from Southern Trench, Guaymas Basin, Mexico - 4870-07-3-4_MG	1528
16	Sorted cell/s from Southern Trench hydrothermal vent microbial mat, Guaymas Basin, Mexico - 6X_4868_18_01	1652
17	Hydrothermal vent microbial mat bacterial communities from Southern Trench, Guaymas Basin, Mexico - 4869-18-0-1_MG	1723
18	Hydrothermal sediment microbial communities from Guaymas Basin, California, USA 4484. Combined assembly of Gp0115313 and Gp0146561	1548
19	Hydrocarbon resource environments microbial communities from Canada and USA	1720
20	freshwater metagenome, University of Waterloo	1610
21	soil metagenome genome assembly	1575
22	groundwater metagenome genome assembly	1237
23	wastewater metagenome [species]	1291
24	oral metagenome	1801
25	Hydrothermal vent metagenome isolate SMAR3k101_15	1322
26	Desulfobulbus sp.	1225
27	Gammaproteobacteria bacterium	1234
28	Desulfobulbus sp.	1258
29	Methylococcaceae bacterium	1235
30	Gammaproteobacteria bacterium	1276
31	Methylobacter tundripaludum	1221
32	Nitrosomonas sp Nm51	1444
359	Syntrophorhabdaceae bacterium PtaU1.Bin034 A4E57_contig000104	1329

FIG. 27C

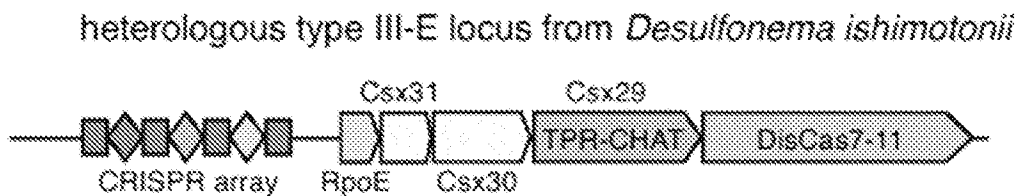
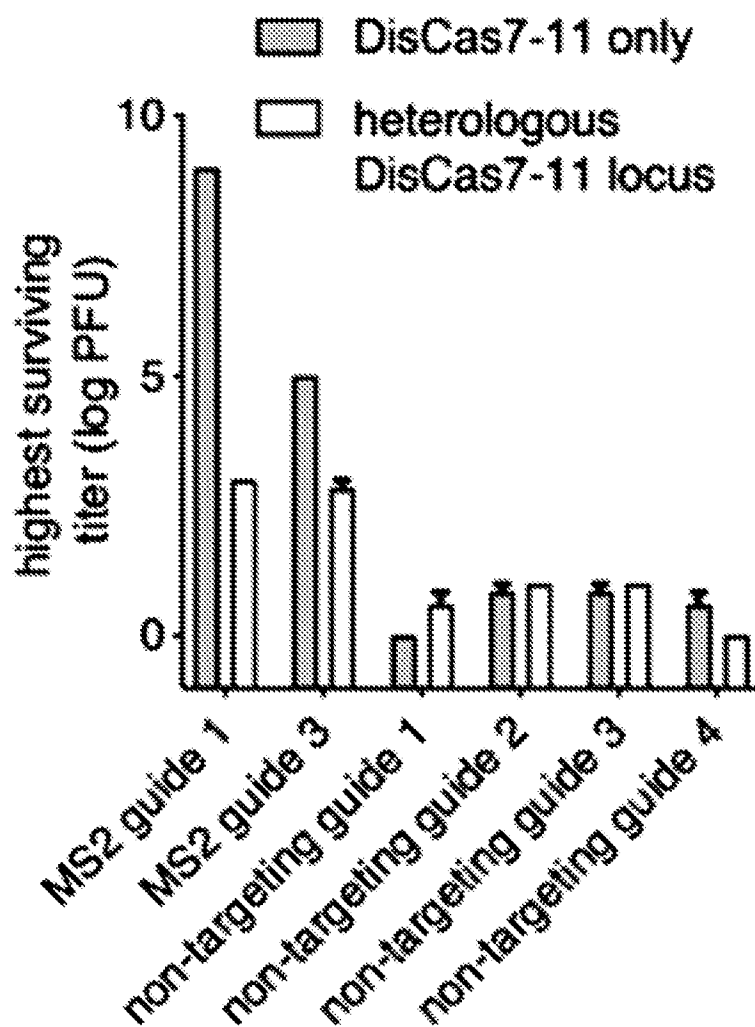
118/142

**FIG. 28A****FIG. 28B**

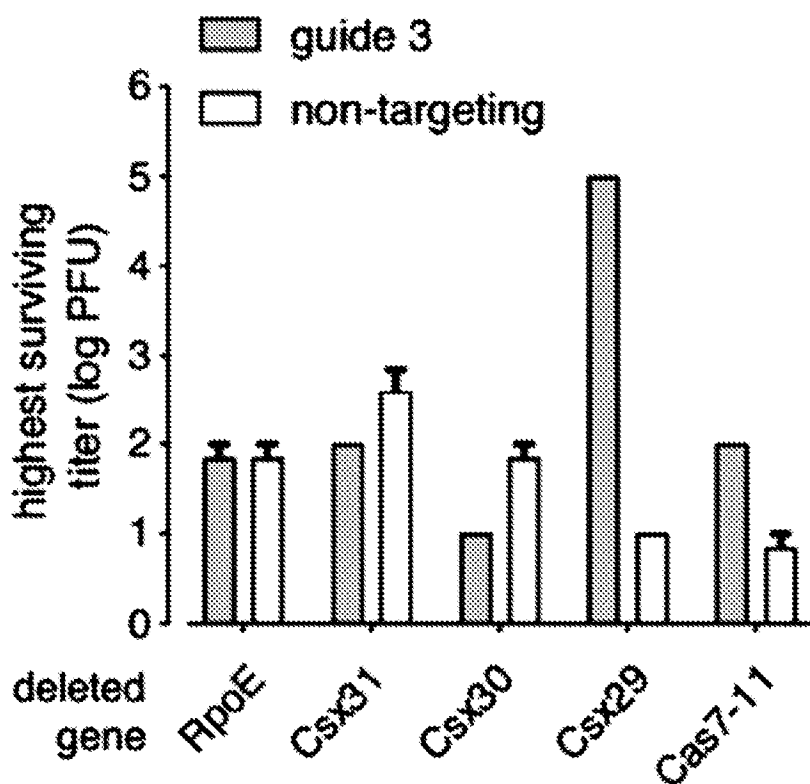
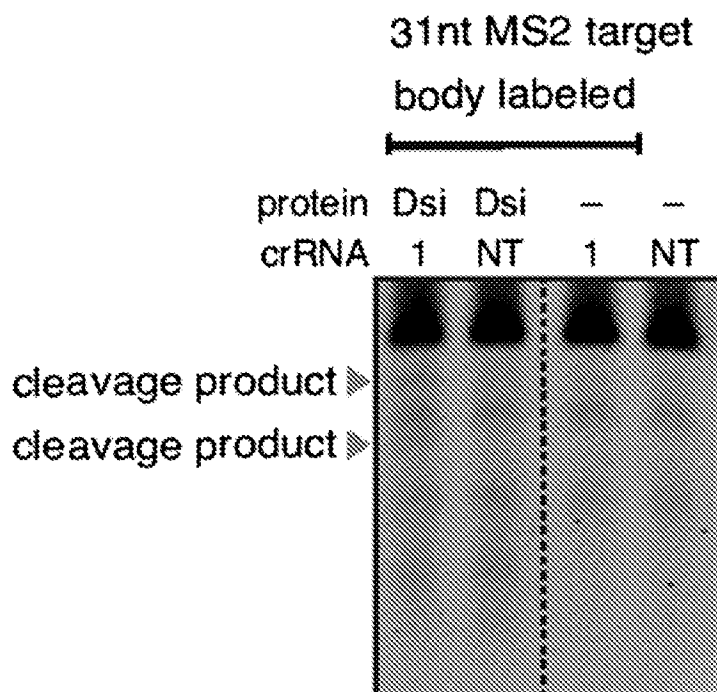
119/142

**FIG. 28C****FIG. 28D**

120/142

**FIG. 28E****FIG. 28F**

121/142

**FIG. 28G****FIG. 29**

122/142

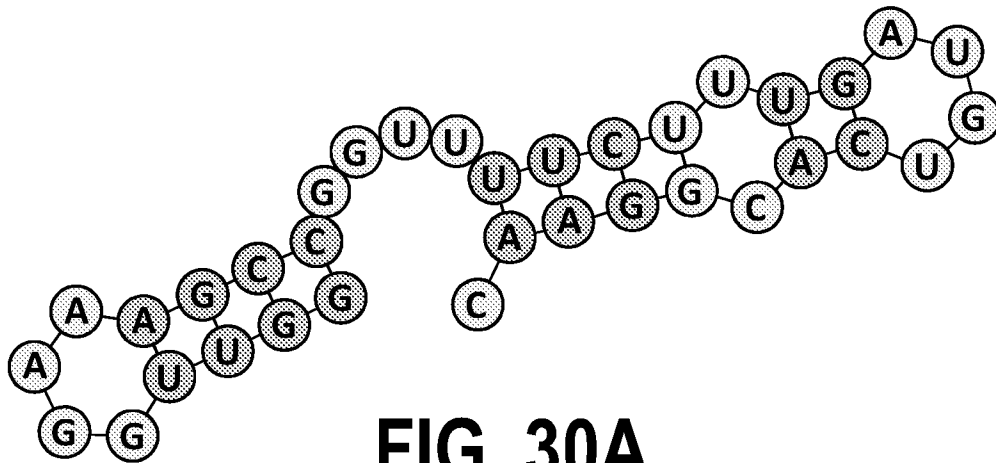


FIG. 30A

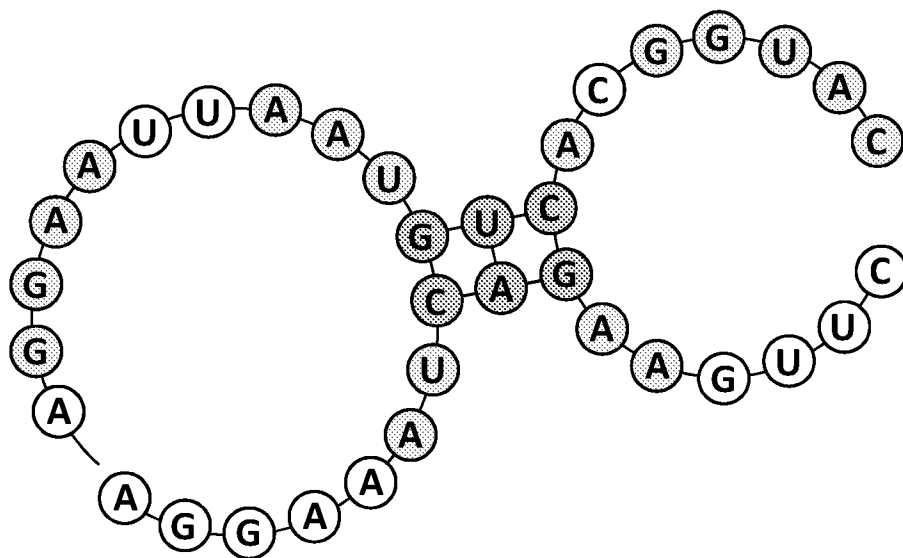


FIG. 30B

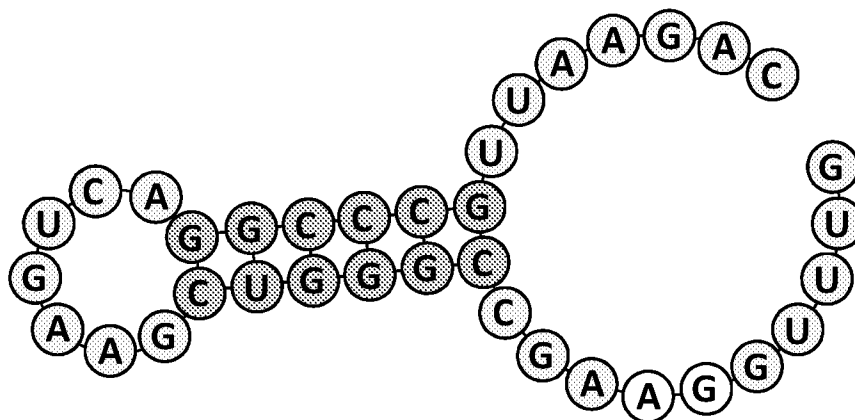
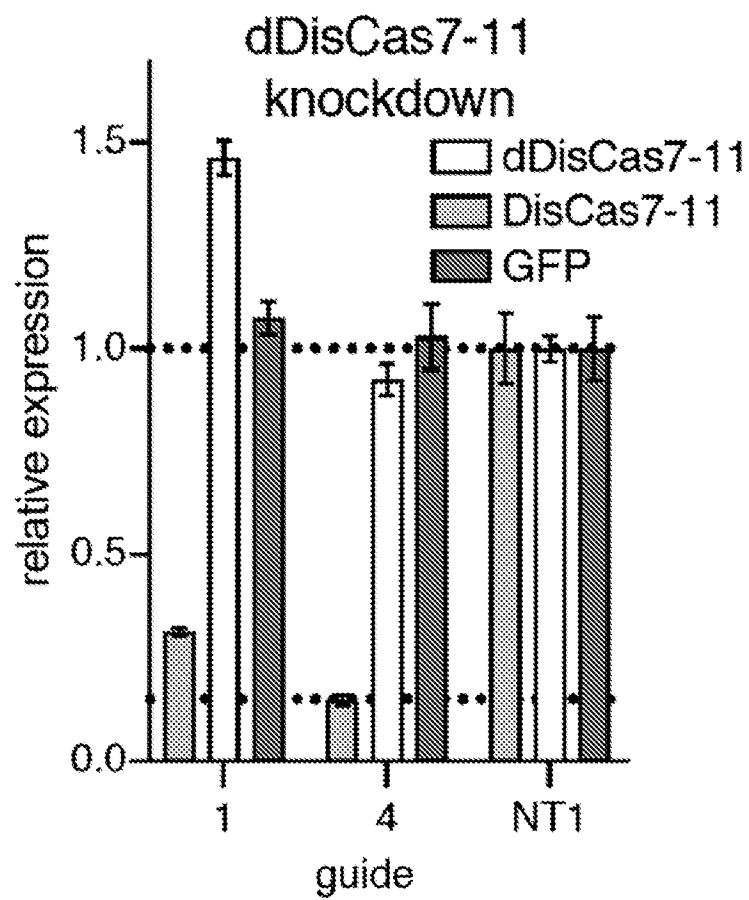


FIG. 30C

123/142

**FIG. 31A**

124/142

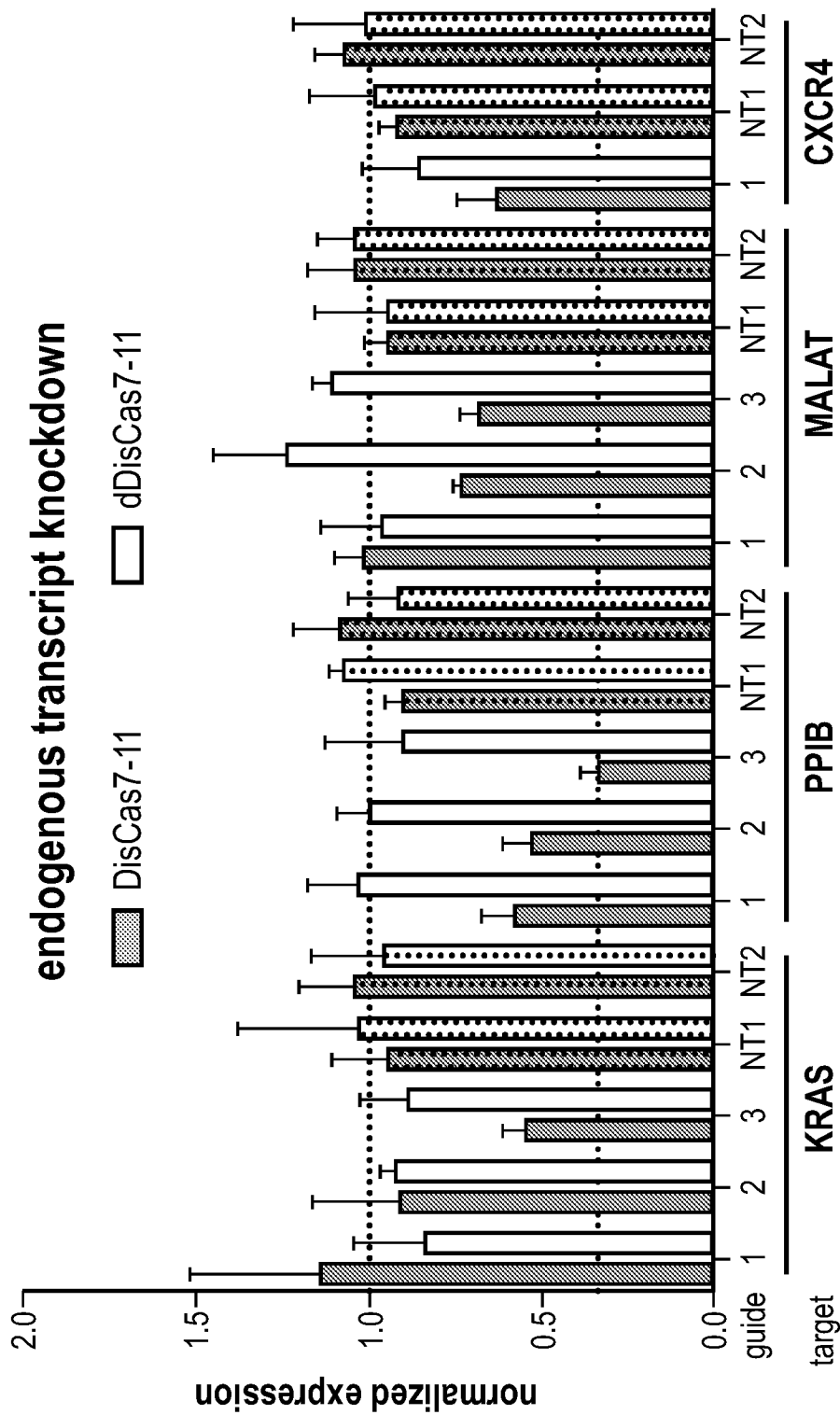
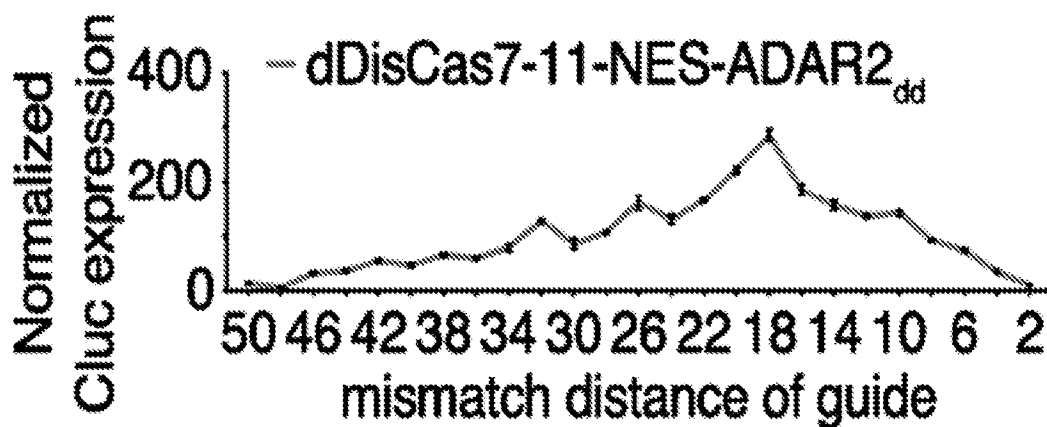
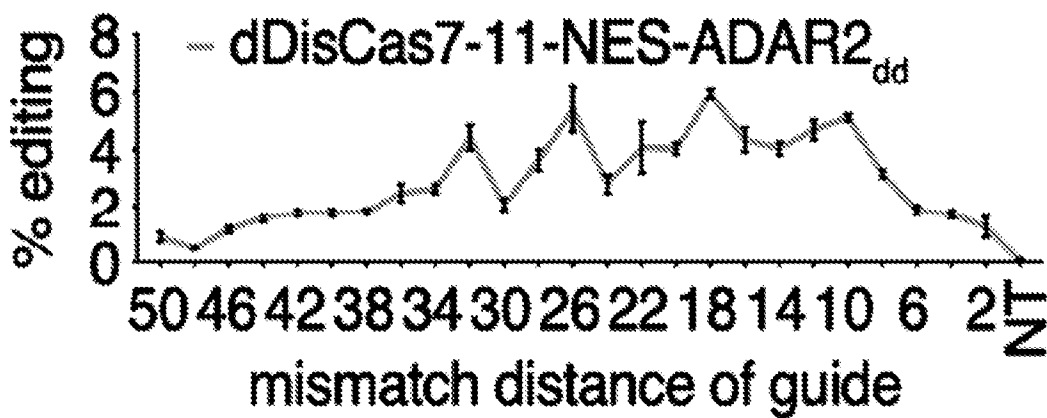
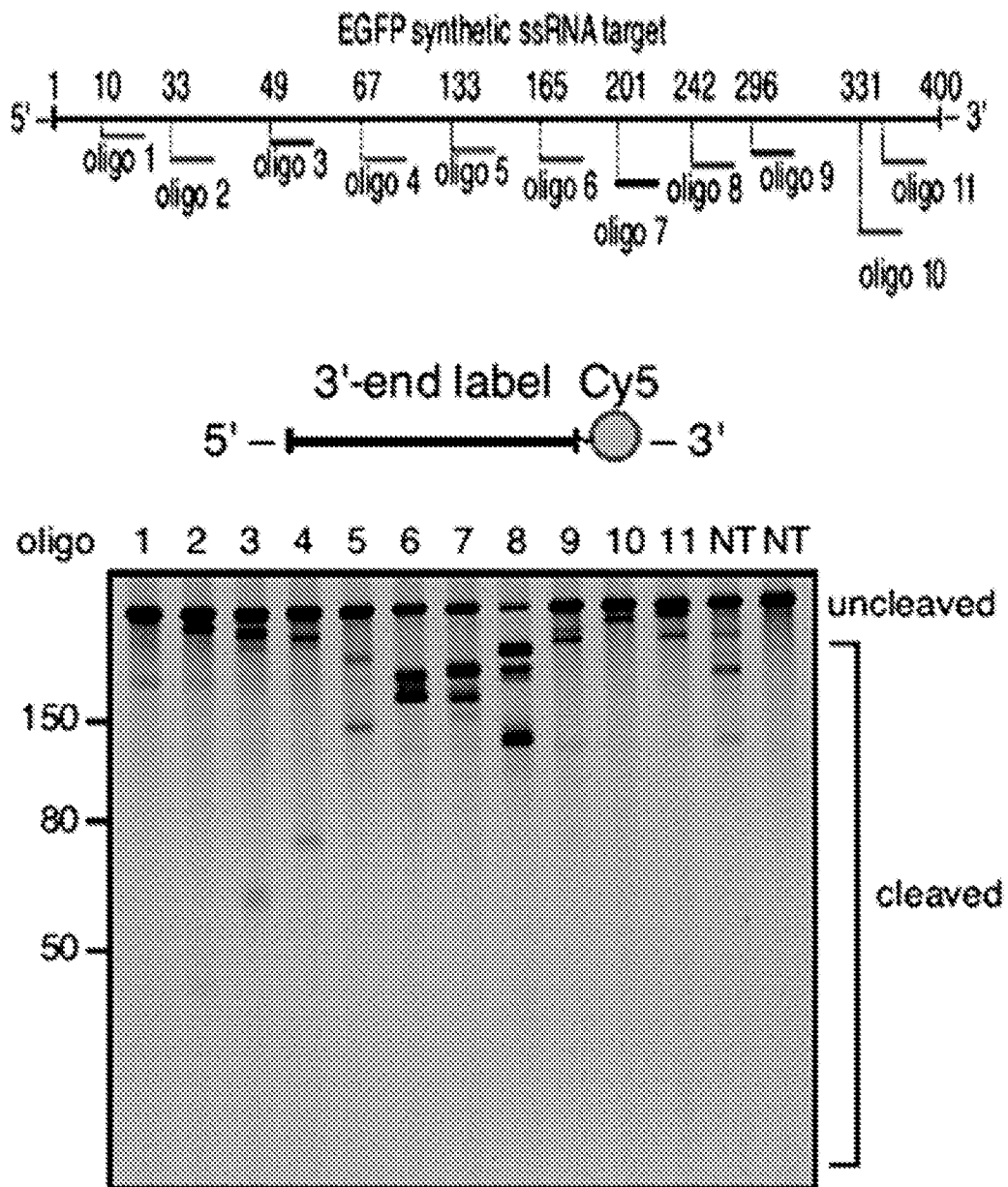


FIG. 31B

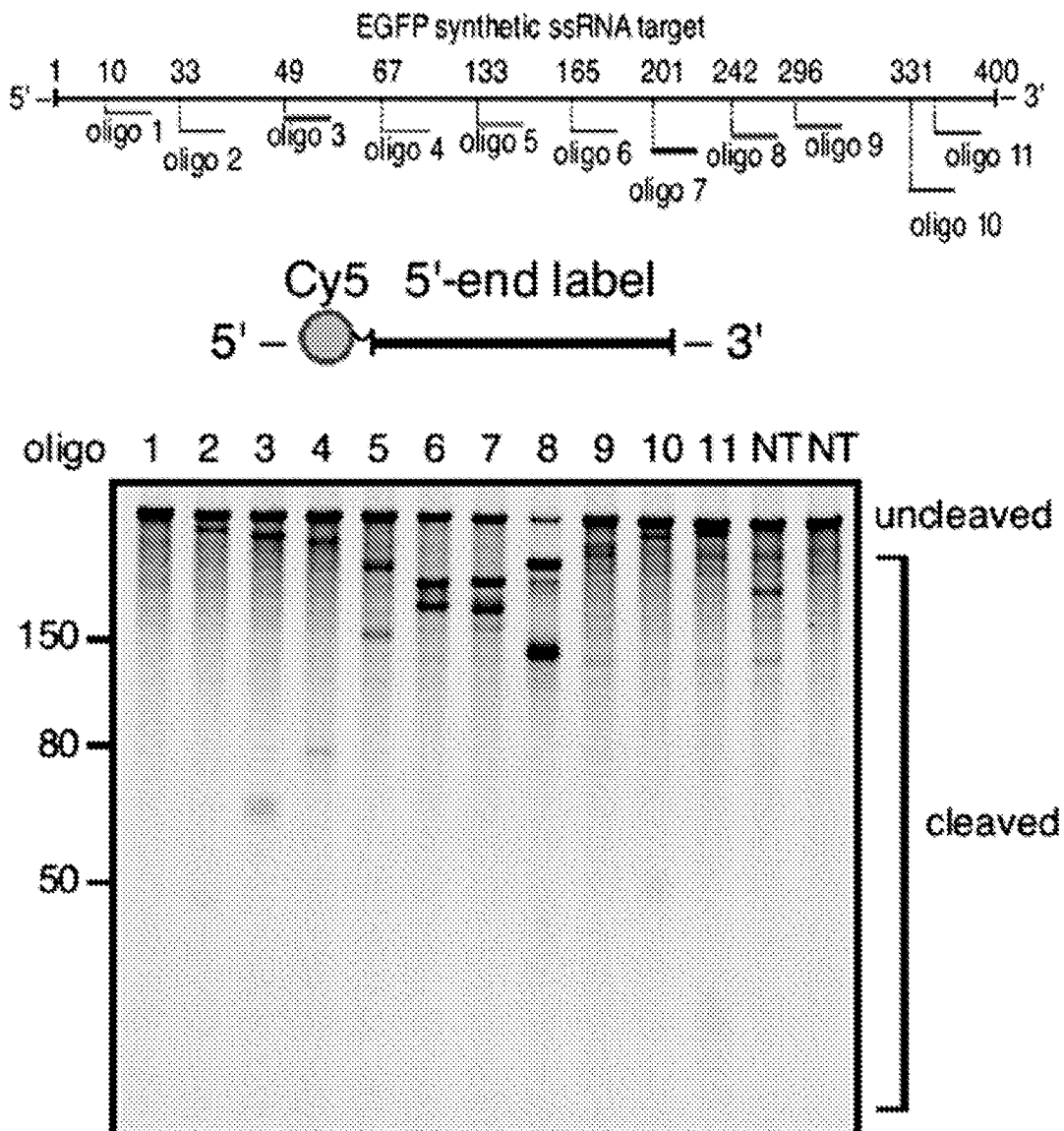
125/142

**FIG. 32A****FIG. 32B**

126/142

**FIG. 33A**

127/142

**FIG. 33B**

СИТРИН

RTTEGKLEKPL	YLRRRD	GFEFNH	AERPC	SEEDP	CPLC	LIQGRF	KILRDE	KQFVRQ	GNIS	SECSVN	FNSLN	ISS	GIKFSWE	EI
RTEGKHORS	FLRRKP	VYTENK	NIQAC	--NK	CPLC	LIQGRD	DRGED	LKRKHYN	GKHV	DNWTV	HFSN	FD	TOATFYWK	ET
KTEGNAPL	FLRRRP	--	RADC	DPQCT	CPLC	LIQGRS	DRVED	LAKRKH	Y	DWSV	HFNK	EA	TERSFNNK	ET
ITPDGOTPKP	HLRRRA	TVRABE	AKDCK	A	DROSA	CPFL	LIQGRF	QVCKD	GGKGL	RDVFRS	NLND	LPKDI	SPRDFDGO	EI
NSQAKKPKP	HLRRRT	TLWQPT	GAKSKD	GRSNAC	PF	LIQGRF	QVCKD	GGKGL		RDVFRS	NLND	LPKDI	SSPKVMSFG	EI
ETEDKRLI	OLRRS	TLRWT	DKNPK	PD	NAETV	CPFL	LIQGRS	GVNGK	AEKKRW	RFRH	FGNLS	LP	GKPTFGGK	EI
LTETDKRPL	OLRRRP	TLRWT	DKNPK	PD	DPFPP	CPFL	LIQGRS	GVNGK	AEKKRW	NSYV	FGNLS	LP	GKPTFGGK	EI
RUENRKRDAM	HLRRS	TLRWT	DKNPK	PD	NGSL	CPFL	LIQGRS	GVNGK	AEKKRW	NSYV	FGNLS	LP	GKPTFGGK	EI
QIDESKGPPL	FLRRH	TLRWT	DKNPK	PD	KREAC	CPFL	LIQGRD	NAGKVR	HERNK	DYDI	HFSN	LD	PLNRVDDPS	EI
ETEDKRDPA	VLRRP	TLRWT	DKNPK	PD	QERKDA	CPFL	LIQGRD	NAGKVR	HERNK	DYDI	HFSN	LD	PLNRVDDPS	EI
ETEREMRKNT	FLRRP	TPAMAT	GKKEIC	KE	THESAC	CPFL	LIQGRS	LHGKED	VNEADPS	CRKP	VGFGNLS	LP	FOPTKROLO	DV
HTEGEPDS	FLRRKP	TVRAEK	DEIC	T	SEBTA	CPFL	LIQGRD	NAGKVR	HERNK	CRKP	VGFGNLS	LP	FOPTKROLO	DV
KGSKAKYKVE	CPRRHR	ATLKW	TNDT	T	DYNAC	CPFL	LIQGRD	NAGKVR	HERNK	CRKP	VGFGNLS	LP	FOPTKROLO	DV
DTKGPST	HLRRP	TLRWT	GKTYC	R	SEADV	CPFL	LIQGRD	NAGKVR	HERNK	CRKP	VGFGNLS	LP	FOPTKROLO	DV
QSENPNPDX	YLRRP	WIELHA	QKEIC	K	TRDTAC	PL	LIQGRD	NAGKVR	HERNK	CRKP	VGFGNLS	LP	FOPTKROLO	DV

FIG. 34A

129/142

	D177	atypical G-rich loop
WmCas7x3		
SER16298.1		
GwCas7x3		
PID64649.1		
WP_031436019.1		
HEB50754.1		
HreCas7-11		
CmaCas7-11		
HvmCas7-11		
SmCas7x3		
oral meta		
DisCas7-11		
DsbaCas7-11		
SstCas7-11		
CsbCas7-11		
CjcCas7-11		
HvsCas7-11		
HsmCas7-11		
FmCas7-11		
DpbaCas7-11		
Domains		
WmCas7x3		
SER16298.1		
GwCas7x3		
PID64649.1		
WP_031436019.1		
HEB50754.1		
HreCas7-11		
CmaCas7-11		
HvmCas7-11		
SmCas7x3		
oral meta		
DisCas7-11		
DsbaCas7-11		
SstCas7-11		
CsbCas7-11		
CjcCas7-11		
HvsCas7-11		
HsmCas7-11		
FmCas7-11		
DpbaCas7-11		
Domains		

FIG. 34B

[illegible]

FIG. 34C

131/142

4	WmCas7x3	WQADIALDADGRVYLPASLKGALKALARRRQVADAPCL-PLFGDLNRGGAPHPDCIPPRTRAGLAERDALQSHAT--QAD--GPD	-----AQPRIA
	SER16298.1	WQADIALDNLKPYPGPASLKGALKALARRYCASNL--NIFGDTI--DTKGDNRKRSVTVQAQAEFLSA--WCVAD--QED--KPF	-----TITRVA
	GwCas7x3	WLEVALDHKGQYIPGASLKGALKALAKRHD-----FNLF--DNKEVDG--FVROAEFLSA--WCVAD--TK--GRL--	-----IQPRVA
	PID64649.1	EVNSVTGKAALPIPGSTIKGRRLRHYSKHSFKDALLNKVFGEES--DATDDQG--RGLAEFHDAKNPEK--NRNLQGRYPYWNTRKTYEVSTA	-----
	WP 031436019.1	KINACIKGSKRPIIPGSTIKGRFLARLBERGVDSALLETFKGGH--NSVSQDQ--RGGAEFHDAEIEPL--TGS--QPWPYWRHEHOAFIAASTA	-----
	HE50754.1	QINALLTGAGGLPIPGSTIKGRFLARLBERGVDSALLETFKGGH--DRETEQD--RGGAEFHDAEIEPL--TGS--RHFYWRPERQIWIKAQTA	-----
	HreCas7-11	--QAILCPGGSYRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	CmaCas7-11	--STIFLNMNGQFCIPRSTIRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	HvmCas7-11	--ATILNRDGYRFLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	SmtCas7x3	--SRTLVDRDGRYRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	oral meta	--RKILLTRDFYRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	DisbaCas7-11	--LQVLLTBNKRYRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	SsbCas7-11	--LMILLDQNHVRIPLRSALRGILRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	CsbCas7-11	--LKVLLDRQNRVRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	CjcCas7-11	--FNILLDKENRYRIPLRSALRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	CbfCas7-11	--FRILMDKKRYRIPLRSALRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	HvsCas7-11	--LRLVMDKKGRFRIPLRSALRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	HsmCas7-11	--GAILLTPDNRFRLPRTALRGILRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	FmCas7-11	--MQILLTKDGRYRLPRSLRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	DpbaCas7-11	--MPTLQDSNDHFRLPRTALRGALRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
	Domains	--LPILLTSRDRFRIPLRSALRGILRRDLRLV--GDSCTPVGS--RV--CYCQVCQ--IMRITVQGG--VSS--YRE--PAE--VRQIR	-----
		G-rich loop	
	WmCas7x3	IDRITGSVVDKKLFHTOTVPVGTFRFSVEI--ILRRADQ--NLAAQLVALLQHGPT-----DPDFRLCAHANLGGGRVGL--YGNIDTRRFGPQQAQ--H	-----
	SER16298.1	IDRVGTAEADRLKLENTRRVNPVGVCFNVI--IIQNACE--TEIQYLLDLRKAQK-----DPFSLGAGANQSQKVRCL--LSSC--VRVFGKQEMH--D	-----
	GwCas7x3	IDRVGTAAQDKKLFQTLTPGTFRFAMKI--VQONAVE--NETADLLGLNLLPD-----DPQFSLGAYANQSGRQVQ--FGKIOTRCFGINEAK--A	-----
	PID64649.1	INRHTGAADKSLHTECVPTGTFVEIKITGSMDDRCA--ALVVALEATQTTGS-----RIFL--CAEDANGNGRIGL--TGKITVKQMDQAHII--Q	-----
	WP 031436019.1	IDRHQVALQSLHYMETVPAQIRKFTFTGVMRDEEA--ALLIAALDSFDKNQ-----QPCF--GVDRANAYGRMEL--HGHLLHVYCMGATEIS--S	-----
	HE50754.1	VDRHGTALRRSLRYTEMVPPGVFRFTITGCMTDAEA--DVLFALLELDGDPQO-----ACSF--GGAGADGNGTMRLL--FGRPEVYCLDRSGIL--G	-----
	HreCas7-11	SNPHTGVEEGALFDLETGQGMTPFPRL-----YFRTRSPYIDRALWLITINHWQ-----GKAI--FGGDIQVGMGRFRL--E--NLQIRSAIDLVSRR--D	-----
	CmaCas7-11	LNCHTGVVEEGALFDMETGQGMIFPFPRL-----YVESKNEIMSQHLVEVLNNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	HvmCas7-11	LNCHTGVVEEGALFDMETGQGMTPFPRL-----YVESKNEIMSQHLVEVLNNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	SmtCas7x3	RNPITGTAADGGLFDTERGPKGAGFPFSL-----RYRGN--KOLEGPIRQLLSWWQO-----GLLF--LGDEFGIGKGRFRL--E--NMKIHRDLRDES--A	-----
	oral meta	VDRSGIIVRDGALFDTYEGIEGVCFPLEI-----RYRGN--KOLEGPIRQLLSWWQO-----GLLF--LGDEFGIGKGRFRL--E--NMKIHRDLRDES--A	-----
	DisCas7-11	INPFTGVAEGALFDMETGQGMIFPFPRL-----YVESKNEIMSQHLVEVLNNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	SsbCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	CsbCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	CjcCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	CbfCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	HvsCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	HsmCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	FmCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	DpbaCas7-11	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----
	Domains	LNPTGTVAEKALFDMELGQGLSDFL-----SYRKG--KTLPKLRLNVLNWTN-----GOAF--FGGEAGTGTGRFRL--L--NNEVFLWEIDGEEED	-----

FIG. 34D

SUBSTITUTE SHEET (RULE 26)

133/142

WmCas7x3	VAQG----	HAVPPVKNNTCP--DPATL----	LFGTAGNRGLLRDTD----	CTGTA----
SER16298.1	IPOG----	HAAPAYDGIARH--DLISL----	LFGTAGNKGIVCASLIHSIPEYALQFNGVRETISDLSDTVKSCIIVDLVKTSTAAEKTTEQLHIRIV	---
GwCas7x3	TPQG----	HTAPAYRKGQPHDDLAVL----	LFGAAGWRGIVQTS-----	CIVED-----
PID64649.1	CCDG----	SINNTCK-----	NPNKNCIACEMFGSGWKTSIEMDPFL-----	CVDRE-----
WP 031436019.1	CCS----	TDDPCPLFDKGLSKLCLACQIFGASGKWTVINIHDFK-----	---	AINKS-----
HEB50754.1	CCD----	TSSPCRLGSSDKVGLCLACQVFGAPGWTTLHIQGT-----	---	CTSV-----
HreCas7-11	-DAPNE----	NGLTIFE--ADHEE--	CDCLLSLFGSKHYQGLKFEDAE-----	LQDEV-----
CmaCas7-11	-EIKLE----	DGKKERIFN--LDHED--	CDCLQCRLFNGVHQOGLRFEADAE-----	ITNKN-----
HvmCas7-11	---	NEELYD--TDHED--	CDCLCRLFGSIHQOGLRFEADAE-----	VQNSV-----
SmCas7x3	---	AGDDLIT--RSHQD--	CKCEICQLFGSEHRAGILRFEDLP-----	PVSPPT-----
oral meta	---	YGISLIN--DEHED--	CDPCPLCKIFGNEHAGMLRFDDMV-----	PVGTW-----
DisCas7-11	-HME----	DGVPLTE--LTHSD--	CECLLCQIFGSEYRAGKIRFEDILV-----	FESPP-----
DsbaCas7-11	-NQGNDDIT--GKNVPLIA--	LTHQD--	CECMLCRFFGSEYRAGRLYFEDLT-----	FESPP-----
SstCas7-11	-HFDKEGKPLDKGKPLLT--	LIHQD--	CECLICRLFGSEHETGRLRFEDLL-----	FDPQP-----
CsbCas7-11	---	TGDLG-K--EDHED--	CTCDMCIIFGNEHSSKIRFEDLE-----	LINGN-----
CjcCas7-11	---	SGDLG-M--DDHDD--	CSTCLCTIFGNEHAGKIRFEDLE-----	VVEEK-----
CbfCas7-11	---	NNLLA-K--NDHDD--	CTCSLCAIFGNEHETGKIRFEDLE-----	VVDKD-----
HvsCas7-11	---	KNVLT-----	CICVLCHLFGNVHETGRLKFEDLK-----	IVSGQ-----
HsmCas7-11	---	EGKLITTE--NEHED--	CTCILCRLFGNEHETGKVRFEDLE-----	LINDS-----
FmCas7-11	---	DALLI-----	KSHED--	CDCLCEAFGSKHHEGKILRFEDLT-----
DpbaCas7-11	---	KNLSS-----	NLHED--	CECLCKIFGSKHQEGNIRFEDMT-----
Domains	---	---	---	---
WmCas7x3	---	---	---	---
SER16298.1	DSAGSLIVHKSENSSWANDTFRDASVVDNFKARLKEIADPQDLSDALRADIKKRAFQLAT-----	PAT-----	LVT-----	LVT-----
GwCas7x3	---	---	---	---
PID64649.1	---	---	---	---
WP 031436019.1	---	---	---	---
HEB50754.1	---	---	---	---
HreCas7-11	---	---	---	---
CmaCas7-11	---	---	---	---
HvmCas7-11	---	---	---	---
SmCas7x3	---	---	---	---
oral meta	---	---	---	---
DisCas7-11	---	---	---	---
DsbaCas7-11	---	---	---	---
SstCas7-11	---	---	---	---
CsbCas7-11	---	---	---	---
CjcCas7-11	---	---	---	---
CbfCas7-11	---	---	---	---
HvsCas7-11	---	---	---	---
HsmCas7-11	---	---	---	---
FmCas7-11	---	---	---	---
DpbaCas7-11	---	---	---	---
Domains	---	---	---	---

FIG. 34F

134/142

	D758	D759	G-rich loop
WmCas7x3	HDMLAIDRFTGGGKDGAKI	LYYA	-----CANAAATPMTALGLITLVRLDLAEGDIPFGHCSAKGYGRCRAOGL
SER16298.1	HEMVAIDRFTGGGKDGAKENVDYI	-----	-----NEDALKP--ALGLSLLRLDLAEGDIAFGGANKGYGQCREHAV
GwCas7x3	HEMLAIDRFTGGGKDGAKENVDYV	-----	-----CGKDALLP--ALGLMTLMLDLAEGDIPFGYGSKGYGQCRASSA
PID64649.1	QEFVAIDRFTGGGKDEAKENAAHY	-----	-----WRK-----GLLALIFRDLAEGDIYFGYGNKGYGAVKKAIEI
WP 031436019.1	QDFVAIDRFTGGGKDGAKENAAHF	-----	-----WKG-----GLLALVLRDMOEGDITFGYGNKGYGGLGESASI
HE50754.1	QTFVAIDRFTGGGKDEGALYIRHA	-----	-----WGR-----GLLALVFRDLREGDITFGYGNKGYGVVDAAVV
HrcCas7-11	CDHVAIDRFTGGGKDEGALYIRHWK	-----	-----DTEK-----EALSKVLTSLKRGVYASLGGISAIGYGRVKELTL
CmaCas7-11	IDHVAIDRFTGGGKDEGALYIRHWK	-----	-----SALK-----EALSKVLTSLKRGVYASLGGISAIGYGRVKELTL
HvmCas7-11	MDHVAIDRFTGGGKDEGALYIRHWK	-----	-----DDID-----DESK-----SALKAFADFRDLGLVSLGGISAIGYGRVKELTL
SmCas7x3	LDHVAIDRFTGGGKDEGALYIRHWK	-----	-----TSGMT-----HOLT-----ELLQAOWRDLAAGHYPVGGKGYGWINSLVW
oral meta	IDHVSCSRFDASVNV	-----	-----ND-----VEIKTALODFADGLYSIGGKGYGWINSLVW
DisCas7-11	LDHVAIDRFTGGGKDEGALYIRHWK	-----	-----EYEC-----KALGKALANVNNGLYPLGGKSAIGYGVQVKSLSGI
DsbaCas7-11	FDHVAIDRFTGGGKDEGALYIRHWK	-----	-----RNEI-----EELGKAFADIRDLGLYPLGGKSAIGYGVQVKSLSGI
SstCas7-11	FDHVAIDRFTGGGKDEGALYIRHWK	-----	-----KELEKPEDER-----EALSKALADIRDLGLYPLGGKSAIGYGVQVKSLSGI
CsbCas7-11	IDHVAIDRFTGGGKDEGALYIRHWK	-----	-----KGFS-----LLITLALSIRDLGLYPLGGKSAIGYGVQVKSLSGI
CjcCas7-11	IDHVAIDRFTGGGKDEGALYIRHWK	-----	-----KDIV-----KDIKAMVDIRDLGLYPLGGKSAIGYGVQVKSLSGI
CbfCas7-11	IDHVAIDRFTGGGKDEGALYIRHWK	-----	-----RDVS-----PDEK-----ARILAFLEIRDLGLYPLGGKSAIGYGVQVKSLSGI
HvsCas7-11	FDHVAIDRFTGGGKDEGALYIRHWK	-----	-----DEAK-----ETLSQAFSDINTGIYLLGANGSIGYGVQVKSLSGI
HsmCas7-11	LDHVAIDRFTGGGKDEGALYIRHWK	-----	-----EELK-----KALSAFADIRDLGLYPLGGKSAIGYGVQVKSLSGI
FmCas7-11	MDHVAIDRFTGGGKDEGALYIRHWK	-----	-----VEMQ-----RALIAAFKEIRDLGLYPLGGKSAIGYGVQVKSLSGI
DpbaCas7-11	FDHVSIDRFTGGGKDEGALYIRHWK	-----	-----EKTQ-----EALSAFADIRDLGLYPLGGKSAIGYGVQVKSLSGI
Domains			
WmCas7x3	PDRWQALEAHFGPNADARALALRAWCRTHATAAL	-----DAPCSLAGSAPTAAAPS-----	-----QOAPAD-----AFHN-----PYHFI-----P-F-SQPDIDRWLSPPDA
SER16298.1	LDNWEERLKKIGAGLTIDGALQALRDVVALEPPAF	-----PPEIEKTTDDNQEPAPDF-----	-----NLKPASN-----GFHN-----PYHFI-----P-L-L-NNPKIGDWPEAKA
GwCas7x3	LGDWAEILLKQHLGADSDTTQALREVLGNPKQELKLD	-----PPSADATQAGVPAQONAAKTQAQAGE-----	-----KFHN-----PYHFI-----P-L-L-SKPDISQWPEPOK
PID64649.1	NPDCN-----ASDFSESDIEAFINKREKKGLYNC	-----NP1KKPGTKVSKNLPALVPLDRDTS-----	-----KFYN-----PYHFI-----P-V-KKPTNTSSWAEKTA
WP 031436019.1	TGIEQ-----ITSDIQAFDRKCVASQPTWLCDEAVKPAQOD	-----KIPPA-----GIOVANS-----	-----GFHN-----PYOFI-----P-S-KEPDTGHWLPVLG
HE50754.1	QDMAE-----LEPVFAFREQCROHQGMADCHSAPS	-----QPLRDHLAEIPPA-----	-----EAPGCE-----TFLN-----PYHFI-----P-I-REPTTGSWLARDE
HrcCas7-11	-----DDGPGW-----	-----LALPERKEDERPQPSFVS-----	-----LGP-VHVNLSKSGSDTADVYYPHYFL-E-PPSQTV-----
CmaCas7-11	-----EEN-----	-----DIIQLTEITESNLNSQSRLS-----	-----LKP-DVKKELSN-----HFYI-----PHYFI-----K-PAPKEV-----
HvmCas7-11	IGGSADW-----	-----LNLKPPEENRTDVPDGRSA-----	-----QGP-EIKISLDAD-----KIYH-----PHFFL-K-PSDKNV-----
SmCas7x3	DGEK-----	-----ITCRPDGDSISLTITVTGDIP-----	-----PRP-ALTPPAG-----AIYV-----PHYFLPNNP-EHKP-----
oral meta	PRSLRKLNSGFREA-----	-----SSIQDALLDSAKEIPLSAPLT-----	-----FTV-VKG-----AVYN-----PYIYL-PFP-AEKP-----
DisCas7-11	KGDKRI-----	-----SRIMNPAFDETDVAVPEKPK-----	-----TDA-EVRIEAE-----KVIY-----PHYFL-E-P-HKKV-----
DsbaCas7-11	VDDSDSDDENPAK-----	-----LLAESMKNASPSLGTPTSLKK-----	-----KDA-GLSLRFDEN-----ADYI-----PHYFL-E-P-EKSV-----
SstCas7-11	KGAGD-----	-----VTKAALQSESSMSASEPEH-----	-----KKP-DGSLKLSFDDKKAVYI-----PHYFL-K-PAAEEV-----
CsbCas7-11	DDNVDPDDFKEMINK-----	-----TEMLPPEEVESNNGPINNDY-----	-----VHP-GHOSPKQDHKNKIYV-----PHYFL-D-S-GSKV-----
CjcCas7-11	LNQSGFQ-----	-----IPVKKDISPEPGTYSTPSH-----	-----STP-SLNGK-----HIYI-----PHYFL-A-P-ANTV-----
CbfCas7-11	DGDAPEAFKEMNSK-----	-----RGQASFEKISFRYPSGAPK-----	-----HIQ-NLKAT-----SFYI-----PHYFL-E-P-GSKV-----
HvsCas7-11	-----LKAPSWLKIK-----	-----KPDNFEKDTSLNISAIMNEF-----	-----KKD-IQTINLD-----KTYL-----PYGFL-K-L-LEKV-----
HsmCas7-11	-----ESGPEWLRLEVEP-----	-----KSSQENVLSPVLSVMDEIF-----	-----NPP-KIDEN-----GVYI-----PYAFL-R-P-LNEV-----
FmCas7-11	-----TNGPDWLNLEVEP-----	-----LPOSTADIPVECTAEPYKPF-----	-----QKP-DLDQN-----AVYI-----PHYFL-Q-P-GKPA-----
DpbaCas7-11	-----VEGPDWLK-----	-----DALSAEKTVAGIASESEY-----	-----KLP-DLPWISLLPKGRAIYN-----PHYFL-GIP-KVTP-----
Domains			

FIG. 34G

135/142

WmCas7x3	H---	RKTGGH-SRYRG	---	LSGRIVCALTTVTPLFVGAAA	RTP-ASDQHPK	---	P---	VAGFALQNPAPATSLRGLSS
SER16298.1	ETLKANREGH-DQYHTGK	---	FSGRIVCSLTQTPLFIGAET	KPS-TSDREPS	---	E---	ARDFKLNKGHAIPATSLRGLSS	
GwCas7x3	L--TEKGHSH-DRYAS	---	LSGRIVCSLTQTPLFIGSEQ	TTP-TNPQAPK	---	S---	LHFFKLNGLAIPATSLRGMIS	
PID64649.1	FGTADSPHSH-GFYRKQTN	---	QOPLYSGRLICMLTSETPFFFIGQA	ESD--PTENE	---	NQASLRHPYQLDGEPAIPATSLRGLISTM		
WP 031436019.1	L-NADSHSH-AFYRDQTDN	---	GEKLYHGRILCCLNTETPIFIGADK	KKD-TEPAEIN	---	N---	YRLNGELAIPATSLRGMIS	
HEB50754.1	L-DSSCCHSH-GFYRQVDD	---	RPLYHGRITCCLTETPFIGATG	DSS-V-PSRIE	---	N---	YRLGNRIAPAAASLRGLSS	
HreCas7-11	S-RELDIISHARTDS	---	GEALLTGRILCRLITRGPFIPTDN	NDNAFLEGGI	G--HKN	---	YRFRINDELAIPGSELRGWSSV	
CmaCas7-11	V-RESPLISHVQCHDT	---	GEFLTGTIKICRLOTLGPLFIANDKDDYFELQHNPG	---	HLN	---	YAFRINDHAIIPGASIRGMISV	
HvmCas7-11	Y-RERELVSHAKKGP	---	GKSLFTGKITCRISTEGPFIPTD	DLGEDYFEMQASHKK	---	HKN	YGFFRINGNVAIPGSSIRGMISV	
SmCas7x3	K-RSDKIIGH-HTFAT	---	PDSTFTGRITCKLEVTPPLIVPTE	G---EQPKDQ	---	HKN	YFPEKINDEIMLPAGLAAVAASQV	
oral meta	E-RCIVPPSH-ARLQS	---	D---RYTGCLTCELETSPVLLPDTG	---	YKE	---	YPSFRLLNTPMIPGAGLRAAASQV	
DisCas7-11	E-REKPCGH-QKFHE	---	G---RLTGKIRCKLITKTPPLIVPTS	NDFFFRPADKEARKEKDEYHKS	---	YAFRLHKOIMIPGSELRGWSSV		
DsbaCas7-11	H-RDPVPPGHEAFRG	---	G---LLTGRTICRLTVTRPLIVPTE	TDDAFNNKEKAG-KKDAYHKS	---	YRFTLNRPVMPGSEIRGMISV		
SstCas7-11	N-RKPIPTGH-ETLNS	---	G---LLTGKIRCKLITKTPPLIVPTE	NDFF--QTGVEG	---	HES	YAFSVNGDIMLPGSELRGWSSV	
CsbCas7-11	Y-REKDIITH-EEFTE	---	E---LLSGKINCKLETLPLIIPDTS	DENGLKQGNKPG	---	HKN	YKFFNINCEIMLPGSELRGMLRTH	
CjcCas7-11	H-REQMIGH-EQFHK	---	E---QKCELLVSGKIVCTLKTPPLIIPDTE	NEDAFGLQNTYS	---	HKN	YQFFHINDEIMVPGSEIRGMISV	
CbfCas7-11	I-REQMIGH-EQFYE	---	E---SYPSGASCEKLLSGRIICSMTHVTPLIIPDTE	---	HAT	---	YDFQMNNAIMIPGSEIRGMISV	
HvsCas7-11	K-RTSSPITH-ERFYE	---	N---HLTGFTIECSLKVLSPLIIPDTE	---	TPEKEENG	---	YHFKIDNKPPIIPGAEIRGA	
HsmCas7-11	K-RTREPIGH-NEWKK	---	S---LISGYLTCRLLETLPLIIPDTS	---	VIKDPENK	---	YHFKIDNKPPIIPGAEIRGA	
FmCas7-11	E-REHVPVSH-DHIDD	---	K---LLTGRLVCTLTTPPLIIPDTS	---	EEVKEKVNNGE	---	HPV	YKFFRLGGHLCIPAAEIRGMISV
DpbaCas7-11	E-REHVPVGH-DRFQT	---	D---LHTGRLICTLTITPLIIPDTE	NDKAFEVENASAD	---	HER	---	YKFFRMDGQAAIPGSAIRGMTSSV
Domains								
E599								
WmCas7x3	FESISGSNLRVLHPT-PYSIRKTTKE	---	---	ALSA-I	---	G	---	
SER16298.1	FESVSNSEFRVLHPE-HYSVRKSLDDIV	---	---	ALSA-M	---	G	---	
GwCas7x3	FESVSNSEFRVLDEK-TYSMRKTMQO	---	---	SLSA-M	---	G	---	
PID64649.1	TEAAANCAMRVLDS-IISYRKPMNP	---	---	ILSA-L	---	G	---	
WP 031436019.1	AEASNSAMRVLDS-LLSYRKATAD	---	---	ALRK-V	---	G	---	
HEB50754.1	AEASNSAMRVLHG-ILSYRKAKN	---	---	ALRE-I	---	G	---	
HreCas7-11	YEALTNSECFRIMEEGRYLSRMADEK	---	---	DFHP	---	G	---	
CmaCas7-11	FETLTHSCFRVMDKKYLTRVPISE	---	---	TQKRSGRYQVEESDPDLP	---	G	---	
HvmCas7-11	FEALTNSECFRVDQERYLSRSEKPDTE	---	---	LTQYYP	---	G	---	
SmCas7x3	YEALTNSECFRVMKQKRLSWRMEADYK	---	---	DFYP	---	G	---	
oral meta	YEALTNSECFRIMDQOTLWRMSTSEHK	---	---	DYOP	---	G	---	
DisCas7-11	YEALTNSECFRIEETKRLSWRMDAQN	---	---	VLQDFLP	---	G	---	
DsbaCas7-11	FEALTNSECFRIEKKYRLSWRMDADYKE	---	---	LEQFKP	---	G	---	
SstCas7-11	YEALTNSECFRVDGEYRLSWRMADEK	---	---	LMQFKP	---	G	---	
CsbCas7-11	FEALTKSCFAIFGEDSTLSWRMADEK	---	---	YKIDSNSIRKMSQRPKYRI	---	DELQKELRNSGNLFNRLYTSE	---	RFRFWSVSNKFNENSIDYKREILR
CjcCas7-11	YEALTNSECFRVDETKYITRLSPEKDK	---	---	ESDNKKSQDDASQKIRK	---	G	---	
CbfCas7-11	YEAMTNSECFRIEHEKQYLTFRISPEDKE	---	---	REFIP	---	G	---	
HvsCas7-11	YEALTNSECFRVEKGVLSWRMEKDAK	---	---	EFMP	---	G	---	
HsmCas7-11	YEALTNSECFRVEKRLISWRMTAEAK	---	---	RPDPKKSEQNRMRFP	---	G	---	
FmCas7-11	FEALTGSCFRVINQKHLSWRINADMAK	---	---	HYRP	---	G	---	
DpbaCas7-11	FEALTNSECFRVLQKSHLSWRMEADDAK	---	---	DYKP	---	G	---	
Domains								

FIG. 34H

WmCas7x3	D998	---	RIVERNGELKYLPL	---	TLPLTHONA-DNAY	---	PVPA	---	RWRKVF-YWESPV	---
SER16298.1		---	RIVDDQGEKLOPL	---	TLPLFGR	---	NVPA	---	KWEKIFGTPSEDD	---
GwCas7x3		---	RIVRHQKLYLLPL	---	TLPLPQGP-HGVY	---	DIGE	---	KWSAVF-DWOP-P	---
PID64649.1		---	MVTKRGEDFWLPL	---	AMPALINDEEHY	---	KLDK	---	RYRTMF-PDGLA	---
WP 031436019.1		---	MVIYVDNKSTIKL	---		---		---	NDAI	---
HEB50754.1		---	MIVLRDGRFLPL	---	VPLM	---		---	EVT	---
HreCas7-11		---	IV--VDGAKIREM	---	KRVRLPLYD-T-P	---	DKTS	---	RTKEMT-CPELFRKDG	---
CmaCas7-11		---	RVQKKGKVKIEKM	---	DEIVRLPLYD-N-F	---	SLVE	---	RIREYH-YSECASYVP	---
HvmCas7-11		---	KVRDGNKFKLKM	---	KDFFRLPLYD-F	---	DFEG	---	EASLR-PNYEDRNEENKGN	---
SmCas7x3		---	RV--LDGGQIKKMG	---	DKAIRMPLYD-D-S	---	TATG	---	SIKDDQ-LISDCCPKSDE	---
oral meta		---	KI--TDNGRKIQPMG	---	KQAIRLPLYD-E	---	VIHVSTPGD	---	TDLEKLAIVLELTRPWKELP	---
DisCas7-11		---	RV--TQGGHIQKF	---	SETARVPFYD-K	---	TQKHFDI	---	EQEAGE	---
DsbaCas7-11		---	RV--ADDGKRIEM	---	KEIRYPFYD-R-T	---	YPER	---	KPVFMW-VKRFIKRLSLVDPAKH	---
SstCas7-11		---	RV--TDNGLRIEM	---	KEIRYPFYD-R-D	---	CSDK	---	NAQNGY-F-RWDARISLTDNSMR	---
CsbCas7-11		---	CAGRPNYKGGIIRQKDSLMAEEL	---	KVHRLPLYD-N-F	---	DIPD	---	KSQRAY-FDEWERSITLDDSL	---
CjcCas7-11		---	IVKTDGFSIIEVERYSMTKGT	---	KLVVKVRLPLYD-S-E	---	AVIA	---	SAYK-ANDHCRKSATCSTSRG	---
CbfCas7-11		---	IVRIINGDVYIEKA	---	EREYRLPLYD-D-V	---	HIT	---	SIQ-FEQYCEKNEKNA	---
HvsCas7-11		---	RVSKKKGKLYMVKM	---	QALRLPLYD-N-P	---	ALAN	---	NYEELE-YEKYIKKNPGRFQ	---
HsmCas7-11		---	RIKKDKKFFAQEM	---	LELRIPVD-N-K	---	D	---	EIRSGS-IYE	---
FmCas7-11		---	RIQNNKMFQPY	---	KMFLPFYA-G-F	---		---	KRNEIS-QNDPTRPS	---
DpbaCas7-11		---	REFKDDKAVIRKF	---	KKKARPFYA-G-P	---		---		---
Domains		---		---		---		---		---
WmCas7x3		---		---		---		---		---
SER16298.1		---		---		---		---		---
GwCas7x3		---		---		---		---		---
PID64649.1		---		---		---		---		---
WP 031436019.1		---		---		---		---		---
HEB50754.1		---		---		---		---		---
HreCas7-11		---		---		---		---		---
CmaCas7-11		---		---		---		---		---
HvmCas7-11		---		---		---		---		---
SmCas7x3		---		---		---		---		---
oral meta		---		---		---		---		---
DisCas7-11		---		---		---		---		---
DsbaCas7-11		---		---		---		---		---
SstCas7-11		---		---		---		---		---
CsbCas7-11		---		---		---		---		---
CjcCas7-11		---		---		---		---		---
CbfCas7-11		---		---		---		---		---
HvsCas7-11		---		---		---		---		---
HsmCas7-11		---		---		---		---		---
FmCas7-11		---		---		---		---		---
DpbaCas7-11		---		---		---		---		---
Domains		---		---		---		---		---

FIG. 34I

137/142

WmCas7x3
SER16298.1
GwCas7x3
PID64649.1
WP 031436019.1
HEB50754.1
HreCas7-11
CmaCas7-11
HvmCas7-11
SmCas7x3
oral_meta
DisCas7-11
DsbCas7-11
SstCas7-11
CsbCas7-11
CjcCas7-11
HvsCas7-11
HsmCas7-11
FmCas7-11
DpbaCas7-11
Domains

PLRVY---FGS-R---KQTYDS-R---QP---HYLP-IQELS-Y---LPND---SDCI---
---FDDIP---SKFSSN-K---RY---FYCKATELKDF---FYN---IKSD---KYFIG---
---RRTYQS-Q---QP---CYMK-LSTVK-Y---SESN---PNQI---
---KLVYLEKAYSNV---MKTFLN-N---EN---TWTLAQSKIH-Y---LPLT---PIQMO---
---KLQOT---YTPGN---MKDFIE-K---SN---SWSPEHTVY-Y---LD---
---KLRHA---YTDPA---MKHFD-D---KN---SWSPRCNRVY-Y---LGRD---
---RPERA---KKFNEEIAKAVONRAYLL-SLDEKERR---EVLLGNREVT-FDCPPDEYSDEYSELKYAQKYDFIAVL
---SVKKA---IDYRMALQAADSREFLY-NHPEAK---SILQKKEVY-Y-ILHKOESKNRGTKKEINPNARYACLTDE
---KVKA---VEFNIRKMGAFKHNQDFLK-KYKEQEI-K---DIFMGKKVY-F---TAGCHKPNEARDNDRIALLTKG
---KLOKA---LATNQKIALAAKHNOEYLA-QLSPDERE---EALQGLKKVY-F---WTESLANNEAPPFLIAKIGEE
FANWHGRELLEFDRFLKDENSCLIKA---ETDTRVIAASVAKNNRDYLF-EIKQDFARYKRIIOGLERP-F---SLRSLAKSETSFOIACLGRL
---ATFIE---QKNGSYFNVVTTNGCTSF-HLWHKPDNFDQEKLEIGQNGE-K---LSCWVRDSRYOKAFOEIPENDP---
---GRYKLVKLKHAERVGGEIPIYSHP---TPTDCKLLSLVGNPQLCR-QDTLVQYRIKHRRGAKPEED-F---MFVGTSPENQK---GHKEN---
GGNITGLIFKYAE-RIGDVPYRHEP---TDTDRMMLSLSEYRNQ-K-SDGKRAYKIKPASKGLGAY-F---MFACTSVENKRICNPACTDKA---
---CGIKV---RDKNRVFLNAANNRQYLN-NIKKSNHDLVLYQLKGEKKIR-F---NSKVITGSESPIDVIAEL---
---KIRAA---IKRNEVIAEVARNNLIIFLR-SLTPPEELK---KVLQGEILVK-F---SLKSGKNPDYLAELH---
---KIKNA---HRFNKNIAEASRNLYLC-SLDRAVRR---EILSGRKNV-F---RLVKVNDKNPDKEAVELC---
---KYKN---SKVEIIFQTVGEGRKFLRGNFNNVEWKK-VLVTGIDPLA-I---LPSQKIPGNDKWKVKNLQ---
---EYNHP---TEPERIFFSNAEKIRNFKL-RNS---NYLHGSTPLL-F---ROWSISNRDYKIA---
---DPNCLSE---KOLLGIEPVK-L---WVKDFVASLVKQPOTDIDIE---
---DTREAFTS---DQIMGEKVT-L---WVKDFEASLTVPDE---

WmCas7x3
SER16298.1
GwCas7x3
PID64649.1
WP 031436019.1
HEB50754.1
HreCas7-11
CmaCas7-11
HvmCas7-11
SmCas7x3
oral_meta
DisCas7-11
DsbCas7-11
SstCas7-11
CsbCas7-11
CjcCas7-11
HvsCas7-11
HsmCas7-11
FmCas7-11
DpbaCas7-11
Domains

---APDQGLRFRSRRDRKFLIGQCPISRYDCPIPETDLP---KLSQERPRY---TR-GWVR-SLWTS---NREK---ELPHTVK-H
---KRTPTV---FKSSSTEKSHLESLEFID-VEKFKAVEN---LEITPGNNPY---IH-GWVR-NLKDE---FRE---DIPDNVK-H
---IAGENLALRFRGNQNTQFLIGQSNQD-ECPTQAEYA---QKSEDERNEY---TS-GWVR-TLVKP---GR---DLPRSVK-H
---NGGINSYNNLRTSPRSNNFLIGQTVAH---GNGIPASGPG---AGMVP---GILR-ILGKE---HRQN---DLPONKH-H
---NNQIPQESYM---NGMKP---GILR-ILGKE---GREQ---EL-ENKL-H
---GNOIPAETRG---AGMRP---GILR-LLGRE---GRHD---AL-QNKH-H
---KKNQKRGYIKFTGPNTAN---KKN---EDAPDKNY---RS-DWDP-FKINI---LLESDPCRVSN---IH
---NTPGRKGFIKFTGPDMT---VNK---ELSKIAPIYDPEWEKIDPWERSNQESNHKYSFILHNEIEMRSSQKK
---SNKKAEGYFKFTGPGMVN---VKA---GVEGEEDFHIDESDPDV-YWNMS---SILPHNQIKWRPSQKK
---RGKPRAGYLKITGPNAN---IAN---TNPPDDGGY---IP-SWKD-QFDYS---FRLGPPRCLPNTKGNR
---RGRLRKGYLKITSGPNKLE---VEI---SGSHSNGY---SD-IWDD-PLDFS---FRLSGKSELRPNTQKTR
---DGWECKEGLHVVGPSKVE---FSDKKGDVINFGDTLPSV---PN---IRTND---FKNRKRNE---
---NDHGGGYLKITSGPNKIE---KEN---VLSGPVPSV---PE-NMGA-VVHNCPP---RLVEVTVRGKRQKEEC
---NKSVMGYLKITSGPNKLE---KYN---ISEPELDGV---PE-DRNCQIHNRI---YLRKIFVANAKKR---K
---NERGQTGFIKLSGLNNSN---KSQ---GNTGTTF---NS-GWDR-FELNI---LLDDLETRPSKSD---
---ENGTERGLIKFTGLNMVN---IKN---V-NEEDKDF---ND-TWDM-EKINI---FHNAHEKRNLSLQ---
---KTGPLEGLVFKFSGLNAV---ISN---LRPTAEEGF---DA-KWDM-WSINI---ILNRMDVNSQKK
---SKISPVRYGFKFTGPNKIE---TKRREEKDEKLTAKNKVSLQ-KD-KWYE-AMHNHVE---YKQDYTPNPSKTEP-L
---LIGNKSQHLKFTGPNKIE---VSE---GTCPKYETIPGRD-EWDK-AVHNYVEPGKFTVISRKKGQKPKA-V
---WKEKIGFVRVTGPNKVE---VDS---SNTDPSLPEC---ES-DWKD-IHITE---DGSTPSKNDR---
---IGMKRGRGYLKVTPGNKVE---IDT---ENISENNPSP---PD-SWQD-VRIND---DGTIPDKNR---

FIG. 34J

138/142

WmCas7x3	QLFI--PDPVE--TPAA--DDL--	LPIP-QGV-LDTHALADLALAGQH--	W-GKDE-TPADD-QLLPFTFAGRQ	
SER16298.1	HVFL--PDTTKRVSP--	LEIP-PSV-KKREHDLADLALAGH--	L-KQGE-TIASPKILPYTPIGRNK	
GwCas7x3	HVFL--PDVFIDAPPV--NDL--	YPIP-DSV-IQRFHDLADQVLSMN--	L-KPEE-IVDST-NLLPYTPVGRRS	
PID64649.1	ELFI--PVPDAFADPK--TFLD--	TATAFLIP-RNV-IDAFEKIAEQTSQKQK-L-KHDE--	ERLPFHHLKGTGR	
WP 031436019.1	ELFI--PVPLEVVDTEN--NKFDYOAYK--	FLYRAIEIP-EPV-LKRYSELADQRTMSOKNKEL-KDD-TCQSV-GWLPHHLKGTGR		
HEB50754.1	EYFI--EIPERYVD--QD--HCFDYRMFTR--	RARNGTLVPSVAMERYHCLAERTLSQNDPEL-REK-ACASL-KWLPHHLKGTGR		
HreCas7-11	CYPR--PLLVC--IK--DKAEYRIHRCEAIFCSI--	GPSDLIDIP-QKV-SNQYRTIL--QDYNNTGK-IVE-IF--	RTQ	
CmaCas7-11	KYPR--PVFIC--KK--NGVEYRMQRCERIFDFT--	KEEEKDKEIVIP-QKV-VSQYNAIL--KDNKENTET-IPG-LF--	NSK	
HvmCas7-11	EYPR--PVLKC--VK--DGTEYVLMKRSEHVFAEA--	SSSEDSYPVP-GKV-RQGFNSIS--RDNVQNTDH-LSS-MF--	QSR	
SmCas7x3	EYPR--PGFTC--VI--DGKEYSLTKRCERIFEDI--	SGGENOVRAVT-ERV-REQYREIL--ASYRANAAG-IAE-GF--	RTR	
oral meta	EYPR--PSFTC--TV--DGQYTVNKRCEVFEDS--	AAPAEILP-RMV-REGYKGIIL--TDYEQNAKH-IPQ-GF--	QTR	
DisCas7-11	DisCas7-11	ED-DKGNYITMAKYCETFFDL--	KENEYEIP-EKA-RIKYKELL--RVYNNPQA-VPESVF--	QSR
DsbaCas7-11	KRKRLVPEYVC--ADPEKVTYTMKRCERIFLEK--	SRRIIPT-NDA-VDKFEILV--KEYRNAEQDTE-AF--	QTI	
SstCas7-11	ERDLVGEFAC--YDPEKVTYTMKRCERIFIKD--	RGRTLPIT-HEA-SELFELV--QYRENAKQDTE-VF--	QTL	
CsbCas7-11	YPR--PRLLE--TK--DQYEVNITKRCERVFED--	KGNKTGYPD-DQI-KKNVEDIL--DSYDGKIDQ-EVAE-RF--	DTE	
CjcCas7-11	GYP--PVLKF--IK--DRVEYTIKRCERIFCIP--	VKNITIEKVS-SKV-CKQYKDV--SDYKFNHGH-INK-IF--	TTK	
HvsCas7-11	EYPR--PALHE--NH--DGKEYTIKRCERVFVRAEAGKRAETGSKYVP--RKV-QEQYONIL--	RDYESNIGH-IDN-TF--	RTL	
HsmCas7-11	ERPRNIPCFVC--SD--KEKIYRMKRCERVFVSL--	GENAPKYEIP-ISA-IKRYEVL--SAYRENWERNKTE-LF--	RTR	
HsmCas7-11	ORRRNVPAFCC--YDYNTRCFVNMKRCERVFVKS--	RDKPKYEIP-PDA-IRRYEHVL--RKYRENWERYDIPE-VF--	RTR	
FmCas7-11	YVRC--QL--KGVTYTVAKWCEAFWKD--	EGKKPIVTV-AEA-INRYHLLM--KSYQDNQPS-PPI-IF--	RSL	
DpbaCas7-11	Domains	KFIC--QY-GTTTYTVAKWCEAFFCDE--	EKDPYELA-PDV-ERKRYLLM--DSYHNNPQA-PPQ-IF--	RSL
WmCas7x3	HDA--DRPRDADROTRLOPGDLVCFD--LG--	D--	DGAVSEISFSSIWREGILAG-KPN-LATTADLLAQ--	V
SER16298.1	LENHHRVENDLTCYMRLLKGDVFFD--VD--	N--	DGQITEISFSSIWRAGI--GTNNK-LQTTADLLSQ--	R
GwCas7x3	DS--CRDTRLOAGDLIFFD--IDPPLHPG-E--	N--	KSQTEISFSSIWRSGI--G-KDH-LLTTPDLLTN--	F
PID64649.1	EQN--HTIQIKTGDVLYFR--PN--A--K--	A--K--	GDEVEEIAFSSIWRGKT--SGTTADFFPD--	K
WP 031436019.1	QID--DKHNVGKIDIDEYDLIYE--AS--	A--	GKEVTEVAFSSIWRGRV--ETNSS-QANKVYSFIP--	G
HEB50754.1	ERD--PENDVCHLSLRHGDVLYFA--EQ--	E--	NRVSEISFSAIWRSV--ETSDSYQAVTVDCFVP--	K
HreCas7-11	IKH--DQ--LTTGDVLYFK--PA--A--	A--	NGQVNAVIVPVSISRKTD--ENPLAKRFKN--	D
CmaCas7-11	MVN--KE--LEDGDVLYFK--YK--	E--	EGKVTETPVAISRKTD--NKPMDKRFKPSINGMKMKPD--	D
HvmCas7-11	RLH--DE--LSHGDVLYFR--HD--	E--	KRKVTDIADVRSRTVD--DRPMDKRFKN--	E
SmCas7x3	MYD--TE--ELRENDVLYFKTAKQ--ADG--	ADG--	KERVVAISPVCISREAD--DRPLGKRLP--	A
oral meta	FSS--YR--ELNDGDVLYFK--TD--S--	S--	QGRVTDLAPVCLSLRAD--DRPLGKRLP--	E
DisCas7-11	VARE--NVEKLKSGVLYFK--HN--	H--	EKIVEDIVPVRISRTVD--DRMIGKRRMS--	A
DsbaCas7-11	LPE--NG--TVNPGDLVLYFR--EE--	E--	KGAAEIVPVRISRKVD--DRHICKRID--	P
SstCas7-11	LPD--NG--RLNPGDLVLYFR--EE--	E--	KGKTEIIPVRSRKID--DSPIGKRLR--	E
CsbCas7-11	TRG--SK--LKVGDLVLYFH--ID--G--	G--	DNKIDSLIPVRSRKA--SKTLGGKLD--	K
CjcCas7-11	IQK--RE--LTDGDVLYFI--PN--EGA--	EGA--	DKTVQAIMPVPLSRITD--SRTLGERLPH--	K
CbfCas7-11	IEN--CG--LNNGDLVLYFK--PD--N--S--	N--S--	RKEVVAITPVKSIRKTD--RLPQGRFPFHTS--	S
HvsCas7-11	LPG--DGTLENDVLYFR--AD--E--	E--	NEKVDIIPVCISRIVD--EVLPLKRLS--	Q
HsmCas7-11	LPD--DGTLENDVLYFR--LD--E--	E--	NNRVLDIIPVRSIRSD--TQYLGRRLP--	D
FmCas7-11	PVLNYKQ--DOKII--GSMIFYR--ES--AKS--	AKS--	DKIVNEIIPVKSIRTD--TELLAKHLPN--	N
DpbaCas7-11	PLFSETG--PKTLEHGDVLYFR--SEV--	SEV--	NKQSQKKQVRERTDIVPVSIRIAN--NOPIKGHIA--	A
Domains				

FIG. 34K

WmCas7x3	SPHLLPLG-----MPGRS-A-----RLSPVEQLFGVVEYRPPQAK--GTRKPTDAPAAVALAGKLQVGFGRP---ARPF	ER-----EP-AVTLKELSTPKPPSPALYFPKA--GDGYVSKAKLASQ--PQ-----DYAPAGRKHLYH--ALRR
SER16298.1	DPNLVLQIG-----MGVRT-KNTDRFKLSAERLFGVVEHRRDDNTTVENVQNDKAQAFAGKVRIGFGLP---DKKT	TV-----NGVS-PVTLKELSSPKPPSPALYLRKN--NDDFVSKKVAEC--SE-----TMTLGRKCYLH--AWRE
GwCas7x3	DVNLPQHG-----MPGRT-Q-----SLSPAEILFLGLVGTQNDQAIT--AYAGKVRIGFGLP---EEGH	NP-----RUDA-RITLKELSSPKPPSPALYFRKKS-GDEYVSKANLADK--PE-----DYILGRKMYLH--AWRK
PID64649.1	E--LLPFN-----RNS-----RVSPAELLFGFTENNPKEMKI--D--RGLAFAGKIRISAGTL--SDKF	SDTTESDLFEP-ETTLKALSSPKPPSPALYFKEKKS-GTYVKKODL-N--PG-----KHEIOGRKYLH--ALRN
WP 031436019.1	E--LLPFN-----ESRK-----KSPAELLFGFTQINKDGSKA--DD--KAQAFAGKVRISAGTI--SEYD	EN--EANLEQ-EVTLKALSTPKLPSPALYFTINGNSAYISKQEL-E--PS-----KHLAKGRKYLH--ALRT
HEB50754.1	E--LRPFN-----RDRR-----AISPAELLFGFVDELSEHST--EKSRYEQMAFAGKVRISAGLP--VEDV	ED--SALLEPKPIVILKALSSPKPPSPFYFMRDG-SCAYIAKKDL-S--PD-----RHRIGRKHLYH--GLRQ
HreCas7-11	S--LRPCA-GLCVE--DNECPARCKK-VADYENPH-R--GLCPACHLFG--TT--FYKGRVRFAGLWLTGEDGAP	RWYKGPD---PC-DSG-KGR-PMTIPLERPRPT--WSIPD--N--SFDIPGRKRYVHHP-YSV
CmaCas7-11	S--LRSCS-HTCTE--GDDCPNLCE-SKDYFKPH-D--GLCPACHLFG--TT--FYKSRVRFAGLWLT--ENNA	KWYISNDFQKDS--KKE-KGG-KLTIPLERPRPT--WSMPN--K--KGSNTEDATIPGRKRYLHHDGWRI
HvmCas7-11	S--LRPCN-HVCVE--GDDECPDRCKE-LEDYFSPH-E--GLCPACHLFG--TT--DYKGRVSFGLGWLH--ESNTP	DD--KQP-ETK-TLTIPLERPRPT--WVLPK--SV--KSEIPGRKRYVHHP-WSV
SmbCas7x3	G--FQPCS-HVCLE--DNTCSAKNCP-VPLYREGWPVN--GLCPACRLFG--AQ--MYKGRVRFSGFAL--P	ST--RSP-QLK-YVTIPLSQRPRPT--WVLPK--S--CKGKEDVPGRKRYLHHDGWRE
oral meta	E--YRPCA-HVCLE--ECDPCTGKDCP-VPIYREGYPAR--GFCPACQLFG--TQ--MYKGRVRFSGFAL--VN--	KWLKED---PA-EPS-QGK-GITLPLERPRPT--WVLPK--S--KFKVPGRKRYVHHPHAWKT
DisCas7-11	E--LRPCH-GWVE--DGLSALNAYP-EKRLLRHP-K--GLCPACRLFG--TG--SYKSRVRFAGLWLT--ENNA	KWLKED---SD-GPS-HGG-PLTIPLERPRPT--WVLPK--ND--KSEIPGRKRYVHHPHAWKT
DsbaCas7-11	E--LRPCH-GWIE--DGLSKLDAYPAEKLLTRHP-K--GLCPACRVFG--TQ--SYKSRVRFAGLWLT--ENNA	EWLI-----TR-GNN-PER-SLTIGVLESPPRA--FSIPD--D--ESEIPGRKRYLHHDGWRI
SstCas7-11	D--LRPCH-GWIE--GDDLSQLSEYP-EKKLSTRNT-E--GLCPACRLFG--TG--AYKGRVRFAGLWLT--ENNDP	KWLT-----ER-ENG-CGG-YVTIPLERPRPT--WSVPS--D--KCDVPGRKRYLHHDGWRI
CsbCas7-11	A--LRPCT-GLS-----D--GLCPACHLFG--TT--FYKGRVRFAGLWLT--ENNDP	KWFR-----KD-EGN-ETC-HITLPLERPRPT--WSMPD--D--TSTIPGRKRYVHHPHAWKT
CjcCas7-11	N--LRPCV-HEVN--EGLSGILDSL-DKLLSIHP-E--GLCPACRLFG--TQ--FYKGRVRFAGLWLT--ENNDP	NW-----KF-REE-QLTIPLERPRPT--WVLPK--R--EVQIPGRKRYLHHDGWRI
HvsCas7-11	E--LRPCV-HEVN--EGDIRMLENSP-FKRLFIHP-E--GLCPACRLFG--TQ--FYKSRVRFAGLWLT--SDGP	KFCQ-----GG-ERP-EDR-AVTIPLERPRPT--WVMPN--E--NSTIPGRKRYLHHDGWRI
HsmCas7-11	H--LRSCV-RECLYEGWGDCKPKLSFP-EKMWIRINP-E--GLCPACHLFG--TQ--FYKGRVRFAGLWLT--SDGP	VW-----AK-GAG-GKD-YFTIPLERPRPT--WTMPN--E--GAKIPGRKRYVHHPHAWKT
FmCas7-11	D--FLPCA-ATCIN--EDCTNAKCTKFLPLYREGYPVN--GLCPACHLFG--TQ--FYKGRVRFAGLWLT--SDGP	
DpbaCas7-11	A--FRPCA-YVCIE--ECEPCDAKTCF-IPVREGYPIK--GLCPACHLFG--TT--FYKGRVRFAGLWLT--SDGP	
Domains		
WmCas7x3	EP-AVTLKELSTPKPPSPALYFPKA--GDGYVSKAKLASQ--PQ-----DYAPAGRKHLYH--ALRR	ER-----EP-AVTLKELSTPKPPSPALYFPKA--GDGYVSKAKLASQ--PQ-----DYAPAGRKHLYH--ALRR
SER16298.1	TV-----NGVS-PVTLKELSSPKPPSPALYLRKN--NDDFVSKKVAEC--SE-----TMTLGRKCYLH--AWRE	TV-----NGVS-PVTLKELSSPKPPSPALYLRKN--NDDFVSKKVAEC--SE-----TMTLGRKCYLH--AWRE
GwCas7x3	NP-----RUDA-RITLKELSSPKPPSPALYFRKKS-GDEYVSKANLADK--PE-----DYILGRKMYLH--AWRK	NP-----RUDA-RITLKELSSPKPPSPALYFRKKS-GDEYVSKANLADK--PE-----DYILGRKMYLH--AWRK
PID64649.1	SDTTESDLFEP-ETTLKALSSPKPPSPALYFKEKKS-GTYVKKODL-N--PG-----KHEIOGRKYLH--ALRN	SDTTESDLFEP-ETTLKALSSPKPPSPALYFKEKKS-GTYVKKODL-N--PG-----KHEIOGRKYLH--ALRN
WP 031436019.1	EN--EANLEQ-EVTLKALSTPKLPSPALYFTINGNSAYISKQEL-E--PS-----KHLAKGRKYLH--ALRT	EN--EANLEQ-EVTLKALSTPKLPSPALYFTINGNSAYISKQEL-E--PS-----KHLAKGRKYLH--ALRT
HEB50754.1	ED--SALLEPKPIVILKALSSPKPPSPFYFMRDG-SCAYIAKKDL-S--PD-----RHRIGRKHLYH--GLRQ	ED--SALLEPKPIVILKALSSPKPPSPFYFMRDG-SCAYIAKKDL-S--PD-----RHRIGRKHLYH--GLRQ
HreCas7-11	RWYKGPD---PC-DSG-KGR-PMTIPLERPRPT--WSIPD--N--SFDIPGRKRYVHHP-YSV	RWYKGPD---PC-DSG-KGR-PMTIPLERPRPT--WSIPD--N--SFDIPGRKRYVHHP-YSV
CmaCas7-11	KWYISNDFQKDS--KKE-KGG-KLTIPLERPRPT--WSMPN--K--KGSNTEDATIPGRKRYLHHDGWRI	KWYISNDFQKDS--KKE-KGG-KLTIPLERPRPT--WSMPN--K--KGSNTEDATIPGRKRYLHHDGWRI
HvmCas7-11	DD--KQP-ETK-TLTIPLERPRPT--WVLPK--SV--KSEIPGRKRYVHHP-WSV	DD--KQP-ETK-TLTIPLERPRPT--WVLPK--SV--KSEIPGRKRYVHHP-WSV
SmCas7x3	ST--RSP-QLK-YVTIPLSQRPRPT--WVLPK--S--CKGKEDVPGRKRYLHHDGWRE	ST--RSP-QLK-YVTIPLSQRPRPT--WVLPK--S--CKGKEDVPGRKRYLHHDGWRE
oral meta	EWLI-----TR-GNN-PER-SLTIGVLESPPRA--FSIPD--D--ESEIPGRKRYLHHDGWRI	EWLI-----TR-GNN-PER-SLTIGVLESPPRA--FSIPD--D--ESEIPGRKRYLHHDGWRI
DisCas7-11	KWLT-----ER-ENG-CGG-YVTIPLERPRPT--WSVPS--D--KCDVPGRKRYLHHDGWRI	KWLT-----ER-ENG-CGG-YVTIPLERPRPT--WSVPS--D--KCDVPGRKRYLHHDGWRI
DsbaCas7-11	KWFR-----KD-EGN-ETC-HITLPLERPRPT--WSMPD--D--TSTIPGRKRYVHHPHAWKT	KWFR-----KD-EGN-ETC-HITLPLERPRPT--WSMPD--D--TSTIPGRKRYVHHPHAWKT
SstCas7-11	NW-----KF-REE-QLTIPLERPRPT--WVLPK--R--EVQIPGRKRYLHHDGWRI	NW-----KF-REE-QLTIPLERPRPT--WVLPK--R--EVQIPGRKRYLHHDGWRI
CsbCas7-11	KFCQ-----GG-ERP-EDR-AVTIPLERPRPT--WVMPN--E--NSTIPGRKRYLHHDGWRI	KFCQ-----GG-ERP-EDR-AVTIPLERPRPT--WVMPN--E--NSTIPGRKRYLHHDGWRI
CbfCas7-11	VW-----AK-GAG-GKD-YFTIPLERPRPT--WTMPN--E--GAKIPGRKRYVHHPHAWKT	VW-----AK-GAG-GKD-YFTIPLERPRPT--WTMPN--E--GAKIPGRKRYVHHPHAWKT
HvsCas7-11		
FmCas7-11		
DpbaCas7-11		
Domains		

FIG. 34L

140/142

WmCas7x3	QG---QVARDNSGHVPSDGSGRPPW-----QSRFDG-----QEDSGNK-----RRVRVEPIPAGETFFHFEDFDNLSPTLEQLCATLIP-HPAFEH
SER16298.1	QNG---NMKLDAGVNSGSGTCKPPMKTHKPAANDQKE-----FEEDKNKFTSQVKIAPISENTPFYFEIDFNNDATLAQCATLQP-APKFEH
GwCas7x3	NE---QVWELSDTGH---DGGVRPPW-----VSKFDE-----SADENK-----RRVSIPIAKDESFEYEDFHNLSRTLAQCATLQP-NEKFEH
PID64649.1	ENN-QNVQRTSQGKFDNAARNTQPW-----VS-----QNEERNH-----LKTCKPLKSGLNFFHIDFNNDLQWELGLLCYALRP-CETFRH
WP 031436019.1	WDN-KVQKLGSGTANGDQSKLPW-----VTH-----NPDERPQ-----LKVIKIPKAE---FISDLDNNTLWELGLLCYALRP-TDSFRH
HE50754.1	RGNPDVQSLDRYGH-ATETAANPPW-----ETC-----HPEERPQ-----IKVRQVPRKTKTFFHLDLSNLSRWELGLLCYVLRP-TACFRH
HreCas7-11	---D-----GI-----NQGN-----QKDISLKPDSDAIKKE-----DGETRTP-----NRTIEPIAEGNEFDIDFNLDWELGLLCYSL-EDSLAH
CmaCas7-11	ENI-K-----NRQFD-----PAK-EKQPD-DVIKPE-----NRTVEPLGKNEFTFEVRFNNDWELGLLCYSL-EDNMAH
HvmCas7-11	DKI-R-----GTNPI-----TG-----ESIKTA-----NNATVEAIMPGATFTDIVCENLDQOELGLLCYSL-EEGMSH
SmCas7x3	VM-A-----DDKPD-----SRPSEE-----CQDIEIGPGKEKHFHFAFNLDKNELGLLCYSL-DAGMNH
oral_meta	IK-D-----GNHPT-----TG-----KAIQSP-----NRTVEALAGNSFSFEIAFENLDKNELGLLCYSL-EKGLAH
DsbaCas7-11	IS-E-----GIHPI-----SG-----ENIPEDE-----NRTVEALDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
SstCas7-11	IN-C-----GCHPT-----TK-----ENIVQNO-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
CsbCas7-11	IR-Q-----KQLEI-----RETQPE-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
CjcCas7-11	VL-R-----NN-----DITPKE-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
CbfCas7-11	VK-K-----NQ-----RTLVKTE-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
HvsCas7-11	IIN-N-----SKK-----EKEKTE-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
HsmCas7-11	IYK-K-----NKK-----NEIKKEK-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
FmCas7-11	IVD-E-----GKNPI-----NG-----DVIEDA-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
DpbaCas7-11	VQ-E-----GKNPI-----DQ-----KAIKPNP-----NRTVEPLDKNRFEVLSFNLEPRELGLLCYSL-EKGLAH
Domains	
WmCas7x3	G-rich loop
SER16298.1	RLGKPKPLGLSVKLAVEGL---LLVDRPRRYAEDEPNAPRRHGW---RAN-ADAG-WPDHLQDSPAAPLEAT-EQP-AALAEARMA-----
GwCas7x3	RLGKPKPLGLSVKIEPVGL---FLNRHQRYTTDSTNCDRYHAW---LKG-EHAANDWPEYFRQNVVTDCTQTFNDFDKLVQNGLA-----
PID64649.1	RLGKPKPLGLSVKIDITPLSL---FLVNRSQRYATDGLKPRYHAW---HTG-TASEPRWPDHLQEQQIAFEGVSTAP-TVMSLAABA-----
WP 031436019.1	KIEMKPKPLGLSVKIDIAVL---QTIDRYARYTDTQDSERNQGA---WIS-QELQNEIPN-QYKKGISNKKGMLSL-EDCRKVME-----
HE50754.1	RIEMKPKPLGLSVKIDIMAL---QTIDRQORYAQDLEENRNRHN---WVN-PPHQPLP---EKAGYSISLSTPLN-EILRATFTK-----
HreCas7-11	KLGMKPKPLGLSVRIDIASL---OLIDRVRRYGTDLTAGRYNMGG---HFN-ASCLDLP---QODSPAPDDSGAAPDP-GTLRODFVK-----
CmaCas7-11	KLGMKPKPLGLSVQINIRGI---SLKN-----GSKG-WDTKTGDDKNQWIK---KGF-----AHLGIDIKE-----
HvmCas7-11	KLGMKPKPLGLSVKIEIKNL---LIKG-----SIND-----ISKAELIK---KGF-----KKLGIDSLEKD-----
SmCas7x3	KLGMKPKPLGLSVKIRIKAEAI---ELRC-----ESA---GQNAELKDKAFAVR---KGF-----EFLEIDKPGEN-----
oral_meta	HLGKPKPLGLSVKIRIVTKL---ERRL-----SDG-----SRREMIPT---KGA-----GLFMTDKVQDALRGLTEG-----
DisCas7-11	KLGMKPKPLGLSVKIDVESV---RLRK-----EPGQ---WRS-EKICTDLPV---TSS-----ELVSSLSKK-----
DsbaCas7-11	KLGMKPKPLGLSVKIDVESV---RVKH-----DWKQ---WRN-GNSEIPNWLG---KGF-----AKLKEWFRD-----
SstCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----RSGE---WDYKDGETVDGWIE---EGK-----RGVAAGKA-----
CsbCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----DSGQ---WKD-ANEQISEWTD---KKG-----KDAGKWFKT-----
CjcCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----NNDK---IKRVQSDIREYN---KGY-----QKLIENSGNSIQKGNVL-----
CbfCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----DGIN---WKP-SENEIGVYVQ---KGR-----EKLVEWTFPSAPHKM-----
HvsCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----EAGK---FHP-CEKADYVLC---KGL-----GKLTTFWFK-----
HsmCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----DPGQ---FQORGQKFRDAVD---KGF-----CVLENRFK-----
FmCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----QGE---WKP-CEKADYVLC---KGL-----TLVSEDKKS-----
DpbaCas7-11	KLGMKPKPLGLSVKIDIEENV---SLRT-----KSGT---WKNE-TSGREKFIQ---SGL-----SQVPSFTKQDEK-----
Domains	

FIG. 34M

141/142

WmCas7x3	--RVPADVRRALQLLGNP--	--GAVAAPVHYPOVKDAQIE--	--EKH--	--YLMFVANDK--	--EKTAGGNRH--	--LPLI--
SER16298.1	--GTDADIKHALQLLGNP--	--QYIGVPVHYPIAGNSTLE--	--NKH--	--FWEFGNDDK--	--ASVLRQKAQANSKNHHYQPKQOATPEEPQYLHTI--	--LSSF--
GwCas7x3	--KVSDDVVKRALELLGNP--	--DEITVPVHYPOLHNGLME--	--SKQ--	--FDWFFVQNDK--	--SGRDQPANNRQH--	--LSPF--
PID64649.1	--TMDADIQRAIEELLGDP--	--GNVTSVHYPPOLDRKNIE--	--TKN--	--YEFWKQNEI--	--EQOV--	--LKPI--
WP_031436019.1	--TMNADIYRTLELLGNP--	--QNVKRPVHYPOVENHNEI--	--QEN--	--YKWFVANDQ--	--GSGKGRNKIDPAEKA--	--LKIL--
HE50754.1	--TMDETVERALDGLNP--	--AHVQRPVHYPOVEMDIE--	--DQT--	--FLWFFVNDK--	--QWKDA--	--LQPL--
HreCas7-11	--ANERPVIKQRELLWVP--	--TGDNLPHVRYPELESTKD--	--VPGYTSLLKE--	--KDL--	--ADR--	--VSLI--
CmaCas7-11	--DLSEYLHIKQRELLWFS--	--DKPVGTIEYPKLENKNTNS--	--RIPSYTDFVOEKDHETGFKNPQYQNL--	--KSR--	--KSR--	--LHIL--
HvmCas7-11	--DPMNFDHIQRELLWFL--	--PENVSANRYPMLEKEDDG--	--TPGYTDFIKQEEFSTGKRNPYLSS--	--EKR--	--EKR--	--RNIL--
SmCas7x3	--GDWHQPHISGLRLLTRY--	--PEIKARYPKLSQEDK--	--EPGYIELKSQKDENGVPI--	--YNP--	--NRE--	--LRVS--
oral meta	--VEERRKLLRLVMTF--	--YKGLTACYPGLEENGR--	--PGYTDLKM--	--ATY--	--DPY--	--RELV--
DisCas7-11	--ET ~~ET~~ FIENLKLRLWFP--	--EGDQAPRVCPYMLRKDD ~~ET~~ PP--	--NGN--	--SGYEELKDGE--	--FKK--	--QKRL--
DsbaCas7-11	--NDLRKLLYLP--	--GEKONPHVHYPTLKKEKKG--	--DPGYEDLKKSEF--	--LNR--	--LNR--	--RKML--
SstCas7-11	--DWEAAEHKLNKLLFLP--	--GEEQNPVVIYPALQKQDIP--	--NSRLPGYEELKKN--	--LNM--	--EKR--	--KEML--
CsbCas7-11	--POHVTIPHIDKLYKLLWVPFLNDSKLEPDVRYPVINEESKGYIEGSDYTYKKLGDKON--					
CjcCas7-11	--EWNGVKHIKDLRLLSIP--	--GDKPTVKYPTLNKDAEG--	--AISDYTYERLSDTKL--			
CbfCas7-11	--EWDREHIQGLRSLQRL--					
HvsCas7-11	--NEINEHMRNLLLFMTYY--	--QNLPKVKYPDF--				
HsmCas7-11	--TNFKI--	--YLNFLQLLYWP--	--NNKVKVRYPYLRQEDDP--	--EKL--		
FmCas7-11	--SWNIHTVNNLRALYYV--	--EDDKIQVNYPKLKDNE--	--DNRPGYVEMKTKTA--			
DpbaCas7-11	--QWNKVEQVKNIRKLLQLSWNKGNAVEPVRYPALREKDE--					
Domains						
WmCas7x3	--HANSPLPLPLRV--	--KREXDHSSNTGKPRK--				
SER16298.1	--TKDSKQISLLKKNKTIENRDQOKHRYSNHR--					
GwCas7x3	--TKDTEKLEPLIRIM--	--RR--				
PID64649.1	--TKNTHLTPFAR--					
WP_031436019.1	--TENSDDLPLSLRD--	--WRDE--				
HreCas7-11	--TSSSTQLPLTRRN--	--KR--				
CmaCas7-11	--KA--	--PWPWKPS--				
HvmCas7-11	--QN--	--PWNAAWKNE--				
SmCas7x3	--QT--	--PWKHLYLIP--				
oral meta	--EN--	--GPLPWFLA--				
DisCas7-11	--VQISNQ--	--PLRPWYEPG--				
DsbaCas7-11	--TT--	--PWTPTWA--				
SstCas7-11	--TT--	--LWEPWHK--				
CsbCas7-11	--TT--	--PWAPWHPIK--				
CjcCas7-11	--TT--	--PWSPWNPFO--				
CbfCas7-11	--RT--	--PWSWNAFV--				
HvsCas7-11	--QN--	--RWIVAS--				
HsmCas7-11	--LFAPRA--	--VWLKWTEMV--				
FmCas7-11	--TT--	--PWPWMPPT--				
DpbaCas7-11	--VS--	--PWAPWHPIK--				
Domains						
WmCas7x3						
SER16298.1						
GwCas7x3						
PID64649.1						
WP_031436019.1						
HreCas7-11						
CmaCas7-11						
HvmCas7-11						
SmCas7x3						
oral meta						
DisCas7-11						
DsbaCas7-11						
SstCas7-11						
CsbCas7-11						
CjcCas7-11						
CbfCas7-11						
HvsCas7-11						
HsmCas7-11						
FmCas7-11						
DpbaCas7-11						
Domains						

FIG. 34N

WmCas7x3	-----
SER16298.1	-----
GwCas7x3	-----
PID64649.1	-----
WP_031436019.1	-----
HEB50754.1	-----
HreCas7-11	EEIFVHRSGVADNSIPKEGQKVGFRIERGARGSHAVEVKAIE----
CmaCas7-11	-----
HvmCas7-11	KDVFVHSSIVGTGFKSLNEGDSVAFKMGVGPKGPCAEKVKKIGN
SmCas7x3	-----
oral_meta	RDIFVHFSAIRGEGYKILEPGEKVRFEIGEGRKGPQAINVIRIR--
DisCas7-11	-----
DsbaCas7-11	-----
SstCas7-11	-----
CsbCas7-11	-----
CjcCas7-11	KEVSISKNSIRGNILLKKGQKVTFHIVQGLIPKAEDIEIAK----
CbfCas7-11	-----
HvsCas7-11	-----
HsmCas7-11	-----
FmCas7-11	NIHFAGNQICRPET--SLQSGDKVKFIEGENYKGPPTALKVERLKG
DpbaCas7-11	-----
Domains	

FIG. 340