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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 December 2022 (22.12.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/266298 A1

(51) International Patent Classification:
C12N 9/22 (2006.01) *C12N 15/10* (2006.01)

(21) International Application Number:
PCT/US2022/033749

(22) International Filing Date:
16 June 2022 (16.06.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/211,610 17 June 2021 (17.06.2021) US

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

(72) Inventors: **JIANG, Kaiyi**; 70 Pacific Street, Cambridge, MA 02139 (US). **VILLIGER, Lukas**; 88 Ames Street, Cambridge, MA 02142 (US). **ABUDAYYEH, Omar, Osama**; 49 Bristol Street, Cambridge, MA 02141 (US). **GOOTENBERG, Jonathan, S.**; 8 Richardson Street, Cambridge, MA 02139 (US).

(74) Agent: **LABEOTS, Laura, A.** et al.; Lathrop GPM LLP, 155 North Wacker Drive, Suite 3000, 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, IQ, IR, IS, IT, JM, 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:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: SYSTEMS, METHODS, AND COMPOSITIONS COMPRISING MINIATURE CRISPR NUCLEASES FOR GENE EDITING AND PROGRAMMABLE GENE ACTIVATION AND INHIBITION

(57) Abstract: This disclosure provides systems, methods, and compositions comprising miniature CRISPR. nucleases for gene editing and programmable gene activation and inhibition. The miniature CRISPR nuclease is a target specific nuclease having a compact structure with a small number of amino acids. The target specific nuclease targets DNA and is directed to a target nucleic acid sequence from the DNA by a guide RNA. In some embodiments, the target specific nuclease exhibits DNA cleavage activity and is directed by a gRNA to a target nucleic acid sequence from a DNA. In some embodiments, the target specific nuclease does not exhibit DNA cleavage activity and is directed by a gRNA to a target nucleic acid sequence from a DNA.

WO 2022/266298 A1

SYSTEMS, METHODS, AND COMPOSITIONS COMPRISING MINIATURE CRISPR NUCLEASES FOR GENE EDITING AND PROGRAMMABLE GENE ACTIVATION AND INHIBITION

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of U.S. Provisional Patent Application Serial No. 63/211,610, filed June 17, 2021. The entirety of this application is hereby incorporated by reference.

FIELD OF INVENTION

The subject matter disclosed herein is generally directed to systems, methods, and compositions comprising miniature CRISPR nucleases for gene editing and programmable gene activation and inhibition.

BACKGROUND

Cluster Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated (Cas) nuclease systems are widely used as genome editing tools. Cas9 and Cas12 are two examples of nucleases that are often used in CRISPR-Cas system to edit genomes. These nucleases are generally more than 1000 amino acids long and can be guided by a guide RNA to edit a single stranded or double-stranded DNA target near a short sequence called protospacer adjacent motif (PAM). However, while these nucleases offer great flexibility, their size remains a significant barrier to their use. For example, gene editing and programmable gene activation and inhibition technologies based on these nucleases can generally not be delivered in mouse models using common methods such as adeno-associated vectors (AAV) because of the large size of the nuclease. Furthermore, development of effective gene and cell therapies requires genome editing tools that can meet the demands for reduced payload sizes and efficient integration of diverse and large sequences, regardless of cell type or active repair pathways. CRISPR associated transposases, such as Cas12k or type I-F directed Tn7 systems, allow for programmable integration in bacteria without the need for repair-pathway dependent editing, but have yet to be

reconstituted in eukaryotic cells for mammalian genome editing. The difficulty in reconstitution of these systems can be due to the sheer number of proteins (4-7 proteins) that must be properly expressed and delivered to the nucleus for proper assembly and DNA targeting. Prime editing was also reported for programmable gene editing independent of DNA repair pathways but is limited to base substitutions or small deletions and insertions (about < 50 bp).

Thus, there is a need for smaller and more compact CRISPR nucleases for gene editing, programmable gene activation and inhibition, and new applications. Smaller and more compact CRISPR nucleases can simplify delivery and extend application, and the additional space on such nucleases can enable fusion with effector domains.

SUMMARY

The present disclosure provides systems, methods, and compositions comprising miniature CRISPR nucleases for gene editing and programmable gene activation and inhibition.

In one aspect, this disclosure pertains to a composition comprising a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, and a guide RNA (gRNA), wherein a target comprises a DNA target. In some embodiments, the DNA target can be a single stranded DNA. In some embodiments, the DNA target can be a double stranded DNA. In some embodiments, the target specific nuclease can have a length less than about 1000 amino acids. In some embodiments, the target specific nuclease can have a length less than about 900 amino acids. In some embodiments, the target specific nuclease can have a length less than about 800 amino acids. In some embodiments, the amino acid sequence can be SEQ ID NO: 1. In some embodiments, the target specific nuclease can comprise an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO: 1, or an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO: 1, or an amino acid sequence 98% identical to the amino acid sequence of SEQ ID NO: 1, an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the nuclease can be the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the target specific nuclease can be selected from the group consisting of Cas12m, Cas12f, and any variants thereof, and optionally the target specific nuclease can be PsaCas12f.

In some embodiments, the gRNA can be a single guide RNA (sgRNA) or a dual guide (dgRNA). In some embodiments, the gRNA can be a sgRNA and the sgRNA can comprise a nucleic acid sequence 75% identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43 and 61-79. In some embodiments, the gRNA can have a spacer region with a sequence comprising a length of about 17 to about 53 nucleotides (nt), optionally the sequence can comprise a length of about 29 to about 53 nt, optionally the sequence can comprise a length of about 40 to about 50 nt, or optionally the sequence can comprise a length of about 22 nt. In some embodiments, the gRNA can have a direct repeat region with a sequence having a length of from about 20 to about 29 nt. In some embodiments, the gRNA can have a tracrRNA region with a sequence having a length of from about 27 to about 35 nt.

In some embodiments, the DNA target can be in a cell. In some embodiments, the cell can be a prokaryotic cell. In some embodiments, the cell can be a eukaryotic cell. In some embodiments, the eukaryotic cell can be a mammalian cell. In some embodiments, the mammalian cell can be a human cell.

In some embodiments, the amino acid sequence can specifically bind to a protospacer-adjacent motif (PAM). In some embodiments, the PAM can be selected from the group consisting of NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

In another aspect, a nucleic acid molecule encoding a target specific nuclease is discussed.

In another aspect, a nucleic acid molecule encoding a guide RNA is discussed.

In another aspect, one or more vectors comprising a nucleic acid molecule encoding a target specific nuclease and/or a guide RNA is discussed.

In another aspect, a cell comprising a composition comprising a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, a target comprises a DNA, and a guide RNA; or a cell comprising a nucleic acid molecule encoding the target specific nuclease; or a cell comprising a

nucleic acid molecule encoding the gRNA; or a cell comprising one or more vectors comprising a nucleic acid molecule encoding the target specific nuclease and/or the guide RNA is discussed. In some embodiments, the cell can be a prokaryotic cell. In some embodiments, the cell can be a eukaryotic cell. In some embodiments, the eukaryotic cell can be a mammalian cell. In some embodiments, the mammalian cell can be a human cell.

In another aspect, a method of inserting or deleting one or more base pairs in a DNA is discussed, the method comprising cleaving the DNA at a target site with a target specific nuclease, the cleavage results in overhangs on both DNA ends, inserting a nucleotide complementary to the overhanging nucleotide on both of the dsDNA ends, or removing the overhanging nucleotide on both of the DNA ends, and ligating the dsDNA ends together, thereby inserting or deleting one or more base pairs in the dsDNA, the nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, and the target specificity of the target specific nuclease is provided by a guide RNA (gRNA). In some embodiments, the target specific nuclease can have a length less than about 1000 amino acids. In some embodiments, the target specific nuclease can have a length less than about 900 amino acids. In some embodiments, the target specific nuclease can have a length less than about 800 amino acids. In some embodiments, the amino acid sequence can be SEQ ID NO: 1.

In some embodiments, the target specific nuclease can comprise an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the target specific nuclease can comprise an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the target specific nuclease can comprise an amino acid sequence 98% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the target specific nuclease can comprise an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO: 1. In some embodiments, the nuclease can be the amino acid sequence of SEQ ID NO: 1.

In some embodiments, the target specific nuclease can be selected from the group consisting of Cas12f, Cas12m, and any variants thereof, and optionally the target specific nuclease can be PsaCas12f.

In some embodiments, the gRNA can be a single guide RNA (sgRNA) or a dual guide RNA (dgRNA). In some embodiments, the gRNA can be a sgRNA comprising a nucleic acid

sequence 70% identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43 and 61-79. In some embodiments, the gRNA comprises a spacer region with a sequence having a length of from about 20 to about 30 nucleotides (nt), about 22 nt; or the gRNA comprises a spacer region with sequence having a length of from about 20 to about 53 nt, or from about 29 to about 53 nt or from about 40 to about 50 nt.

In some embodiments, the DNA target can be in a cell. In some embodiments, the cell can be a prokaryotic cell. In some embodiments, the cell can be a eukaryotic cell. In some embodiments, the eukaryotic cell can be a mammalian cell. In some embodiments, the mammalian cell can be a human cell.

In some embodiments, the amino acid sequence can specifically bind to a protospacer-adjacent motif (PAM). In some embodiments, the PAM can be selected from the group consisting of NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

In another aspect, a method of detecting a DNA target is discussed, the method comprising coupling the DNA target with a reporter to form a DNA-reporter complex, mixing the DNA-reporter complex with a target specific nuclease and a guide RNA (gRNA), cleaving the DNA-reporter complex, and measuring a signal from the reporter, thereby detecting the DNA target. In some embodiments, the target specific nuclease can be selected from the group consisting of Cas12f, Cas12m, and any variants thereof, and optionally the target specific nuclease can be PsaCas12f. In some embodiments, the target specific nuclease can be complexed with a crRNA. In some embodiments, the reporter can be a fluorescent reporter.

In another aspect, a method for activating or inhibiting the expression of a gene is discussed, the method comprising mixing a composition with one or more transcription factors, the composition comprising a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, a DNA target, and a guide RNA (gRNA), the target specific nuclease lacks endonuclease ability, and the target DNA comprises the gene, thereby activating the gene.

In another aspect, a method for nucleic acid base editing is discussed, the method comprising mixing a composition, the composition comprising a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the

group consisting of SEQ ID NOs: 1-19, a DNA target, and a guide RNA (gRNA), the target specific nuclease is a nickase or a nuclease coupled to a deaminase, thereby editing the nucleic acid base from the target DNA.

In another aspect, a method for activating or inhibiting the expression of a gene is discussed, the method comprising mixing a composition comprising a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, and a guide RNA (gRNA), a target comprises a DNA target, with one or more epigenetic modifiers, the target specific nuclease lacks endonuclease activity, the target DNA comprises the gene, and modifying the target DNA or one or more histones associated to the target DNA, thereby activating or inhibiting the gene. In some embodiments, the epigenetic modifier can comprise KRAB, DNMT3a, DNMT1, DNMT3b, DNMT3L, TET1, p300, any variants thereof, or any combinations thereof.

These aspects and embodiments, as well as others, are disclosed in further detail herein.

BRIEF DESCRIPTION OF THE DRAWINGS

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

FIG. 1A shows a schematic diagram illustrating the computational identification of novel miniature CRISPR nucleases from metagenomic samples according to embodiments of the present teachings;

FIG. 1B shows a simulated tree of Cas orthologs according to embodiments of the present teachings;

FIG. 1C shows the size distribution of Cas12a ortholog according to embodiments of the present teachings;

FIG. 1D shows the size distribution of CasM ortholog according to embodiments of the present teachings;

FIG. 1E shows the secondary structure prediction of PasCas12f direct repeat according to embodiments of the present teachings;

FIG. 1F shows the secondary structure prediction of putative PasCas12 tracrRNA according to embodiments of the present teachings;

FIG. 2 shows a schematic diagram illustrating the screening of smaller CRISPR nucleases for functional activity via LASSO and TXTL according to embodiments of the present teachings;

FIG. 3A shows a vector map depicting single-vector activators, base editors, or homology directed repair (HDR) enabled by smaller CRISPR nucleases according to embodiments of the present teachings;

FIG. 3B shows a schematic diagram illustrating *in vivo* modification via single-vector activators, base editors, or HDR with AAV according to embodiments of the present teachings;

FIG. 3C shows the optimization of small CRISPR effectors for mammalian single-vector delivery according to embodiments of the present teachings;

FIG. 4 shows the testing of PsaCas12f sgRNA constructs in human mammalian cells according to embodiments of the present teachings;

FIG. 5A shows the testing of PsaCas12f NLS constructs according to embodiments of the present teachings;

FIG. 5B shows the editing with PsaCas12f (NLS14) with sgRNA 13 according to embodiments of the present teachings;

FIG. 5C shows the editing with PsaCas12f (NLS14) with non-targeting guide according to embodiments of the present teachings;

FIG. 5D shows the editing with PsaCas12f (no NLS) with sgRNA 14 according to embodiments of the present teachings;

FIG. 5E shows the editing with PsaCas12f (no NLS) with non-targeting guide according to embodiments of the present teachings;

FIG. 6A shows a process for optimal guide RNA prediction according to embodiments of the present teachings;

FIG. 6B shows predicted energy landscape for different RNA designs according to embodiments of the present teachings;

FIG. 6C shows *in vitro* cleavage with PsaCas12f using different sgRNA scaffolds generated by *in silico* optimization according to embodiments of the present teachings;

FIG. 7A shows a diagram of luciferase indel reporter for engineering novel CRISPR effectors like PsaCas12f for mammalian genome editing according to embodiments of the present teachings;

FIG. 7B shows genome editing data with PasCas12f in HEK293FT cells showing about 0.05% indel activity that is 100 times higher than background detection, wherein activity is detected with N-terminal NLS Cas12f expression and natural guide scaffold according to embodiments of the present teachings;

FIG. 7C shows a bar graph of gene editing with PasCas12f in HEK293FT cells according to embodiments of the present teachings;

FIG. 7D shows allele plot of Cas12f EMX1 cleavage showing indels at target according to embodiments of the present teachings;

FIG. 7E shows a bar graph of the sgRNA and DR/tracr optimization for Cas12f, wherein the luciferase reporter for indels reveals key sgRNA and tracrRNA/DR combos that have indel activity in HEK293FT cells according to embodiments of the present teachings;

FIG. 8A shows a schematic of PsaCas12f expression locus according to embodiments of the present teachings;

FIG. 8B shows the PasCas12f PAM determined by *in vitro* cleavage according to embodiments of the present teachings;

FIG. 8C shows the putative crRNA determined by small RNA sequencing according to embodiments of the present teachings;

FIG. 8D shows the validation of PasCas12f PAM *in vitro* cleavage with recombinant protein according to embodiments of the present teachings;

FIG. 9A shows PsaCas12f coupled to MiniVPR for CRISPR activation (CRISPRa) using dead PsaCas12f according to embodiments of the present teachings;

FIG. 9B shows a bar graph of the RLU for PsaCas12f coupled to VPR and MiniVPR, demonstrating that gene activation using MiniVPR and VPR can be achieved with catalytically dead PsaCas12f, wherein pDF235 and EMX1v2 reporters are different luciferase reporters for measuring gene activation according to embodiments of the present teachings;

FIG. 9C shows a bar graph of the RLU of PsaCas12f coupled with small linker sequences (5-10aa) at 6 different positions according to embodiments of the present teachings; and

FIG. 9D shows a bar graph of the fluorescence for PasCas12f based on target specific collateral activity, which can be used for diagnostics according to embodiments of the present teachings.

FIG. 10A illustrates the resulting sgRNA secondary structure derived from an *in silico* secondary structure determination with stem loop 1-3 boxed (SL1-3) predicted using via http://rna.tbi.univie.ac.at/. Stem loop 4 (SL4, interacts with crRNA) and stem loop 5 (SL5) were informed by Takeda et al., Mol Cell, 81(3):558-570 (2021).

FIG. 10B displays the annotated stem-loop sequence for the sgRNA stem-loop variants which were mutated to analyze the impact of gene editing efficiencies. Red denotes nucleobase changes that were introduced, orange denotes nucleobases that form stems, and violet denotes loops that were added to allow recruitment of MS2 coat/proteins.

FIG. 10C shows a bar graph of the RLU using PsaCas12f with the different sgRNA stem-loop variants demonstrating that modifications to the secondary structure of the sgRNA impacts gene editing efficiencies.

FIG. 11A shows a bar graph of the RLU using PsaCas12f with a panel of sgRNA variants which each have a combination of the modifications derived from single modification sgRNA stem-loop variants.

FIG. 11B shows a bar graph of the percent indel formation at the EMX1 genomic locus using PsaCas12f with a panel of sgRNA variants which each have a combination of modifications derived from the single sgRNA stem-loop variants (4x combinations, left panel and 2x combinations, right panel).

FIG. 11C shows a bar graph of the RLU using a panel of thirty mutant PsaCas12f with the two best sgRNA combination stem-loop variants (named scaffold version 3.1 and scaffold version 3.2) demonstrating the robustness of the sgRNA scaffold version 3.2.

FIG. 12A is a schematic of the sgRNA scaffold named version 3.2 which highlights the position of the spacer sequence at the 3'end.

FIG. 12B shows a bar graph of the RLU using PsaCas12f with a panel of version 3.2 sgRNA scaffolds which have varying spacer lengths (2, 3, 18, 19, 20, 21, 22, 23, 24, and 25 base pairs).

FIG. 13 shows the percent indel formation at two different positions within the HBB and the RNF genomic loci (HBB g1, HBB h2, RNF g4, and RNF g6) using either the PsaCas12f with the sgRNA scaffold version 3.2 or the Un1Cas12f1 with nbt scaffold.

FIG. 14 shows a bar graph of the percent indel formation at the EMX genomic locus using a panel of PsaCas12 variants (intra-protein NLS constructs 1-6) where the NLS sequence

derived from SV40 was fused at random positions in the PsaCas12f sequence (as shown in bottom schematic).

FIG. 15 shows a bar graph of the percent indel formation at the RUNX1 genomic locus using a PsaCas12f with a sgRNA scaffold (has a flanking SV40 NLS) which was delivered to cells via AAV particles.

FIG. 16A shows a bar graph of the RLU using a panel of 12 circular permutated PsaCas12f mutants (named cpPsaCas12_1-12). The bottom schematic depicts how the PsaCas12f sequence can be split at different positions to create new N- and C- termini by inserting a $(GGS)_6$ peptide linker.

FIG. 16B shows a bar graph of the percent indel formation at the RUNX1 genomic locus using a panel of 12 circular permutated PsaCas12f mutants (cpPsaCas12_1-12).

FIG. 17 shows a bar graph of the percent indel formation at the RNF2 genomic locus using a panel of PsaCas12f mutants obtained from a machine learning model which predicted point mutations which could result in higher gene editing efficiencies. PsaCas12f variant with a point mutation at position 333 dramatically increased cleavage efficiency.

DETAILED DESCRIPTION

It will be appreciated that for clarity, the following disclosure will describe various aspects of embodiments. 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 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 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.

Definitions

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 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. 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 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).

As used herein, the singular forms "a", "an," and "the" include both singular and plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells.

As used herein, 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.

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

As used herein, the term "about" or "approximately" refers 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 disclosed.

As used herein, the term "polypeptide" and the likes refer to an amino acid sequence including a plurality of consecutive polymerized amino acid residues (e.g., at least about 2 consecutive polymerized amino acid residues). "Polypeptide" refers to an amino acid sequence, oligopeptide, peptide, protein, enzyme, nuclease, or portions thereof, and the terms "polypeptide," "oligopeptide," "peptide," "protein," "enzyme," and "nuclease," are used interchangeably.

Polypeptides as described herein also include polypeptides having various amino acid additions, deletions, or substitutions relative to the native amino acid sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain non-conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure. In some embodiments, polypeptides that are homologs of a polypeptide of the present disclosure contain conservative changes of certain amino acids relative to the native sequence of a polypeptide of the present disclosure, and thus may be referred to as conservatively modified variants. A conservatively modified variant may include individual substitutions, deletions or additions to a polypeptide sequence which result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well-known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the disclosure. The following eight groups contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (see, e.g., Creighton, Proteins (1984)). A modification of an amino acid to produce a chemically similar amino acid may be referred to as an analogous amino acid.

The term "variant" as used herein means a polypeptide or nucleotide sequence that differs from a given polypeptide or nucleotide sequence in amino acid or nucleic acid sequence by the addition (e.g., insertion), deletion, or conservative substitution of amino acids or nucleotides, but that retains some or all the biological activity of the given polypeptide (e.g., a variant nucleic acid could still encode the same or a similar amino acid sequence). A conservative substitution of an amino acid, i.e., replacing an amino acid with a different amino acid of similar properties (e.g., hydrophilicity and degree and distribution of charged regions) is recognized in the art as typically involving a minor change. These minor changes can be identified, in part, by considering the hydropathic index of amino acids, as understood in the art (see, e.g., Kyte et al., J. Mol. Biol., 157: 105-132 (1982)). The hydropathic index of an amino acid is based on a consideration of its hydrophobicity and charge. It is known in the art that amino acids of similar hydropathic indexes can be substituted and still retain protein function. The present disclosure provides amino acids having hydropathic indexes of ±2 that can be substituted. The hydrophilicity of amino acids also can be used to reveal substitutions that would result in proteins retaining some or all biological functions. A consideration of the hydrophilicity of amino acids in the context of a peptide permits calculation of the greatest local average hydrophilicity of that peptide, a useful measure that has been reported to correlate well with antigenicity and immunogenicity (see, e.g., U.S. Pat. No. 4,554,101). Substitution of amino acids having similar hydrophilicity values can result in peptides retaining some or all biological activities, for example immunogenicity, as is understood in the art. The present disclosure provides substitutions that can be performed with amino acids having hydrophilicity values within ±2 of each other. Both the hydrophobicity index and the hydrophilicity value of amino acids are influenced by the particular side chain of that amino acid. Consistent with that observation, amino acid substitutions that are compatible with biological function are understood to depend on the relative similarity of the amino acids, and particularly the side chains of those amino acids, as revealed by the hydrophobicity, hydrophilicity, charge, size, and other properties.

The term "variant" also can be used to describe a polypeptide or fragment thereof that has been differentially processed, such as by proteolysis, phosphorylation, or other post-translational modification, yet retains some or all its biological and/or antigen reactivities. Use of "variant" herein is intended to encompass fragments of a variant unless otherwise contradicted by context.

The term "protospacer-adjacent motif" as used herein refers to a DNA sequence immediately

following a DNA sequence targeted by a nuclease. Examples of protospacer-adjacent motif include, without limitation, NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

Alternatively, or additionally, a "variant" is to be understood as a polynucleotide or protein which differs in comparison to the polynucleotide or protein from which it is derived by one or more changes in its length or sequence. The polypeptide or polynucleotide from which a protein or nucleic acid variant is derived is also known as the parent polypeptide or polynucleotide. The term "variant" comprises "fragments" or "derivatives" of the parent molecule. Typically, "fragments" are smaller in length or size than the parent molecule, whilst "derivatives" exhibit one or more differences in their sequence in comparison to the parent molecule. Also encompassed modified molecules such as but not limited to post-translationally modified proteins (e.g., glycosylated, biotinylated, phosphorylated, ubiquitinated, palmitoylated, or proteolytically cleaved proteins) and modified nucleic acids such as methylated DNA. Also, mixtures of different molecules such as but not limited to RNA-DNA hybrids, are encompassed by the term "variant". Typically, a variant is constructed artificially, by gene-technological means whilst the parent polypeptide or polynucleotide is a wild-type protein or polynucleotide. However, also naturally occurring variants are to be understood to be encompassed by the term "variant" as used herein. Further, the variants usable in the present disclosure may also be derived from homologs, orthologs, or paralogs of the parent molecule or from artificially constructed variant, provided that the variant exhibits at least one biological activity of the parent molecule, *i.e.*, is functionally active.

Alternatively, or additionally, a "variant" as used herein can be characterized by a certain degree of sequence identity to the parent polypeptide or parent polynucleotide from which it is derived. More precisely, a protein variant in the context of the present disclosure exhibits at least 80% sequence identity to its parent polypeptide. A polynucleotide variant in the context of the present disclosure exhibits at least 70% sequence identity to its parent polynucleotide. The term "at least 70% sequence identity" or the like is used throughout the specification with regard to polypeptide and polynucleotide sequence comparisons. This expression refers to a sequence identity of at least 70%, at least 71%, at least 72%, at least 73%, at least 74%, at least 75%, at least 76%, at least 77%, at least 78%, at least 79%, at least 80%, at least 81%, at least 82%, at

least 83%, at least 84%, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% to the respective reference polypeptide or to the respective reference polynucleotide.

The similarity of nucleotide and amino acid sequences, *i.e.*, the percentage of sequence identity, can be determined via sequence alignments. Such alignments can be carried out with several art-known algorithms, with the mathematical algorithm of Karlin and Altschul (Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5877), with hmmalign (HMMER package, hmmer.wustl.edu/) or with the CLUSTAL algorithm (Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) Nucleic Acids Res. 22, 4673-80) available e.g. on www.ebi.ac.uk/Tools/clustalw/ or on www.ebi.ac.uk/Tools/clustalw2/index.html or on npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_clustalw.html. Some parameters used are the default parameters as they are set on www.ebi.ac.uk/Tools/clustalw/ or www.ebi.ac.uk/Tools/clustalw2/index.html. The grade of sequence identity (sequence matching) may be calculated using *e.g.*, BLAST, BLAT or BlastZ (or BlastX). A similar algorithm is incorporated into the BLASTN and BLASTP programs of Altschul et al. (1990) J. Mol. Biol. 215: 403-410. To obtain gapped alignments for comparative purposes, Gapped BLAST is utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25: 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs can be used. Sequence matching analysis may be supplemented by established homology mapping techniques like Shuffle-LAGAN (Brudno M., Bioinformatics 2003b, 19 Suppl 1:I54-I62) or Markov random fields. When percentages of sequence identity are referred to in the present application, these percentages are calculated in relation to the full length of the longer sequence, if not specifically indicated otherwise.

As used herein, the term "miniature CRISPR nuclease" and the like refer to a "target specific nuclease" having a compact structure with a small number of amino acids.

As used herein, the term "target specific nuclease" and the like refer to a nuclease that targets DNA and is directed to a target nucleic acid sequence from the DNA by a guide RNA (gRNA). The DNA can be a single stranded DNA or a double stranded DNA.

As used herein, the term "guide RNA" (gRNA) and the like refer to an RNA that guides the editing, activation or inhibition of one or more genes of interest or one or more nucleic acid

sequences of interest into a target genome. A gRNA is capable of targeting a nuclease to a target nucleic acid or sequence in a genome. The gRNA can also refer to a prime editing guide RNA (pegRNA), a nicking guide RNA (ngRNA), a single guide RNA (sgRNA), i.e., a fusion of two noncoding RNAs, a synthetic CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA), and a dual guide RNA (dgRNA). In some embodiments, the term "gRNA molecule" or the like refer to a nucleic acid encoding a gRNA. In some embodiments, a gRNA molecule is non-naturally occurring. In some embodiments, a gRNA molecule is a synthetic gRNA molecule.

As used herein, the term "target" or the like refer to a polynucleotide or polypeptide that is targeted. In some embodiments, the target is a DNA target. In some embodiments, the DNA target is associated with one or more histones. In some embodiments, the DNA target is a double-stranded DNA target. In other embodiments, the DNA target is a single-stranded DNA target.

As used herein, the terms "circular permutation," "circularly permuted," and "(CP)," refer to the conceptual process of taking a linear protein, or its cognate nucleic acid sequence, and fusing the native N- and C-termini (directly or through a linker, using protein or recombinant DNA methodologies) to form a circular molecule, and then cutting the circular molecule at a different location to form a new linear protein, or cognate nucleic acid molecule, with termini different from the termini in the original molecule. Circular permutation thus preserves the sequence, structure, and function of a protein (other than the optional linker), while generating new C- and N-termini at different locations that, in accordance with one aspect of the invention, results in an improved orientation for fusing a desired polypeptide fusion partner as compared to the original ligand. Circular permutation also includes any process that results in a circularly permutated straight-chain molecule, as defined herein. In general, a circularly permuted molecule is de novo expressed as a linear molecule and does not formally go through the circularization and opening steps.

It is noted that all publications and references cited herein are expressly incorporated herein by reference in their entirety. The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication.

Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Overview

The embodiments disclosed herein provide non-naturally occurring or engineered systems, methods, and compositions comprising miniature CRISPR nucleases for gene editing and programmable gene activation and inhibition. The miniature CRISPR nuclease is a target specific nuclease having a compact structure with a small number of amino acids. The target specific nuclease targets single stranded or double stranded DNA and is directed to a target nucleic acid sequence from the DNA by a guide RNA (gRNA). The gRNA can be a single-guide RNA, i.e., a fusion of two non-coding RNA: a synthetic CRISPR RNA (crRNA) and a trans-activating CRISPR RNA (tracrRNA). The crRNA and tracrRNA aid in directing the target specific nuclease to a target nucleic acid sequence, and these RNA molecules can be specifically engineered to target specific nucleic acid sequences. Certain aspects of the present teachings involve a target specific nuclease that exhibits DNA cleavage activity and is directed to a target nucleic acid sequence from a DNA by a gRNA. Certain aspects of the present teachings involve a target specific nuclease that does not exhibit DNA cleavage activity and is directed to a target nucleic acid sequence from a DNA by a gRNA molecule. Certain aspects of the present teachings involve a target specific nuclease for diagnostic applications.

Miniature CRISPR Nucleases

Some embodiments disclosed herein are directed to non-naturally occurring or engineered 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 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. 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.

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.

The CRISPR-Cas systems disclosed herein comprise a miniature CRISPR nuclease. The miniature CRISPR nuclease is a target specific nuclease that has a compact structure with a small number of amino acids and targets DNA. The target specific nuclease disclosed herein can be for example, without limitation, Cas12f, Cas12m, and any variants thereof, and optionally the target specific nuclease can be PsaCas12f. In some embodiments, the target specific nuclease is a nuclease that edits a single stranded or double stranded DNA. In some embodiments, the target specific nuclease is a nuclease that edits a single-stranded DNA (ssDNA). In some embodiments, a target specific nuclease is a nuclease that edits a double-stranded DNA. In some embodiments, the target specific nuclease is a nuclease that edits DNA in the genome of a cell.

The CRISPR-Cas systems disclosed herein can comprise one or more epigenetic modifiers. Examples of epigenetic modifiers include, without limitation, KRAB, DNMT3a, DNMT1, DNMT3b, DNMT3L, TET1, p300, any variants thereof, and any combinations thereof.

The target specific nuclease can comprise an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19. For example, the target specific nuclease comprises an amino acid sequence at least 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%, or 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19.

In some embodiments, the target specific nucleases include tags such as for example, without limitation, 3xFlag, nuclear localization sequence (NLS), and the combination of 3xFlag and NLS.

The CRISPR-Cas systems disclosed herein comprise a guide RNA (gRNA). The gRNA directs the target specific nuclease to a target nucleic acid sequence from a single stranded or double stranded DNA targeted by the nuclease. In some embodiments, the gRNA is a single-guide RNA (sgRNA). In some embodiments, the gRNA comprises a CRISPR RNA (crRNA), a trans-activating CRISPR RNA (tracrRNA), or a combination thereof. The crRNA and tracrRNA aid in directing the target specific nuclease to a target nucleic acid sequence, and these RNA molecules can be specifically engineered to target specific nucleic acid sequences.

In general, a guide sequence from the gRNA is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a target specific nuclease to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 52%, 54%, 56%, 58%, 60%, 62%, 64%, 66%, 68%, 70%, 72%, 74%, 76%, 78%, 80%, 82%, 84%, 86%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 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, ClustalX, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. In some embodiments, the guide RNA has a spacer region with a sequence having a length of from about 17 to about 53 nucleotides (nt), from about 25 to about 53 nt, from about 29 to about 53 nt or from about 40 to about 50 nt. In some embodiments, the

guide RNA has a spacer region with a sequence having a length of about 20 nt, about 21 nt, about 22 nt, about 23 nt, about 24 nt, about 25 nt, about 26 nt, about 27 nt, about 28 nt, about 29 nt, about 30 nt, about 31 nt, about 32 nt, about 33 nt, about 34 nt, about 35 nt, about 36 nt, about 37 nt, about 38 nt, about 39 nt, about 40 nt, about 41 nt, about 42 nt, about 43 nt, about 44 nt, about 45 nt, about 46 nt, about 47 nt, about 48 nt, about 49 nt, about 50 nt, or within any ranges that are made of any two or more points in the above list. In some embodiments, the guide RNA has a direct repeat region with a sequence having a length of about 15 nt, about 16 nt, about 17 nt, about 18 nt, about 19 nt, about 20 nt, about 21 nt, about 22 nt, about 23 nt, about 24 nt, about 25 nt, about 26 nt, about 27 nt, about 28 nt, about 29 nt, about 30 nt, about 31 nt, about 32 nt, about 33 nt, about 34 nt, about 35 nt, about 36 nt, about 37 nt, about 38 nt, about 39 nt, about 40 nt, about 41 nt, about 42 nt, about 43 nt, about 44 nt, about 45 nt, about 46 nt, about 47 nt, about 48 nt, about 49 nt, about 50 nt, or within any ranges that are made of any two or more points in the above list. In some embodiments, the guide RNA has a tracrRNA region having a sequence with a length of about 15 nt, about 16 nt, about 17 nt, about 18 nt, about 19 nt, about 20 nt, about 21 nt, about 22 nt, about 23 nt, about 24 nt, about 25 nt, about 26 nt, about 27 nt, about 28 nt, about 29 nt, about 30 nt, about 31 nt, about 32 nt, about 33 nt, about 34 nt, about 35 nt, about 36 nt, about 37 nt, about 38 nt, about 39 nt, about 40 nt, about 41 nt, about 42 nt, about 43 nt, about 44 nt, about 45 nt, about 46 nt, about 47 nt, about 48 nt, about 49 nt, about 50 nt, or within any ranges that are made of any two or more points in the above list. The ability of a guide sequence to direct sequence-specific binding of a target specific nuclease to a target sequence may be assessed by any suitable assay.

In some embodiments, the gRNA comprises a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43 and 61-79. For example, the sgRNA can comprise a nucleic acid sequence at least 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%, or 99% identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43 and 61-79.

Discovery of Miniature CRISPR Nucleases

A major challenge for in vivo genome engineering is the size of tools, which are prohibitive for viral delivery, especially with applications such as base editing, activation, inhibition, and HDR. The most commonly used Cas9 ortholog is Streptococcus pyogenes SpCas9, a large, 1368 amino acid length protein. Smaller CRISPR nucleases with lengths less than about 1000 amino acids can result in base editors and transcriptional activators that can fit within the 4.7 kb limit of AAV vectors. Smaller CRISPR nucleases can be discovered through metagenomic mining and innovative screening methods. Protein and guide RNA engineering can be used to boost the activity of these smaller nucleases for robust mammalian cell applications.

Cas12f and Cas12h nucleases are among the smallest DNA-targeting Cas12 families characterized to date, with Cas12f having between about 400 and about 700 residues and Cas12h having between about 870 and about 933 residues. However, these enzymes have not been engineered for high efficiency genome editing, with unquantified editing rates by Cas12f in mammalian cells and genome editing not yet demonstrated with Cas12h.

Cas12f, Cas12h and novel Cas12 systems can be mined across diverse prokaryotic genomes to identify shorter proteins. Using families of known Cas12f/h orthologs to seed hidden Markov model (HMM) alignment algorithms, NCBI and JGI databases of prokaryotic genomes and metagenomes can be searched to discovered new enzymes. The computational identification of novel miniature CRISPR nucleases from metagenomic samples is illustrated in FIG. 1A. The JGI database is particularly suitable for this search because it contains more than about 100,000 genomes and metagenomes and over about 54 billion protein coding genes, with continual rapid growth.

Single-effector CRISPR enzyme families lacking homology to classified enzymes can be found by searching for CRISPR arrays across aggregated genomes and CRISPR selecting nearby single-effector proteins, which can be putative new subtypes of Class 2 CRISPR systems. Additional sources of data from novel metagenomic sources can be used to supplement this approach, including urban-sampled metagenomes from diverse subways and microbiomes from non-western cohorts, which have been demonstrated to possess numerous additional uncharacterized genes.

CRISPR arrays as seed markers can be used to select genes within the proximity of these arrays and to develop neighborhoods of CRISPR-associated genes. HMM profiles for CRISPR-

associated proteins can be generated from the literature and these profiles can be applied to filter out known systems. All remaining genes in the dataset can be clustered with linear-time clustering algorithms, such as LinClust. To select single effectors, the co-association of different protein clusters with each other can be investigated and filtered for clusters that either associate only with CRISPR arrays, or with known CRISPR adaptation machinery such as for example, without limitation, Cas1, Cas2, and Cas4. These putative single effector clusters can then be annotated for function via HMM-based alignment to assembled pfams. Clusters can be initially selected based on the presence or similarity to known nuclease domains such as for example, without limitation, RuvC and HNH, and if they are below about 800 residues in length. These candidates can be iteratively searched in a unified dataset to guarantee that "shorter" CRISPR nucleases are not misannotated truncations of larger nucleases due to loss of coverage in sequencing or homologs of larger nucleases that were truncated and inactivated. Results from panning for small CRISPR nucleases are shown in FIGS. 1B-1D and describe in Example 1 below.

Characterization of Miniature CRISPR Nucleases

Small CRISPR nuclease systems found during computational discovery can be screened *in vitro* and *in vivo*. DNA synthesis can allow the large-scale synthesis of primers to clone gene clusters from metagenomic samples. For select candidates, the corresponding CRISPR effector gene and any accessory RNAs for testing activity can be synthesized. Although this approach can scale to tens of orthologs, complementary approaches are necessary for screening hundreds to thousands of potential orthologs for screening. Next generation DNA synthesis can allow large scale synthesis of primers to clone gene clusters from metagenomic samples. Small CRISPR nucleases can be amplified from urban sample metagenomes, either in isolation or in context of their neighboring genes and cloned into plasmids for biochemical sampling in bulk using transcription-translation (TXTL) in microfluidic droplets. Biochemical assays can profile sequence constraints or cleavage activity of the CRISPR enzymes. Profiling can enable the engineering of these qualities for subsequent use in mammalian cells.

Small CRISPR nucleases can be cloned using covalently-linked primers (Long Adapter Single-Stranded Oligonucleotide or LASSO) generated via pooled DNA synthesis, allowing cloning of hundreds of thousands of gene candidates. Because these enzymes are selected to be

small, they can easily be reconstituted in TXTL systems, allowing for rapid screening of millions of candidates in a controlled biochemical setting with no purification. When small RNAs can be expressed in TXTL system, as crRNA directionality needs to be determined for each CRISPR system, the pooled candidate library can be initially express via RNA sequencing to determine crRNA direction and processing. A second set of LASSO primers that amplify the candidate systems can then be synthesized and a synthetic CRISPR array targeting a synthetic target site can be appended on the plasmid along with a gene specific barcode. Pools of these constructs can be cloned into vectors containing the target site for the synthetic CRISPR array flanked by randomized sequences to accommodate all possible PAMs. In the TXTL system, successful cleavage events can result in a double-stranded break next to the PAM sequence, which can be captured by ligation of an adaptor. Subsequent PCR amplification can produce amplicons containing both the cleaved PAM sequence and the gene-specific barcode. Pooled sequencing of this library can reveal top candidates capable of cleavage and their corresponding sequence preferences. Additionally, the pooled TXTL assay can be performed at different timepoints to profile cleavage kinetics and select orthologs with highest activity. Once top candidates are identified, each of the enzymes can be individually cloned and the cleavage activity can be tested in individual TXTL reactions on fixed PAM targets. The candidates that are the most active and have optimal PAMs that are not too restrictive can then be confirmed.

Existing orthologs of Cas12f/h can also be screened to maximize successful identification of smaller nucleases for genome editing. This may result in issues with expression of candidate nucleases in TXTL systems. For example, base sequence biases can limit expression. If unsatisfactory results in TXTL assays are found, pooled LASSO can be used for assaying constructs heterologously in E. coli cells. Candidates can be screened targeting the synthetic guides towards a ccdB toxin plasmid with a degenerate PAM library, allowing positive selection of gene candidates with activity and facile sequencing of the candidate barcode and PAM sequence by picking surviving clones. Examples of protospacer-adjacent motif include, without limitation, NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

Guide RNA Discovery for Miniature CRISPR Nucleases

Some embodiments disclosed herein requires a gRNA comprising a tracrRNA. Small RNA sequencing studies can be performed to determine the molecular identity of the tracrRNA and associated crRNAs. However, further optimization of small RNAs is often necessary to reach levels of activity required for DNA cleavage and genome editing in mammalian cells. These designs can be informed by secondary structure algorithms to predict both optimal hybridization and tracrRNA structures with ideal hairpins for protein binding. *In vitro* cleavage assays can be performed with both panels of crRNAs carrying varying DR and spacer lengths as well as tracrRNAs with different architectures. These models can be further optimized across the design space *in silico* by progressive truncations of putative tracrRNA or crRNA and simulations of folding, resulting in an energy landscape that can be validated with *in vitro* cleavage reactions (**FIG. 6A** and **FIG. 6B**). Upon finding good candidates, crRNAs and tracrRNAs can then be combined into single-guide RNAs (sgRNAs) using a combination of potential loops and linkers to find the optimal sgRNA design. For Cas12 orthologs without tracrRNAs, crRNA designs can just be screened to find the optimal design. As an example, PsaCas12f was tested with different crRNA/tracrRNA designs as disclosed in Example 4 and **FIG. 6C**.

With optimal crRNA and sgRNA designs, mutagenesis studies can be performed to find mutations that can optimally stabilize the protein and boost cleavage activity. It was found that mutations, insertions, and deletions can drastically change the editing activity of a CRISPR enzyme. *In vitro* cleavage screens can be performed to find optimal sgRNA and crRNA mutants for efficient enzymatic activity. Top designs can then be tested in bacteria for confirmation of cellular DNA cleavage activity by these top orthologs.

Characterization of Genome Editing by Miniature CRISPR Nucleases

Miniature CRISPR nucleases can serve as a rich base for a new toolbox of easily-deliverable genome engineering tools. As their small size permits delivery with AAV, they can be used for genome editing *in vivo*. Furthermore, the additional space that is allowed by these miniature proteins can enable fusion with numerous effector domains, including transcriptional activators, repressors, and deaminases, and single vector HDR delivery (**FIG. 3A**). Miniature CRISPR nucleases can be engineered for mammalian genome editing and editing efficiency can be improved through multiple optimizations of the proteins. The small editors can be fused with

transcriptional activators to create miniature, programmable activators capable of *in vivo* delivery with AAV constructs. These miniature activators can be used to demonstrate selective gene activation to activate the Pdx1 gene *in vivo* and treat a mouse model of Type I diabetes.

Initially, a set of miniature CRISPR nucleases can be engineered, drawn from both new nucleases and previously characterized Cas12 members, to enable genome editing. The novel nucleases can be human-codon optimized and cloned into mammalian expression constructs for genome editing on luciferase reporter constructs in HEK293FT cells. In this model, indels can inactivate the luciferase gene, allowing editing efficiency to be quantified by loss of luciferase signal (**FIG. 7A**). As localization of CRISPR enzymes can be a significant factor in their efficiency, top candidates can be selected and a panel of nuclear localization signals (NLS) can be fused on either the N-terminus, the C-terminus, or both to determine the effects on editing efficiency. Localization can be further verified by tagging of constructs with small HA epitope tags, which can then be interrogated using immunofluorescence microscopy. Beyond demonstrating evidence of localization, the accessibility of these tags can provide insights into the accessibility of the N- and C-termini of the protein, which can inform the engineering of activators.

Furthermore, as sgRNA expression and localization can be different in mammalian contexts than *in vitro*, the top sgRNA designs can be compared to further tune the efficiency of editing. Flexible insertions into the sgRNA can also be engineered, and the effects on cleavage efficiency can be tested to determine potential areas where binding loops can be inserted. Constructs with high cleavage efficiency can be validated against the disease-relevant endogenous gene EMX1. For example, editing tests from PsaCas12f family members for indel generation at EMX1 were performed as disclosed in Example 5 and **FIG. 7B**. Optimization of PsaCas12f in terms of codon, optimization expression, stabilization, and localization can allow for further increases in mammalian activity.

It is essential that genome editing tools such as CRISPR nucleases are active in a variety of contexts. Once the optimized enzyme and sgRNA constructs for mammalian editing are determined, these constructs can be tested for robust editing over a panel of cell lines and additional endogenous genes TRAC, VEGF, and Pdx1. As the specificity of these enzymes is an important factor into their use, both as basic research tools as well as potential future therapies, unbiased methods for profiling genome-wide specificity can be used. The best performing

candidate can be subjected to a GUIDE-Seq genome-wide profiling pipeline. After knowing that these enzymes are effective and specific, they can be further engineered for activation-based applications.

Engineering of Miniature CRISPR Activators for Programmable Gene Activation and Inhibition

Conversion of miniature CRISPR nucleases to programmable binding platforms for applications such as editing requires catalytic inactivation. To this end, conserved catalytic residues can be mutated in the RuvC domains of these type V effectors and loss of cleavage can be tested. The maintenance of binding activity can be validated by fusing an HA tag to the effector and determining binding locations by CHIP-Seq. If binding is still maintained in these catalytically inactivated mutants, CHIP signal should correspond to locations targeted by the sgRNA. Upon validation of binding in mammalian cells, this minimal programmable binding platform can be used to develop programmable activators.

To reconstitute programmable activators from the minimal CRISPR nucleases in mammalian cells, two parallel and synergistic approaches to recruit transcriptional activators can be taken. First, sets of transcriptional activators can be fused to the effector protein at either the N- or C- terminus. These fusions can be drawn from known sets of effectors, including VP64, p65, HSF1, and RTA, and these effectors can be tested in isolation or in combination of up to three effectors. In parallel, the sgRNA can be engineered to contain MS2 hairpin loops, which can bind the MCP protein. MS2 loops can then be inserted into potential predetermined accessible areas. These loops can bind MCP-activator fusions, such as MCP-VP64 or p65. These constructs can then be tested in isolation or in combination with the fusion activators to optimize the potency of activation. In order to conserve the size of constructs and avoid the need for a second promoter, a P2A fusion linker can be used to express both the minimal CRISPR nuclease and MCP-activators from a single promoter.

Candidates for transcriptional activation can be tested on luciferase reporter constructs in HEK293FT cells with a secreted luciferase downstream of a minimal promoter. This assay can allow screening of different activator constructs in throughput over multiple rounds to determine the most active construct. Importantly, the result construct from these rounds of optimization can be selected to be small enough for packaging into AAV. The activity of these constructs can be validated on endogenous genes through RT-qPCR. As recruitment of transcriptional activators

and the resulting transcriptional machinery can be dependent on cell state, the optimal construct can be tested in a variety of cell types to guarantee robust activation *in vivo*. Lastly, the specificity of this activation system can be profiled by targeting the HBG gene in HEK293FT cells and measuring transcriptome-wide gene expression. If the activator is specific, the activation of HBG and no off-target activation should be observed. If the activator construct is specific, it can be prepared for *in vivo* delivery.

Transcriptional activators of the present disclosure may be targeted to specific target nucleic acids to induce activation/expression of the target nucleic acid. In some embodiments, the transcriptional activator polypeptide is targeted to the target nucleic acid via a heterologous DNA-binding domain. In this sense, a target nucleic acid of the present disclosure is targeted based on the particular nucleotide sequence in the target nucleic acid that is recognized by the targeting portion of the DNA-binding domain. In some embodiments, transcriptional activators activate expression of a target nucleic acid by being targeted to the nucleic acid with the assistance of a guide RNA (via CRISPR-based targeting). With CRISPR-based targeting, a target nucleic acid of the present disclosure can be targeted based on the particular nucleotide sequence in the target nucleic acid that is recognized by the targeting portion of the crRNA or guide RNA that is used according to the methods of the present disclosure.

Various types of nucleic acids may be targeted for activation of expression. The target nucleic acid may be located within the coding region of a target gene or upstream or downstream thereof. Moreover, the target nucleic acid may reside endogenously in a target gene or may be inserted into the gene, e.g., heterologous, for example, using techniques such as homologous recombination. For example, a target gene of the present disclosure can be operably linked to a control region, such as a promoter, which contains a sequence that can be recognized by e.g., a crRNA/tracrRNA and/or a guide RNA of the present disclosure such that a transcriptional activator of the present disclosure may be targeted to that sequence. In some embodiments, the target nucleic acid is not a target of and/or does not naturally associate with the naturally-occurring transcriptional activator polypeptide.

The target specific nucleases disclosed herein can be used with various CRISPR gene activation methods (see *e.g.*, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015 Jan

29;517(7536):583-8. doi: 10.1038/nature14136. Epub 2014 Dec 10. PMID: 25494202; PMCID: PMC4420636; David Bikard, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, Luciano A. Marraffini, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system, Nucleic Acids Research, Volume 41, Issue 15, 1 August 2013, Pages 7429--7437, doi.org/ 10.1093/ nar/ gkt520; Perez-Pinera, P., Kocak, D., Vockley, C. et al. RNA-guided gene activation by CRISPR-Cas9--based transcription factors. Nat Methods 10, 973--976 (2013). doi.org/10.1038/nmeth.2600; Marvin E. Tanenbaum, Luke A. Gilbert, Lei S. Qi, Jonathan S. Weissman, Ronald D. Vale, "A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging," RESOURCE| VOLUME 159, ISSUE 3, P635-646, OCTOBER 23, 2014, DOI: doi.org/ 10.1016/ j.cell.2014.09.039; Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. 2015 Jan 29;517(7536):583-8. doi: 10.1038/nature14136. Epub 2014 Dec 10. PMID: 25494202; PMCID: PMC4420636; Chavez, A., Scheiman, J., Vora, S. et al. Highly efficient Cas9-mediated transcriptional programming. Nat. Methods 12, 326--328 (2015). doi.org/ 10.1038/ nmeth.3312; Chavez, A., Tuttle, M., Pruitt, B. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods* **13,** 563--567 (2016). doi.org/ 10.1038/ nmeth.3871; and Sajwan, S., Mannervik, M. Gene activation by dCas9-CBP and the SAM system differ in target preference. *Sci Rep* **9,** 18104 (2019). doi.org/ 10.1038/ s41598-019-54179-x, which are incorporated herein by reference in their entirety).

Examples of CRISPR gene activation methods include, without limitation, dCas9-CBP CRISPR gene activation method, SPH CRISPR gene activation method, Synergistic Activation Mediator (SAM) CRISPR gene activation method, Sun Tag CRISPR gene activation method, VPR CRISPR gene activation method, and any alternative CRISPR gene activation methods therein. The dCas9-VP64 CRISPR gene activation method uses a nuclease lacking endonuclease ability and fused with VP64, a strong transcriptional activation domain. Guided by the nuclease, VP64 recruits transcriptional machinery to specific sequences, causing targeted gene regulation. This can be used to activate transcription during either initiation or elongation, depending on which sequence is targeted. The SAM CRISPR gene activation method uses engineered sgRNAs to increase transcription, which is done through creating a nuclease/VP64 fusion protein engineered with aptamers that bind to MS2 proteins. These MS2 proteins then recruit additional

activation domains (HS1 and p65) to then activate genes. The Sun Tag CRISPR gene activation method uses, instead of a single copy of VP64 per each nuclease, a repeating peptide array to fused with multiple copies of VP64. By having multiple copies of VP64 at each loci of interest, this allows more transcriptional machinery to be recruited per targeted gene. The VPR CRISPR gene activation method uses a fused tripartite complex with a nuclease to activate transcription. This complex consists of the VP64 activator used in other CRISPR activation methods, as well as two other potent transcriptional activators (p65 and Rta). These transcriptional activators work in tandem to recruit transcription factors.

The target specific nucleases disclosed herein can be used as base editors for base editing (see e.g., Anzalone, A.V., Koblan, L.W. & Liu, D.R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat Biotechnol* **38,** 824–844 (2020), which is incorporated herein by reference in its entirety). There are generally three classes of base editors: cytosine base editors (CBEs), adenine base editors (ABEs), and dual-deaminase editor (also called SPACE, synchronous programmable adenine and cytosine editor). Base editing requires a nickase or nuclease fused or coupled to a deaminase that makes the edit, a gRNA targeting the nuclease to a specific locus, and a target base for editing within the editing window specified by the nuclease.

Cytosine base editors (CBEs) uses a cytidine deaminase coupled with an inactive nuclease. These fusions convert cytosine to uracil without cutting DNA. Uracil is then subsequently converted to thymine through DNA replication or repair. Fusing an inhibitor of uracil DNA glycosylase (UGI) to a nuclease prevents base excision repair which changes the U back to a C mutation. To increase base editing efficiency, the cell can be forced to use the deaminated DNA strand as a template by using a nuclease nickase, instead of a nuclease. The resulting editor can nick the unmodified DNA strand so that it appears "newly synthesized" to the cell. Thus, the cell repairs the DNA using the U-containing strand as a template, copying the base edit.

Adenine base editors (ABEs) can convert adenine to inosine, resulting in an A to G change. Creating an adenine base editor requires an additional step because there are no known DNA adenine deaminases. Directed evolution can be used to create one from the RNA adenine deaminase TadA. While cytosine base editors often produce a mixed population of edits, some ABEs do not display significant A to non-G conversion at target loci. The removal of inosine

from DNA is likely infrequent, thus preventing the induction of base excision repair. In terms of off-target effects, ABEs also generally compare favorably to other methods.

Suitable target nucleic acids will be readily apparent to one of skill in the art depending on the particular need or outcome. The target nucleic acid may be in a region of euchromatin (e.g., highly expressed gene), or the target nucleic acid may be in a region of heterochromatin (e.g., centromere DNA). Use of transcriptional activators according to the methods described herein to induce transcriptional activation in a region of heterochromatin or other highly methylated region of a plant genome may be especially useful in certain embodiments. A target nucleic acid of the present disclosure may be methylated, or it may be unmethylated.

The target gene can be any target gene used and/or known in the art. Exemplary target genes include, without limitation, Pdx1 and any variants thereof.

Delivery of Miniature CRISPR Nucleases

In some embodiments, the target specific nuclease and/or peptide sequence are introduced into a cell as a nucleic acid encoding each protein. The nucleic acid introduced into the eukaryotic cell is a plasmid DNA or viral vector. In some embodiments, the target specific nuclease and/or peptide sequence are introduced into a cell via a ribonucleoprotein (RNP).

Delivery is in the form of a vector which may be a viral vector, such as a lenti- or baculo- or adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. The viral vector may be selected from a variety of families/genera of viruses, including, but not limited to Myoviridae, Siphoviridae, Podoviridae, Corticoviridae, Lipothrixviridae, Poxviridae, Iridoviridae, Adenoviridae, Polyomaviridae, Papillomaviridae, Mimiviridae, Pandoravirusa, Salterprovirusa, Inoviridae, Microviridae, Parvoviridae, Circoviridae, Hepadnaviridae, Caulimoviridae, Retroviridae, Cystoviridae, Reoviridae, Birnaviridae, Totiviridae, Partitiviridae, Filoviridae, Orthomyxoviridae, Deltavirusa, Leviviridae, Picornaviridae, Marnaviridae, Secoviridae, Potyviridae, Caliciviridae, Hepeviridae, Astroviridae, Nodaviridae, Tetraviridae, Luteoviridae, Tombusviridae, Coronaviridae, Arteriviridae, Flaviviridae, Togaviridae, Virgaviridae, Bromoviridae, Tymoviridae, Alphaflexiviridae, Sobemovirusa, or Idaeovirusa.

A vector may mean not only a viral or yeast system (for instance, where the nucleic acids of interest may be operably linked to and under the control of (in terms of expression, such as to ultimately provide a processed RNA) a promoter), but also direct delivery of nucleic acids into a host cell. For example, baculoviruses may be used for expression in insect cells. These insect cells may, in turn be useful for producing large quantities of further vectors, such as AAV or lentivirus adapted for delivery of the present invention. Also envisaged is a method of delivering the target specific nuclease and/or peptide sequence comprising delivering to a cell mRNAs encoding each.

One of the values of miniature transcriptional activators is their capacity to be packaged in AAV. To this end, the optimal activators that are discovered can be cloned into AAV packaging vectors, and AAV2 containing the minimal activator can be purified. The activity of these AAV can be confirmed by delivery to HepG2 cells to confirm both liver targeting and activity. If titering or expression is found to be low, various liver-specific promoters can be tested, including the albumin and TBG promoters, to find minimal promoters with high expression to optimize delivery.

After confirming the delivery of the minimal construct in cell culture, expression in mice by hydrodynamic injection of promoter-less luciferase constructs can be assessed and followed by the tail-vein injection of minimal activator-AAV targeting the upstream region of these luciferase constructs. Luciferase expression can only be induced in the liver in the presence of successful activation, which can be measured by bioluminescence imaging.

To test the activation in a less perturbative model, Pdx1 can be activated. Pdx1 is a target of *in vivo* activation that had been performed with Cas9 activators in a Cas9-mouse model (*see* PMC5732045). Pdx1 overexpression in the liver can transdifferentiate hepatic cells *in vivo* to generate insulin-secreting cells. Pdx1 activation can be tested in cell culture using Hepa1-6 cells and expression can be measured by RT-qPCR to determine the optimal guide. These optimal Pdx1-targeting guides can be injected into mice via tail vein injection. These mice can be harvested 2 weeks post-injection to determine changes in Pdx1 expression as well as genes downstream from Pdx1 such as for example, without limitation, insulin and Pcsk1. To validate the phenotypic effects of Pdx1 targeting, mice can be treated with streptozotocin to produce hyperglycemia. The introduction of the Pdx1 activators can be tested to determine it can reduce

blood glucose levels and increase serum insulin, as it has been found for Cas9 activators in a Cas9-mouse model.

Combinations of transcriptional activators can lead to successful activation. However, these combinations can be too large. If this is the case, activators can be truncated to find essential domains that allow for activation but have reduced size. Truncation of the guide RNA to modulate binding of novel Cas effectors and to quantitatively tune gene activation can be also assessed.

In some embodiments, expression of a nucleic acid sequence encoding the target specific nuclease and/or peptide sequence may be driven by a promoter. In some embodiments, the target specific nuclease is a Cas. In some embodiments, a single promoter drives expression of a nucleic acid sequence encoding a Cas and one or more of the guide sequences. In some embodiments, the Cas and guide sequence(s) are operably linked to and expressed from the same promoter. In some embodiments, the CRISPR enzyme and guide sequence(s) are expressed from different promoters. For example, the promoter(s) can be, but are not limited to, a UBC promoter, a PGK promoter, an EF1A promoter, a CMV promoter, an EFS promoter, a SV40 promoter, and a TRE promoter. The promoter may be a weak or a strong promoter. The promoter may be a constitutive promoter or an inducible promoter. In some embodiments, the promoter can also be an AAV ITR, and can be advantageous for eliminating the need for an additional promoter element, which can take up space in the vector. The additional space freed up by use of an AAV ITR can be used to drive the expression of additional elements, such as guide sequences. In some embodiments, the promoter may be a tissue specific promoter.

In some embodiments, an enzyme coding sequence encoding a target specific nuclease and/or peptide sequence is codon-optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, mouse, rat, rabbit, dog, or non-human primate. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms)

often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database", and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, Pa.), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas protein correspond to the most frequently used codon for a particular amino acid.

In some embodiments, a vector encodes a target specific nuclease and/or peptide sequence comprising one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the Cas protein comprises about or more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., one or more NLS at the amino-terminus and one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Typically, an NLS consists of one or more short sequences of positively charged lysines or arginines exposed on the protein surface, bur other types of NLS are known. In some embodiments, the NLS is between two domains, for example between the Cas12 protein and the viral protein. The NLS may also be between two functional domains separated or flanked by a glycine-serine linker.

In general, the one or more NLSs are of sufficient strength to drive accumulation of the target specific nuclease and/or peptide sequence in a detectable amount in the nucleus of a

eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the target specific nuclease and/or other peptide sequences, the particular NLS used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the target specific nuclease and/or peptide sequence, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Examples of detectable markers include fluorescent proteins (such as green fluorescent proteins, or GFP; RFP; CFP), and epitope tags (HA tag, FLAG tag, SNAP tag). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly.

In some respects, the invention provides methods comprising delivering one or more polynucleotides, such as one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some respects, the invention further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a Cas protein in combination with (and optionally complexed) with a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding a target specific nuclease and/or a blunting enzyme to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g., a transcript of a vector described herein), naked nucleic acid, nucleic acid complexed with a delivery vehicle, such as a liposome, and ribonucleoprotein. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-8313 (1992); Navel and Felgner, TIBTECH 11:211-217 (1993); Mitani and Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer and Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and Bohm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

The target specific nuclease and/or peptide sequence can be delivered using adeno-associated virus (AAV), lentivirus, adenovirus, or other viral vector types, or combinations thereof. In some embodiments, Cas protein(s) and one or more guide RNAs can be packaged into one or more viral vectors. In some embodiments, the targeted trans-splicing system is delivered via AAV as a split intein system, similar to Levy et al. (Nature Biomedical Engineering, 2020, DOI: doi.org/10.1038/s41551-019-0501-5). In other embodiments, the target specific nuclease and/or peptide sequence can be delivered via AAV as a trans-splicing system, similar to Lai et al. (Nature Biotechnology, 2005, DOI: 10.1038/nbt1153). In some embodiments, the viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the viral delivery is via intravenous, transdermal, intranasal, oral, mucosal, intrathecal, intracranial or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector chosen, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo), or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues. Viral-mediated in vivo delivery of Cas13 and guide RNA provides a rapid and powerful technology for achieving precise mRNA perturbations within cells, especially in post-mitotic cells and tissues.

In certain embodiments, delivery of the target specific nuclease and/or peptide sequence to a cell is non-viral. In certain embodiments, the non-viral delivery system is selected from a ribonucleoprotein, cationic lipid vehicle, electroporation, nucleofection, calcium phosphate

transfection, transfection through membrane disruption using mechanical shear forces, mechanical transfection, and nanoparticle delivery.

In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, VA). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences.

Diagnostics

The present disclosures provide target specific nucleases for diagnostic applications. The diagnostic applications include for example and without limitation molecular, amino acid, nucleic acid, and derivatives thereof diagnostics (see e.g., Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, Cofsky JC, Kyrpides NC, Banfield JF, Doudna JA. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. Science. 2018 Nov 16;362(6416):839-842. doi: 10.1126/science.aav4294. Epub 2018 Oct 18. PMID: 30337455; PMCID: PMC6659742; and Xiang X, Qian K, Zhang Z, Lin F, Xie Y, Liu Y, Yang Z. CRISPR-cas systems based molecular diagnostic tool for infectious diseases and emerging 2019 novel coronavirus (COVID-19) pneumonia. J Drug Target. 2020 Aug-Sep;28(7-8):727-731. doi: 10.1080/1061186X.2020.1769637. Epub 2020 May 26. PMID: 32401064; PMCID: PMC7265108, which are incorporated herein by reference in their entirety). In one example, the target specific nuclease can be used with DETECTR, a DNA endonuclease-targeted CRISPR trans reporter technology for molecular diagnostics. This technique achieves high sensitivity for DNA detection by combining the activation of non-specific single-stranded deoxyribonuclease of Cas12 ssDNase with isothermal amplification that enables fast and specific detection of biologicals such as viruses. In this assay, a crRNA-Cas12a complex binds to a target DNA and induces an indiscriminate cleavage of ssDNA that is coupled to a fluorescent reporter. In another example, the target specific nuclease can be combined with a fluorescence-based point-of-care (POC) device. In this example, Cas12a/crRNA detects and binds to a targeting DNA, the

Cas12a/crRNA/DNA complex then becomes activated and degrades a fluorescent ssDNA reporter to generate a signal.

Kits

The present disclosure provides kits for carrying out a method. The present disclosure provides the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the kit comprises a vector system comprising regulatory elements and polynucleotides encoding the target specific nuclease and/or peptide sequence. In some embodiments, the kit comprises a viral delivery system of the target specific nuclease and/or peptide sequence. In some embodiments, the kit comprises a non-viral delivery system of the target specific nuclease and/or peptide sequence. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kit includes instruction in one or more languages, for examples, in more than one language.

In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g., in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide sequence and a regulatory element.

Sequences

Sequences of target specific nucleases, guides, and nuclear localization signal (NLS) can be found in Table 1 below.

TABLES

Table 1

SEQ ID NO / DESCRIPTION / SOURCE	SEQUENCE
SEQ ID NO: 1 PsaCas12f (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDK
SEQ ID NO: 2 Cas12f ortholog 160429_1003 (Artificial sequence)	MEEENFDNAEVTTGIKFKLKLNSETREKLNNYFNEYGKAINFAVRIIQKQLAD DRFAGKAKLDENKKQLLDEDGKKIWDFPSESCSCGKQVVRYVNGKPFCQEC YRNKFSENGIRKRMYSAKGRKAEYDINIKNSTNRISKTHFNYAIREAFILDKSI KKQRKERFRRLNDMMRKLQEFIDIREGKRLVCPKIERQKVERYIHPAWINKE KKIEEFRGYSLSVVNSKIKALDRNIKREEKSLKEKGQINFKARRLMLDKSVKF TDTNKVSFTISKSLPKEYELDLPKKEKRLNWLKEKIEIIKNQKPKYAYLLRRG DDFYLQYTLQTKPEIKTTHSGAVGIDRGISHIAVYTFVSNDGKNERPLFLSSSE ILRLKNLQKERDKFLRRKHNKIRKKSNMRNIEDKIQLILHNYSKQIVDFAKEK NAFIVFEKLEKPKKSRSKMSKKEQYKLSLFTFKKLSDLVDYKAKREGIKVIYI EPAYTSKECSHCGEKVNTQRPFNGNYSLFKCNKCGIILNSDYNASINIAKKGL NIFNI
SEQ ID NO: 3 Cas12f ortholog 176283_308 (Artificial sequence)	MAKGEKNNDVLYRAVKFEIRPTLNQETILQRISSNLRLIWNEAWKERQDRYE IFFKPIYERIYNAKKKALEKGFTDLWEKEVAKFSQQVLVKRGFPLQLVLEQKS LFAELKKAFEEHGITLYDQINALTAKRSLNTEFGLIPRNWQEETLDALDGSFK SFFALRKRGDKDAKPPSERTNEDSFYKIPGRSGFKVTDDGKVIVSFGKLSETL VGRIPEYQQEKLSHAKNLKKFEIVRDERDMAKSGCFWISIAYEIPKPPELPFNP SKAVFLAIGASWIGHSPRGEFCWRMPRPDFHWKPKINAVDERLKRVSKGSIK WKRLIFARSKMFAIMARQQKQHGQYEVIKRLLELGVCFVVTDLKVRSKEGS LADSSKAERGGSPFGANWSAQNTGNIANLVAKLTDHVSALGGMVIKRKSPE LLVEEKRLPQEKRKILLAQKLKDEFLSLN
SEQ ID NO: 4 Cas12f ortholog 176287_13 (Artificial sequence)	MKNTKEEKWMQTYCFDLTDEEFGAENIRLATHISDSLVPLFNEVLLQVLKGD ETIKELKQEVKLRGRALKQKAKEAQMEDLWDRENNEIDDEEWLERGYDQE VIKEHRDYVDEIAKLYSENKVTAFDHNYHYAQENLEAIGCTAPYVNISAGLR RGAIKNCHGAVDSWRKHLATGDYKSKPPGQQEVGKFYMLRCEPGCAVTKD RKNVRISLGDRKSSPVFELPGLDNSKNKPLHMLLRSDAKVKSFTLSRRSARNP DKKESDLQKPGVWRISINFELPLPEKKPATEYNTVALVIGSNYLGVALHDSER NFPLNLPLPHKHWFPIIGDIEGRANVPWRKKGSKKWRRKMFGVQKKHSGGR QACYRYMARQQKQGEYETIADHLIGCGVHFVVSKPSINHPKGLADASAPDR GGDTGPNRIISSTGVNSLVLKLKQKVKEFGGSVTEMEAPPLPERFRFWDSGPK KVIVAQLLRNQYLAQKK
SEQ ID NO: 5 Cas12f ortholog 209659_1510	MVKQTTFFCKECNKNINIPRNIIKKLESNHISQDQAIKKAKERHNKKKHSLILG IKFKLYVKNKEDKEKLNSYFEEYAKAVTFAAQIIDKIKSGYLPQWKKDKKLK RIIFPKGKCDFCGTKTEIGWISKRGKKICKNCYSKEYGENGIRKKLYATRGRK VNPSYNIFNATKKLAATHYNYAIREAFQLLEANRKQRQERIRRLLRDKKRLR

(Artificial sequence)	EFEDLIEKPDRRIELPMKTRQREKRYIHISQKDKINELRGYTLHKIKEKIRILRR NTEREERALRKKTPIIFKGNRIMLFPQGIKFDKENNKVKITIAKNLPKEFIFSGT NVANKHGRRFFKEKLNLISQQKPKYAYLIRKQTKNSKKITDYDYYLQYTIET VYKIRKNYDGIIGIDRGINNLACLVLLEKNQEKPCGVKFYKGKEINALKIKRR KQLYFLRRKHNRKQKQKRIRRIEPKINQILHIISKEIVELAKEKNFAIGLEQLEK PKKSRFRQRRKERYFLSLFNFKTLSTFIEYKAKKEGIRVIYIPPERTSQICSHCAI KGDVHTNTIRPYRKPNAKKSSSSLFKCKKCGVELNADYNAAFNIAQKSLKIL ST
SEQ ID NO: 6 Cas12f ortholog 213082_2246 (Artificial sequence)	MKIKEQSEVRELLKAYKYRIYPNKEQRLYLAKTFGCTRFIYNKMLSDRIKVY EENKDLDIKKVKYPTPAQYKKEFTWLKEVDSLALANAQMNLDKAYKNFFR DKSMGFPKFKSKKVNYYSYTTNNQKGTVYIEDGYIKLPKLKTMIKIKQHRKF NGLIKSCTISKTPSNKYYISILVYTENKQLPKVDKKVGIDVGLKEFAITSNGEF FSNPKWLRKSEKRLRKLQKDLSRKQKGSNNRCKARLKVAKLHEKITNQRKN FLHKLSIKLIRENQSIVIEDLKVKNMLQNHKLAKAISEVSWYEFRTMLEYKAD WYGRELIIAPSNYASSQICSNCGYKNKEVKNLELREWVCPKCGIHHHRDINA SKNLLKLAI
SEQ ID NO: 7 Cas12f ortholog 238436_2949 (Artificial sequence)	MLVFEAKLRGTKEQYERLDEAIRTARFVRNSCLRYWMDNKGEKVGRYELS AYCAVLAKEPPWAKKLNSMARQASAERAWTAIARFYDNCKKKVSGKKGFP KFKKYKTRDSVEYKTSGWKLSEDRRTITFTDGFKAGSFKTWGTRDLHFYQL KQIKRVRVVRRADGYYVQFCIDQDRVEKREPTGTAIGLDVGLNHFYTDSDG QTVENPRHLRKSEKALKRLQRRLAKTQKGSKNRQKARNRLGRKHLKVSRQ RKDFAVKTALCVVQSNDLVAYEDLKVRNMVKNHNLAKSISDAAWSTFRQW MEYFGKVFGVATVAVPPQYTSQNCSNCGEKIQKSLSTRTHRCPHCGFVADR DHNAAINILELGLSTVGHTETHASGDIDLCLGGETPQSKSSRRKRKPHQ
SEQ ID NO: 8 Cas12f ortholog 265253_1259 (Artificial sequence)	MDQHKGVKLRLYPNRGQKDKLWQMFGNDRFVWNQMLSMAKTRYQNNPR ASFINGYGMDTLLKVLKNEYPFLKESDSTSLQVVNHKLNQSFQMLFKHRGG YPRFKSRKATKQAYTGKSKVSVVAKRCLKLPKIGYIKTSKTNQLVDTKIKRY TVSYDATGRYYLSLQVEVPAPELLPKTGKVVGLDVGLADLAISSDGVKYGTF NAKWLDKQVNKWQSAYAKRKYRATIAVRQWNHNHKTVKEELNDYQNWQ RARRYKARYQAKVANKRQDNLQKLTTELVKQYDVIVIEDLKTKNLQKNHH LAKSIANASWYQLRTMLEYKCAWYGRQLIIVKPNYTSQICSSCGYHNGPKPL KIREWTCSKCGVHHDRDINAAINILHKGLKANG
SEQ ID NO: 9 Cas12f ortholog 325997_390 (Artificial sequence)	MTSNKCAEEGQKKVSVTPITFNFWLTKVKDRIFELEDQTTVLLKDVSVDLSR QVLKMLAGAWQSYFELRKRGDTEARPPSPKKEGWFQTMAWSNFTVRQGSIF VPGYQKNRIEIKLGDYLKRMVEDKEVAYVTLYRDRFSGEFNLSVVVKNPAP KHIEHPKVIRAIDLGAGDIAVSDSSGAEYLIPARRPDKHWMPLIAQVEHRAER CIKGSRAYKRRMKARRVMHEKSGNQKDSYQRKLARALFSGEVEAIVIGKGK TRLGLAQSESGTPDQHYGAQNTGYLFRQLLYIKEKAKERGIPVVEFPDPQRK GELEDSQKKFFASRELLSLGCKKFKIEVPNSFVQGEFIFNQGKGGKPKVA
SEQ ID NO: 10 Cas12m ortholog 58610_1188_protein_locus_of_contig_LFOD01000003_-_Query_protein_(58610_1188)_translation_(5) Protein locus genbank annotated by CrisprCasFinder for protein 58610_1188 from file 58610 (Artificial sequence)	MAITVHTAGVHYRWTDNPPEQLMRQLRLAHDLREDLVTLQLDYETAKAGI WSSYPAVAAAETELADAESAAEQAAAAVSEERTKLRTKRITGPLAQKLTAA RKRVREARSTRRAAISEVHEEAKGRLVDASDALKAQQKALYKTYCQDGDLF WATFNDVLDHHKAAVKRIGQMRAAGQPAQLRHHRFDGTGSIAVQLQRQAG QPQRTPELIADVDGKYGRVLSVPWVQPDRWERIPRRERRMIGRVTVRMRAG QLSGEPQWLDIPVQQHRMLPLDADITGARLTVTRTAGTLRAQISVTAKIPDPE PVTDGPDVAVHLGWRNTDTGVRVARWRSTEPIEVPFDFRDTLTVDPGGRSG EIFVPEAVPRRVERAHLIASHRADRMNELRARLVDYLAETGPRPHPSREGEEL GAGNVRMWKSPNRFAWLARVWADDESVSTDIREALAQWRHQDWISWHHQ EGGRRRSAAQRLDVYRQVAAVLVSQAGRLVLDDTSYADIAQRSATTKTEEL PNETAARINRRRAHAAPGELRQTLVAAADRDAVPVDTVSHTGVSVVHAKCG HENPSDGRFMSVVVACDGCGEKYDQDESALTHMLTRAVQSAA

SEQ ID NO: 11 Cas12m ortholog 63461_4106_protein_locus_of_contig_LSKL01000323_-_Query_protein_(63461_4106)_translation_(4) Protein locus genbank annotated by CrisprCasFinder for protein 63461_4106 from file 63461 (Artificial sequence)	MTTMTVHTMGVHYKWQIPEVLRQQLWLAHNLREDLVSLQLAYDDDLKAIW SSYPDVAQAEDTMAAAEADAVALSERVKQARIEARSKKISTELTQQLRDAKK RLKDARQARRDAIAVVKDDAAERRKARSDQLAADQKALYGQYCRDGDLY WASFNTVLDHHKTAVKRIAAQRASGKPATLRHHRFDGSGTIAVQLQRQAGA PPRTPMVLADEAGKYRNVLHIPGWTDPDVWEQMTRSQCRQSGRVTVRMRC GSTDGQPQWIDLPVQVHRWLPADADITGAELVVTRVAGIYRAKLCVTARIGD TEPVTSGPTVALHLGWRSTEEGTAVATWRSDAPLDIPFGLRTVMRVDAAGTS GHVVPATIERRLTRTENIASSRSLALDALRDKVVGWLSDNDAPTYRDAPLEA ATVKQWKSPQRFASLAHAWKDNGTEISDILWAWFSLDRKQWAQQENGRRK ALGHRDDLYRQIAAVISDQAGHVLVDDTSVAELSARAMERTELPTEVQQKID RRRDHAAPGGLRASVVAAMTRDGVPVTIVAAADFTRTHSRCGHVNPADDR YLSNPVRCDGCGAMYDQDRSFVTLMLRAATAPSNP
SEQ ID NO: 12 Cas12m ortholog 21566_3969_protein_locus_of_contig_BAFB01000202_-_Query_protein_(21566_3969)_translation_(4) Protein locus genbank annotated by CrisprCasFinder for protein 21566_3969 from file 21566 (Artificial sequence)	MPDQLTQQLRLAHDLREDLVTLEYEYEDAVKAVWSSYPAVAALEAQVAEL DERASELASTVKEEKSRQRTKRPSHPAVAQLAETRAQLKAAKASRREAIASV RDEATERLRTISDERYAAQKQLYRDYCTDGLLYWATFNAVLDHHKTAVKRI AAHRKQGRAAQLRHHRWDGTGTISVQLQRQATDPARTPAIIADADTGKWRS SLIVPWVNPDVWDTMDRASRRKAGRVVIRMRCGSSRNPDGTKTSEWIDVPV QQHRMLPADADITAAQLTVRREGADLRATIGITAKIPDQGEVDEGPTIAVHLG WRSSDHGTVVATWRSTEPLDIPETLRGVITTQSAERTVGSIVVPHRIEQRVHH HATVASHRDLAVDSIRDTLVAWLTEHGPQPHPYDGDPITAASVQRWKAPRR FAWLALQWRDTPPPEGADIAETLEAWRRADKKLWLESEHGRGRALRHRTDL HRQVAAYFAGVAGRIVVDDSDIAQIAGTAKHSELLTDVDRQIARRRAIAAPG MLRAAIVAAATRDEVPTTTVSHTGLSRVHAACGHENPADDRYLMQPVLCDG CGRTYDTDLSATILMLQRASAATSN
SEQ ID NO: 13 Cas12m ortholog 633299_527_protein_locus_of_contig_Scfld15_-_Query_protein_(633299_527)_(4) Protein locus genbank annotated by CrisprCasFinder for protein 633299_527 from file 633299 (Artificial sequence)	MLRAYKYRIYPTDEQKVLFAKTFGCCRFVYNWALNLKITAYKERKETLGNV YLTNLMKSELKVEHEWLSEVNSQSLQSSLRNLDTAYTNFFRNTKAVGFPRFK SRKDKQSFLCPQHCRVDFEKGTITIPKAKDIPAVLHRRFKGTVKTVTISMTPSG RYFASVLVDTSMQEMKPSEPMRDTTVGIDLGIKSLAVCSDGRTFANPKNLQR SLDRLKLLQKRLSRKQKGSANRNKARIRVARLQEHIANSRKDSLHKITHALT HDSQVRTICMEDLNVKGMQRNHHLAQAVGDASFGMFLTLLEYKCSWYGVN LIKIDRFAPSSKTCGKCGHVYKGLNLSERSWTCPECGTHHDRDFNAACNIKEF GLKALPTERGKVKPVDCPLVDDRPRVLKSNGRKKQEKRGGIGISEAAKSLV
SEQ ID NO: 14 Cas12m ortholog	PQGIKFDKENNKVKITIAKNLPKEFIFSGTNVANKHGRRFFKEKLNLISQQKPK YAYLIRKQTKNSKKITDYDYYLQYTIETVYKIRKNYDGIIGIDRGINNLACLVL LEKNQEKPCGVKFYKGKEINALKIKRRKQLYFLRRKHNRKQKQKRIRRIEPKI NQILHIISKEIVELAKEKNFAIGLEQLEKPKKSRFRQRRKERYFLSLFNFKTLST

209658_13971_protein_locus_of_contig_Ga0190333_1001561_-_Query_protein_(209658_13971)_(2) Protein locus genbank annotated by CrisprCasFinder for protein 209658_13971 from file 209658 (Artificial sequence)	FIEYKAKKEGIRVIYIPPERTSQICSHCAIKGDVHTNTIRPYRKPNAKKSSSSLFKCKKCGVELNADYNAAFNIAQKSLKILST
SEQ ID NO: 15 Cas12m ortholog 209657_57738_protein_locus_of_contig_Ga0190332_1015597_-_Query_protein_(209657_57738)_(2) Protein locus genbank annotated by CrisprCasFinder for protein 209657_57738 from file 209657 (Artificial sequence)	DRGINNLACLVLLEKNQEKPCGVKFYKGKEINALKIKRRKQLYFLRRKHNRKQKQKRIRRIEPKINQILHIISKEIVELAKEKNFAIGLEQLEKPKKSRFRQRRKERYFLSLFNFKTLSTFIEYKAKKEGIRVIYIPPERTSQICSHCAIKGDVHTNTIRPYRKPNAKKSSSSLFKCKKCGVELNADYNAAFNIAQKSLKILST
SEQ ID NO: 16 Cas12m ortholog 209660_51257_protein_locus_of_contig_Ga0190335_1015156_-_Query_protein_(209660_51257)_(2) Protein locus genbank annotated by CrisprCasFinder for protein 209660_51257 from file 209660 (Artificial sequence)	LLEKNQEKPCGVKFYKGKEINALKIKRRKQLYFLRRKHNRKQKQKRIRRIEPKINQILHIISKEIVELAKEKNFAIGLEQLEKPKKSRFRQRRKERYFLSLFNFKTLSTFIEYKAKKEGIRVIYIPPERTSQICSHCAIKGDVHTNTIRPYRKPNAKKSSSSLFKCKKCGVELNADYNAAFNIAQKSLKILST
SEQ ID NO: 17 Cas12m ortholog 466065_250_protein_locus_of_contig_SFKR01000004.1_-_Query_protein_(4660	MEYSYKFRVYPTAAQAEQIQRTFGCCRFVWNHYLALRKDLYEQDGKTMNYNACSGDMTQLKKTLLWLREVDATALQSSLRDLDTAYQNFFRRVKKGEKPGYPKFKSKHHSKKSYKSKCVGTNIKVLDKAVQLPKLGLVKCRISKEVKGRILSATISQNPSGKYFVAICCTDVELEPLTSTGAVAGIDMGLKAFAITSDGVEYPNHKYLTKSQKKLAKLQRQLSRKSKGSKRRREKARIQVARLHEHVANQRQDMLHKLSTDLVRNYDLIAIEDLAPSNMVKNHMLAKAISDASWGEFPRQLKYKAEWHGKKVVTVGRFFPSSQLCSNCGAQWSGTKDLSVRQWTCPVCGAIHDRDMNAARNILNEGLRLMA

65_250) Protein locus genbank annotated by CrisprCasFinder for protein 466065_250 from file 466065 (Artificial sequence)	
SEQ ID NO: 18 Cas12m ortholog 8971_2857_protein_locus_of_contig_OEJQ01000083.1_-_Query_protein_(8971_2857) Protein locus genbank annotated by CrisprCasFinder for protein 8971_2857 from file 8971 (Artificial sequence)	VYNYFLSQRKEQYRLTGKSDNYYAQAKTLTALKKQEETAWLKEVNAQTLQFAIKSLESAYTNFFKKSAKFPKFKSKHSKNSFTVPQSASVAGGRLFIPKFTEGIKCSVHREIKGKIGKVTITKSPSGKYFVSVFTEEEYITQLEKTGKSIGLDMGLKDLLITSEGEIFNNNRYTRRYECKLAKAQRHLSRKKKGSRGFENQRLKVARLHEKIVNSRTDYLHKCSISLVRRYDIICIEDLNVKGMTKNHHLAKSITDASWGKFVSMLTYKAEWNNKKVVDVDRYFPSSQTCNVCGYVNKQIKDLSVREWECPHCHTHHDRDKNAAINILRIGLNNNISAGTVDYTGGEEVRTDLLESHSSVKPEANEPLVHG
SEQ ID NO: 19 Cas12m ortholog 9265_901_protein_locus_of_contig_OEFX01000005.1_-_Query_protein_(9265_901) Protein locus genbank annotated by CrisprCasFinder for protein 9265_901 from file 9265 (Artificial sequence)	MLAKHFGCSRFVYNYFLSQRKEQYRLTGKSDNYYAQAKTLTALKKQEETAWLKEVNAQTLQFAIKSLESAYTNFFKKSAKFPKFKSKHSKNSFTVPQSASVAGGRLFIPKFTEGIKCSVHREIKGKIGKVTITKSPSGKYFVSVFTEEEYITQLEKTGKSIGLDMGLKDLLITSEGEIFNNNRYTRRYECKLAKAQRHLSRKKKGSRGFENQRLKVARLHEKIVNSRTDYLHKCSISLVRRYDIICIEDLNVKGMTKNHHLAKSITDASWGKFVSMLTYKAEWNNKKVVDVDRYFPSSQTCNVCGYVNKQIKDLSVREWECPHCHTHHDRDKNAAINILRIGLNNNISAGTVDYTGGEEVRTDLLESHSSVKPEANEPLVHG
SEQ ID NO: 20 sgRNA 1 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 21 sgRNA 2 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 22 sgRNA 3	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGAATGCAAC

(Artificial sequence)	
SEQ ID NO: 23 sgRNA 4 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCT GAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATC CTTACCTATTGAAATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 24 sgRNA 5 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAC CCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 25 sgRNA 6 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAG TAATAGGTCAAGGAATGCAAC
SEQ ID NO: 26 sgRNA 7 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAA TAGGTCAAGGAATGCAAC
SEQ ID NO: 27 sgRNA 8 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATA ATAGGTCAAGGAATGCAAC
SEQ ID NO: 28 sgRNA 9 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATA GGTCAAGGAATGCAAC
SEQ ID NO: 29 sgRNA 10 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAA GGAATGCAAC
SEQ ID NO: 30 sgRNA 11 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGA ATGCAAC
SEQ ID NO: 31 sgRNA 1	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGG AATGCAAC

(Artificial sequence)	
SEQ ID NO: 32 sgRNA 13 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTG CCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAGGT CAAGGAATGCAAC
SEQ ID NO: 33 sgRNA 14 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTG CCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGA ATGCAAC
SEQ ID NO: 34 sgRNA 15 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTT GCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGAAT GCAAC
SEQ ID NO: 35 sgRNA 16 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTG CCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGGAA TTGCAAC
SEQ ID NO: 36 sgRNA 17 (Artificial sequence)	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTC TGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAG GTCAAGGAATGCAAC
SEQ ID NO: 37 sgRNA 18 (Artificial sequence)	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTC TGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAG GAATGCAAC
SEQ ID NO: 38 sgRNA 19 (Artificial sequence)	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTC TGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGAA TGCAAC
SEQ ID NO: 39 sgRNA 20 (Artificial sequence)	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTC TGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGGA ATGCAAC
SEQ ID NO: 40 sgRNA 21 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGA AAACCCAAAGTAATAGGTCAAGGAATGCAAC

SEQ ID NO: 41 sgRNA 22 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGA AAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 42 sgRNA 23 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGA AAAATAGGTCAAGGAATGCAAC
SEQ ID NO: 43 sgRNA 24 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGA AATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 44 n-terminal NLS SV40 large T antigen (from plasmid) (Artificial sequence)	EGAPKKKRKVGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAID RIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRY TKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAF YECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 45 n-terminal NLS SV40 large T antigen (Artificial sequence)	PKKKRKVGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIV DIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTK DIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGL TGTEYAMAIRKAISILRSFEKRRRNAERRIEYEKSKKEYLELIDDVEKGKTNK IVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKH PTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKK SIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGI DRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAK KLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVL EDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVM IIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYE CPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 46 n-terminal NLS c-myc (Artificial sequence)	PAAKRVKLDGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDR IVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYT KDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKA FYECPTFRWEEKLHAYVCSEPDK

SEQ ID NO: 47 n-terminal NLS TUS (Artificial sequence)	KLKIKRPVKGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYT KDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKA FYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 48 n-terminal NLS NLP (Artificial sequence)	AVKRPAATKKAGQAKKKKLDGGSMPSETYITKTLSLKLIPSDEEKQALENYF ITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFK FKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWK VSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEY LELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVK DIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFAS ISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTV KVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENR YKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQI VEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDM IKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNAD LNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 49 c-terminal NLS SV40 large T antigen (from plasmid) (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSEGAPKKKRKV
SEQ ID NO: 50 c-terminal NLS SV40 large T antigen (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSPKKKRKV
SEQ ID NO: 51 c-terminal NLS c-myc (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV

	GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSPAAKRVKLD
SEQ ID NO: 52 c-terminal NLS TUS (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSKLKIKRPVK
SEQ ID NO: 53 c-terminal NLS NLP (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSAVKRPAATKKAGQAKKKKLD
SEQ ID NO: 54 n- and c-terminal NLS SV40 large T antigen (from plasmid) (Artificial sequence)	EGAPKKKRKVGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAID RIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRY TKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKA FYECPTFRWEEKLHAYVCSEPDKGGSEGAPKKKRKV
SEQ ID NO: 55 n- and c-terminal NLS SV40 large T antigen (Artificial sequence)	PKKKRKVGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIV DIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTK DIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGL TGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNK IVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKH PTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKK SIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGI DRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAK KLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVL EDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVM IIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYE CPTFRWEEKLHAYVCSEPDKGGSPKKKRK
SEQ ID NO: 56	PAAKRVKLDGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDR IVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYT

n- and c-terminal NLS c-myc (Artificial sequence)	KDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKA FYECPTFRWEEKLHAYVCSEPDKGGSPAAKRVKLD
SEQ ID NO: 57 n- and c-terminal NLS TUS (Artificial sequence)	KLKIKRPVKGGSMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYT KDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAP GLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKT NKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKA FGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAM AKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVI VLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVP VMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKA FYECPTFRWEEKLHAYVCSEPDKGGSKLKIKRPVK
SEQ ID NO: 58 n- and c-terminal NLS NLP (Artificial sequence)	AVKRPAATKKAGQAKKKKLDGGSMPSETYITKTLSLKLIPSDEEKQALENYF ITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFK FKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWK VSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEY LELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVK DIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFAS ISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTV KVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENR YKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQI VEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDM IKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNAD LNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSAVKRPAATKKAGQAK KKKLD
SEQ ID NO: 59 pCMV – hu191034_6034 Cas14 C (term msfGFP) (Artificial sequence)	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDKGGSVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDAT NGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPE GYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEY NFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLL PDNHYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK
SEQ ID NO: 60	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNK NEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHIC KTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIR KAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQ

pCMV – hu191034_6034 Cas14 C (no NLS) (Artificial sequence)	RVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVEL HKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAV GCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKI RLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRK CSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEK LHAYVCSEPDK
SEQ ID NO: 61 EMX1 5' G guides sgRNA 1 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAA CCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 62 EMX1 5' G guides sgRNA 2 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCT GAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATC CTTACCTATTGAAAAATAGGTCAAGGAATGCAAC
SEQ ID NO: 63 EMX1 5' G guides sgRNA 3 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAA GTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 64 EMX1 5' G guides sgRNA 4 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAA ATAGGTCAAGGAATGCAAC
SEQ ID NO: 65 EMX1 5' G guides sgRNA 5 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATA ATAGGTCAAGGAATGCAAC
SEQ ID NO: 66 EMX1 5' G guides sgRNA 6 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATA GGTCAAGGAATGCAAC
SEQ ID NO: 67 EMX1 5' G guides sgRNA 7 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAA GGAATGCAAC
SEQ ID NO: 68	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGA ATGCAAC

EMX1 5' G guides sgRNA 8 (Artificial sequence)	
SEQ ID NO: 69 EMX1 5' G guides sgRNA 9 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 70 EMX1 5' G guides sgRNA 10 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 71 EMX1 5' G guides sgRNA 11 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGATGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 72 EMX1 5' G guides sgRNA 12 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 73 EMX1 5' G guides sgRNA 13 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 74 EMX1 5' G guides sgRNA 14 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAACCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 75 EMX1 5' G guides sgRNA 15 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGAATGCAAC
SEQ ID NO: 76 EMX1 5' G guides sgRNA 16 (Artificial sequence)	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAATAGGTCAAGGAATGCAAC
SEQ ID NO: 77	GTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAATAATAGGTCAAGGAATGCAAC

EMX1 5' G guides sgRNA 17 (Artificial sequence)	
SEQ ID NO: 78 EMX1 5' G guides sgRNA 18	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAACCCAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 79 EMX1 5' G guides sgRNA 19 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAA
SEQ ID NO: 80 DR only 1 (Artificial sequence)	GACCCAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCAT TG
SEQ ID NO: 81 DR only 2 (Artificial sequence)	GAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 82 DR only 3 (Artificial sequence)	GAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 83 DR only 4 (Artificial sequence)	GTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 84 Tracr only 1 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCT GAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATC CTTACCTA
SEQ ID NO: 85 Tracr only 2 (Artificial sequence)	GCTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCG CTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTA
SEQ ID NO: 86 Tracr only 3 (Artificial sequence)	GGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGT CTGCCCACCTCAGAGTGGGTATCCTTACCTA

SEQ ID NO: 87 Tracr only 4 (Artificial sequence)	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTG CCCACCTCAGAGTGGGTATCCTTACCTA
SEQ ID NO: 88 Tracr only 5 (Artificial sequence)	GATTGTATTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCT GAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATC CTTACCTA
SEQ ID NO: 89 Tracr only 6 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTA
SEQ ID NO: 90 Tracr only 6 (Artificial sequence)	G---------- TGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCT GCCCACCTCAGAGTGGGTATCCTTACCTA
SEQ ID NO: 91 5pr_trunc_4 (Artificial sequence)	GTTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGG AGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTAC CTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 92 5pr_trunc_5 (Artificial sequence)	GTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGA GGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACC TATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 93 5pr_trunc_6 (Artificial sequence)	GATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAG GATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCT ATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 94 5pr_trunc_7 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 95 SL1_modification_1 (Artificial sequence)	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 96 SL1_modification_2 (Artificial sequence)	GCTCCACTTTACTAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC

SEQ ID NO: 97 SL1_modification_3 (Artificial sequence)	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 98 SL1_modification_4 (Artificial sequence)	GCTCCACTTTAATAAGTGGAGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 99 SL1_modification_5 (Artificial sequence)	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 100 SL1_modification_6 (Artificial sequence)	GTGCTCCACTTTAATAAGTGGTGCATTCCAAAGCTATATGCTGAGGGAGG ATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTA TTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 101 SL1_modification_7 (Artificial sequence)	GCTCCACTTGTAATCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGG ATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTA TTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 102 SL1_modification_8 (Artificial sequence)	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGA GGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACC TATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 103 SL1_modification_9 (Artificial sequence)	GCTCCACTTGGCTAATGCCAAGTGGTGCCTTCCAAAGCTATATGCTGAGG GAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTA CCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 104 SL1_modification_1 (Artificial sequence)	GCTCCACTTGGCATAATTGCCAAGTGGTGCCTTCCAAAGCTATATGCTGA GGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCT TACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 105 SL1_MS2_hp (Artificial sequence)	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTAT ATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGG GTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 106 SL2_modification_1 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTAATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 107 SL2_modification_2	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTAAATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCAAC

(Artificial sequence)	
SEQ ID NO: 108 SL2_modification_3 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCCTATATGGCTGAGGGAGG ATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTA TTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 109 SL2_modification_4 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGA GGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACC TATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 110 SL2_modification_5 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGG GAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTA CCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 111 SL2_modification_6 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTTATATAGCAGCTGA GGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCT TACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 112 SL2_modification_7 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTGTATATCAGCAGCT GAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATC CTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 113 SL2_MS2_hp (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCACATGAGGATCACCCATG TGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGG TATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 114 increase_interaction_w_crRNA_13 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGATTGCAAC
SEQ ID NO: 115 increase_interaction_w_crRNA_14 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCACGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAGTGCAAC
SEQ ID NO: 116 increase_interaction_w_crRNA_15 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGACTGCAAC
SEQ ID NO: 117 increase_interaction_w_crRNA_16	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATGCAAC

(Artificial sequence)	
SEQ ID NO: 118 increase_interaction_w_crRNA_17 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTCGATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATCGAAC
SEQ ID NO: 119 increase_interaction_w_crRNA_18 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGAGTGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAACTCAAC
SEQ ID NO: 120 increase_interaction_w_crRNA_19 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCGTGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAACGCAAC
SEQ ID NO: 121 increase_interaction_w_crRNA_20 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGTATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATACAAC
SEQ ID NO: 122 increase_interaction_w_crRNA_21 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCGGC
SEQ ID NO: 123 increase_interaction_w_crRNA_22 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCCGC
SEQ ID NO: 124 increase_interaction_w_crRNA_23 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCGGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAACGCAAC
SEQ ID NO: 125 increase_interaction_w_crRNA_24 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGTAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGA AAAGTAATAGGTCAAGGAATACAAC
SEQ ID NO: 126 increase_interaction_w_crRNA_25 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATGCGGC

SEQ ID NO: 127 increase_interaction_w_crRNA_26 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG AAAAGTAATAGGTCAAGGAATGCCGC
SEQ ID NO: 128 increase_interaction_of_SL4_3 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACGCTAGACGTGGGTATCCTTACCTAT TGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 129 increase_interaction_of_SL4_4 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACTGCTAGACAGTGGGTATCCTTACCT ATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 130 increase_interaction_of_SL4_5 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTGCTAGACAGGTGGGTATCCTTAC CTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 131 increase_interaction_of_SL4_6 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACGCTCAGACGTGGGTATCCTTACCTA TTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 132 increase_interaction_of_SL4_7 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACTGCTCAGACAGTGGGTATCCTTACC TATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 133 increase_interaction_of_SL4_8 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTGCTCAGACAGGTGGGTATCCTTA CCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 134 increase_interaction_of_SL4_9 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACGCTGCTCAGACAGCGTGGGTATCCT TACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 135 increase_interaction_of_SL4_10 (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACTGCTGCTCAGACAGCAGTGGGTATC CTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC

SEQ ID NO: 136 SL3_MS2_hp (Artificial sequence)	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACACATGAGGATCACCCATGTGTGGGT ATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 137 increase_interaction_of _SL5_4 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTA AAAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 138 increase_interaction_of _SL5_5 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG GAAAAGCTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 139 increase_interaction_of _SL5_6 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGC TAAAAGAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 140 increase_interaction_of _SL5_7 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGT GAAAAGCATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 141 increase_interaction_of _SL5_8 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGC TGAAAAGCAGTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 142 increase_interaction_of _SL5_9 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTG GCTGAAAAGCAGCTAATAGGTCAAGGAATGCAAC
SEQ ID NO: 143 increase_interaction_of _SL5_10 (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGT GCTGAAAAGCAGCATAATAGGTCAAGGAATGCAAC
SEQ ID NO: 144 SL4_MS2_hp (Artificial sequence)	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATG GGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAC ATGAGGATCACCCATGTAATAGGTCAAGGAATGCAAC

The percent identity of Cas12ms to other Cas12 orthologs can be found in Tables 2-13 below.

Table 2

	Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	Cas12i2	Cas12i1	Cas12g1	Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	CasY5
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180		18.819	5.239	5.689	10.024	7.355	6.225	4.971
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	18.819		5.027	4.978	8.197	6.75	6.78	4.996
Cas12i2	5.239	5.027		4.944	5.939	5.899	4.155	4.478
Cas12i1	5.689	4.978	4.944		4.46	5.688	4.461	6.058
Cas12g1	10.024	8.197	5.939	4.46		7.375	7.576	5.483
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	7.355	6.75	5.899	5.688	7.375		10.271	4.31
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	6.225	6.78	4.155	4.461	7.576	10.271		3.457
CasY5	4.971	4.996	4.478	6.058	5.483	4.31	3.457	
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	8.029	7.91	3.986	4.859	6.178	6.734	6.186	3.336
CasY6	5.089	5.319	4.61	6.114	4.878	4.6	4.351	6.205
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	5.415	7.185	4.476	4.6	6.072	7.925	6.364	6.332
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	6.218	7.407	3.864	3.727	5.315	7.65	6.347	3.843
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	6.371	5.585	3.575	3.022	5.478	7.386	6.088	3.274
Cas12a_UPI00094EEDB4	3.643	3.157	5.548	4.833	4.397	3.972	4.869	5.552
Cas12a_UPI000B4235CE	4.519	3.519	6.326	5.434	4.604	5.118	4.828	5.773
Cas12a_UPI000818CC52	4.525	3.451	6.335	5.512	4.535	5.126	4.758	5.71

Cas12a_UP80007B78B7F	4.519	3.519	6.326	5.505	4.604	5.118	4.828	5.773
Cas12a_UP8008B4235F9	4.519	3.519	6.326	5.501	4.604	5.118	4.828	5.773
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	5.204	5.391	3.425	3.51	4.439	5.663	5.627	3.501
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	6.039	6.595	4.207	3.321	6.144	4.903	6.19	3.298
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	3.808	5.292	4.429	3.337	4.581	6.917	5.538	2.681
CasY4	6.058	4.651	5.598	3.922	6.556	4.348	3.766	6.522
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	7.333	5.063	3.626	3.053	5.27	6.97	5.952	3.469
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	5.767	7.752	4.511	4.255	6.195	6.031	5.381	4.825
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	6.307	8.258	4.444	4.089	5.457	7.386	5.706	4.474
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	5.696	6.349	4.178	3.815	5.402	6.036	4.654	3.616
Cas12b1	6.801	6.015	5.403	5.47	6.919	6.586	4.432	5.237
CasX1	7.116	5.52	6.421	6.225	6.724	6.571	5.714	5.849
CasX2	7.033	5.592	5.867	5.341	6.796	6.522	5.28	6.061
CasY1	6.31	4.979	7.038	4.286	6.423	4.376	4.513	6.407
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	7.628	7.483	4.688	4.377	6.883	9.741	9.105	2.842
Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006	8.531	7.733	2.921	3.03	5.952	5.855	4	2.743
Cas14u.8|3300005660.a|Ga0073984_10021651|765..1943	7.341	5.992	3.891	3.39	5.812	6.317	4.341	2.741
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	6.137	5.615	3.783	3.491	5.797	8.841	3.797	3.527
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	7.444	5.898	4.051	3.707	6.045	11.318	9.486	3.495
Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|revcom	7.459	7.246	3.961	4.864	6.021	6.156	4.859	3.163
CasY3	5.921	4.781	6.715	4.958	5.753	4.456	3.918	6.795
633299_527_protein_locus_of_contig_Scf8d15 - Query protein (633299_527) (4)	6.853	7.057	4.203	3.491	6.109	5.819	5.28	3.815
8971_2857_protein_locus_of_contig_QEJQ010000083.1 - Query protein (8971_2857)	6.677	6.14	5.263	2.944	5.579	4.866	4.53	3.704

9265_901_protein_locus_of_contig_QEFX01000005.1 - Query protein (9265_901)	6.567	6.043	5.203	3.012	5.493	4.942	4.444	3.759
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	7.317	8.101	4.094	2.993	6.806	6.484	5.663	3.206
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	7.007	6.564	4.187	3.868	6.729	5.271	6.688	3.439
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	6.191	4.78	3.349	5.14	4.666	7.069	6.923	3.578
CasY2	5.34	5.364	5.168	6.993	5.294	5.448	4.297	5.865
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	9.517	7.923	5.44	4.995	7.417	7.339	5.346	3.767
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	7.921	7.629	5.186	4.857	8.052	7.891	8.1	3.733
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	7.983	7.422	5.442	4.447	7.403	6.98	7.944	3.534
Cas14b.4|cg3_0.2_scaffold_785_c_curated|32521..34155	9.986	9.823	4.608	4.135	8.105	8.739	5.295	3.826
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	9.655	8.243	5.366	4.846	6.839	8.204	6.818	4.074
Cas14u.2|3300002172.a|3GI247303267401002785|496..1605|revcom	6.828	7.084	4.02	3.425	7.723	5.91	5.854	3.209
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	9.904	9.511	4.701	5.446	7.245	6.619	7.362	4.093
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	9.218	9.078	5.352	4.843	7.324	7.122	7.355	4.227
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	9.986	8.071	4.931	5.104	7.029	7.069	7.199	4.029
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	10.125	9.029	4.931	4.915	7.427	7.806	8.764	3.491
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	10.028	8.038	4.322	5.239	8.216	7.932	7.207	5.446
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	10.633	8.311	5.604	5.365	7.97	7.402	6.149	5.013
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	10.852	9.041	5.408	5.07	8.503	8.146	6.147	4.732
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	11.434	8.289	5.032	4.11	5.732	8.818	6.2	3.591

209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	21.344	13.074	9.571	5.621	12.261	16.216	10.046	6.757
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	20.661	13.91	9.31	5.288	12.295	16.588	10.096	6.516
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	8.04	7.412	4.074	4.384	7.067	6.771	5.842	3.704
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	8.09	8.85	4.356	4.093	8.864	7.084	6.723	3.951
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	8.391	7.859	4.906	6.029	7.915	6.409	6.349	4.228
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	8.545	9.06	4.72	5.326	7.65	7.711	6.46	3.887
Cas14b.16|3300005573.a|Ga0078972_1001015a|33758..35627	8.607	6.86	5.529	5.009	9.554	8.604	8.247	3.53
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	8.969	9.031	4.981	6.187	8.217	7.255	6.647	4.974
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	9.151	7.513	4.803	5.097	8.805	7.714	7.829	4.01
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	7.801	6.658	2.761	3.636	6.992	6.535	7.085	2.599
Cas12c1	3.749	5.389	5.444	5.339	5.582	4.362	3.803	5.334
Cas12c2	5.609	5.178	5.988	4.403	5.954	4.676	4.073	5.778
Cas12a_UPI001113398F	4.949	5.412	7.131	5.547	5.649	5.709	5.372	7.105
Cas12b_UPI001113398F	4.949	5.412	7.131	5.547	5.649	5.709	5.372	7.105
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	4.818	5.541	7.248	5.708	5.434	5.585	5.461	7.186
Cas12a_UPI00083514A7	5.013	5.917	5.824	5.837	5.986	5.254	5.085	6.941
Cas12b_UPI00083514A7	5.013	5.917	5.824	5.837	5.986	5.254	5.085	6.941
Cas12a_UPI00097159F1	4.865	6.396	6.03	5.934	5.845	5.1	5.743	6.921
Cas12b_UPI00097159F1	4.865	6.396	6.03	5.934	5.845	5.1	5.743	6.921
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.865	6.396	6.03	5.934	5.845	5.1	5.743	6.921
Cas12a_UPI0009715A14	4.865	6.396	6.03	5.934	5.935	5.1	5.743	6.838
Cas12b_UPI0009715A14	4.865	6.396	6.03	5.934	5.935	5.1	5.743	6.838
Cas12a_UPI00097159CF	4.865	6.396	6.03	5.934	5.935	5.1	5.743	6.915
Cas12b_UPI00097159CF	4.865	6.396	6.03	5.934	5.935	5.1	5.743	6.915
Cas12a_UPI000832F6D2	4.861	6.218	6.114	6.008	5.75	5.369	6.011	6.843
Cas12b_UPI000832F6D2	4.861	6.218	6.114	6.008	5.75	5.369	6.011	6.843
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	5.122	5.959	5.946	5.692	5.93	5.096	6.011	7.076

OspCas12c	5.082	6.075	5.914	5.588	5.657	5.251	3.54	4.853
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	6.658	8.752	4.39	4.128	9.103	8.21	7.283	5.804
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	5.931	7.333	3.933	2.982	6.91	7.211	6.686	4.204
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	6.989	8.614	3.599	3.458	6.914	7.487	7.55	4.856
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	6.465	7.995	3.937	3.451	8.56	6.098	6.676	4.668

Table 3

	Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	CasY6	Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	Cas14f.2|3300009991.a|Ga01050421_100140|1624..3348	Cas14a.6|3300012359.a|Ga0137385_1000156|41289..42734	Cas12a_UPI00094EEDB4	Cas12a_UPI000B4235CE	Cas12a_UPI00098 18CC52
Cas14g.1|RBG_13_scaffold_1491_curated|15949..18189	8.029	5.089	5.415	6.218	6.371	3.643	4.519	4.525
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	7.91	5.319	7.185	7.407	5.585	3.157	3.519	3.451
Cas12i2	3.986	4.61	4.476	3.864	3.575	5.548	6.326	6.335
Cas12i1	4.859	6.114	4.6	3.727	3.022	4.833	5.434	5.512
Cas12g1	6.178	4.878	6.072	5.315	5.478	4.397	4.604	4.535
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	6.734	4.6	7.925	7.65	7.386	3.972	5.118	5.126
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	6.186	4.351	6.364	6.347	6.088	4.869	4.828	4.758
CasY5	3.336	6.205	6.332	3.843	3.274	5.552	5.773	5.71
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffol		4.691	5.862	5.07	9.029	4.555	4.758	4.758

d_20906_curated|649..2829								
CasY6	4.691		6.434	3.704	3.819	6.452	6.443	6.452
Cas14f.3|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	5.862	6.434		23.19	6.846	3.92	4.278	4.278
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	5.07	3.704	23.19		6.352	2.595	2.961	2.966
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	9.029	3.819	6.846	6.352		2.313	3.241	3.241
Cas12a_UPI00094EEDB4	4.555	6.452	3.92	2.595	2.313		41.921	41.996
Cas12a_UPI000B4235CE	4.758	6.443	4.278	2.961	3.241	41.921		99.618
Cas12a_UPI000818CC52	4.758	6.452	4.278	2.966	3.241	41.996	99.618	
Cas12a_UPI0007B78B7F	4.758	6.443	4.278	2.961	3.241	42.07	99.771	99.847
Cas12a_UPI000B4235F9	4.758	6.443	4.278	2.961	3.241	42.039	99.466	99.389
Cas14e.2|rifcsphowo2_01_scaffold_81231_curated|976..2217	6.259	3.609	6.964	8.233	6.705	2.73	3.191	3.191
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	5.817	3.6	5.93	6.777	7.529	2.886	3.183	3.183
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	7.083	3.852	6.868	6.623	6.936	3.196	3.658	3.658
CasY4	4.635	9.225	6.672	4.25	3.466	5.765	6.089	6.098
Cas14h.3|3300009698.a|Ga0116216_10000905|80059504	7.077	3.424	8.026	8.847	8.672	3.248	2.877	2.877
Cas14h.1|3300005602.a|Ga0070762_10001740|73779071|revcom	5.875	4.481	7.652	7.413	7.333	3.752	3.979	3.979
Cas14h.2|3300005921.a|Ga0070766_10011912|3842083	5.643	3.633	7.477	7.362	6.588	3.379	3.991	3.991
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	6.472	2.96	6.95	8.05	8.818	3.414	3.104	3.104
Cas12b1	5.527	6.121	6.416	5.131	4.61	3.627	5.205	5.066
CasX1	5.443	6	5.825	3.887	5.123	5.151	6.204	6.213
CasX2	6.279	7.645	5.859	3.854	6.515	4.716	5.564	5.572
CasY1	5.178	6.381	6.047	3.874	4.736	5.234	5.688	5.626
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	7.945	4.077	7.343	6.518	9.524	3.73	3.026	3.026
Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006	7.448	2.927	7.542	9.769	8.554	2.846	3.007	3.007
Cas14u.8|3300005660.a|Ga0073984_10021651|765..1943	7.26	3.712	7.972	8.099	8.704	2.771	3.075	3.075

Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	5.761	2.776	4.33	7.317	10.2	3.082	3.077	3.077
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	6.389	3.772	7.412	7.026	11.132	3.991	4.372	4.372
Cas14e.2|3300001245.a|JGI12048J13642_1020128 6|4257..5489|revcom	7.191	2.675	6.658	5.415	9.312	2.822	3.351	3.351
CasY3	5.481	8.333	5.316	3.772	3.416	5.999	6.877	6.887
633299_527_protein_locus_of_contig_Scfld15 - Query protein (633299_527) (4)	6.474	3.323	7.832	7.679	9.298	3.009	3.236	3.236
8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	6.922	3.078	7.059	8.098	10.478	2.659	3.223	3.223
9265_901_protein_locus_of_contig_OEFX010000005.1 - Query protein (9265_901)	6.812	3.133	6.946	7.934	10.222	2.716	3.195	3.195
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	6.292	3.917	9.655	10.224	6.623	2.868	4.189	4.189
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	6.936	2.76	9.272	9.324	10.23	2.679	2.518	2.518
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.658	2.441	5.27	4.647	6.549	3.966	5.004	5.008
CasY2	4.878	6.471	4.818	2.85	4.903	6.557	6.424	6.362
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	12.273	4.194	7.65	7.267	17.056	4.855	3.909	3.909
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	12.188	5.436	6.827	7.401	19.342	3.801	4.425	4.425
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	11.523	5.485	6.426	7.049	19.923	3.395	4.17	4.17
Cas14b.4|cgt_0.2_scaffold_785_c_curated|32521..34155	7.367	3.512	6.711	7.764	8.305	3.807	3.106	3.106
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	8.713	3.816	7.662	8.75	8.819	4.338	3.464	3.464
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1685|revcom	7.022	2.718	5.618	5.965	8	2.644	2.638	2.638
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	8.647	3.987	8.422	7.75	10.616	4.439	4.5	4.507
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	10.57	4.19	8.56	6.615	8.848	4.471	4.15	4.15
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	10.497	4.093	8.548	7.373	10.067	4.766	4.29	4.29

Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	10.083	3.692	7.87	7.988	9.564	4.375	4.29	4.29
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	8.482	3.92	6.937	6.202	10.282	3.724	4.267	4.267
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	9.707	4.124	7.412	6.724	9.35	4.08	3.92	3.926
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	10.174	5.044	8.524	7.364	9.076	4.405	4.099	4.099
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	8.733	3.531	6.709	7.4	11.616	2.914	3.265	3.265
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	13.531	5.979	12.057	10.37	16.667	5.092	6.061	6.061
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	12.329	5.696	12.546	10.811	16.129	4.792	5.992	5.992
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	7.393	3.543	5.728	5.503	8.423	3.917	3.514	3.514
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	7.345	4.282	6.633	4.809	9.56	4.012	5.174	5.174
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	7.078	3.909	5.122	4.492	6.076	3.474	4.502	4.508
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	7.441	3.876	6.034	5.232	6.378	3.479	5.469	5.477
Cas14b.16|3300005573.a|Ga0078972_1001015a|33758..35627	7.294	4.444	8.161	7.123	9.385	5.104	5.097	5.104
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	8.621	4.167	7.412	7.613	8.661	4.224	4.587	4.671
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	6.974	4.567	7.263	6.589	9.291	4.228	4.82	4.904
Cas14u.1|3300009029.a|Ga0066793_10010091|37..3113|revcom	7.865	2.972	6.276	7.279	8.884	2.422	3.04	3.04
Cas12e1	3.943	7.076	5.155	3.681	3.421	7.387	7.064	7.074
Cas12c2	4.396	6.856	4.448	3.598	4.153	5.411	6.555	6.564
Cas12a_UPI0011133988	3.91	7.015	6.356	4.2	2.899	5.679	5.297	5.233
Cas12b_UPI0011133988	3.91	7.015	6.356	4.2	2.899	5.679	5.297	5.233
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	3.747	6.942	6.394	4.259	2.893	5.575	5.323	5.259

Cas12a_UPI00083514A7	4.391	6.428	6.014	4.541	4.159	6.026	5.583	5.448
Cas12b_UPI00083514A7	4.391	6.428	6.014	4.541	4.159	6.026	5.583	5.448
Cas12a_UPI00097159F1	5.165	6.133	6.324	4.558	2.69	6.82	6.017	5.882
Cas12b_UPI00097159F1	5.165	6.133	6.324	4.558	2.69	6.82	6.017	5.882
Cas12b_sp|T0D7A2|CS12B_ALIAG	5.165	6.133	6.324	4.558	2.69	6.82	6.017	5.882
Cas12a_UPI0009715A14	5.165	6.058	6.324	4.649	2.69	6.82	6.017	5.882
Cas12b_UPI0009715A14	5.165	6.058	6.324	4.649	2.69	6.82	6.017	5.882
Cas12a_UPI00097159CF	5.165	6.133	6.324	4.558	2.69	6.82	6.017	5.882
Cas12b_UPI00097159CF	5.165	6.133	6.324	4.558	2.69	6.82	6.017	5.882
Cas12a_UPI000832F6D2	5.33	6.502	6.416	4.831	2.966	6.671	5.87	5.735
Cas12b_UPI000832F6D2	5.33	6.502	6.416	4.831	2.966	6.671	5.87	5.735
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	5.161	6.353	6.416	4.649	2.966	6.671	5.941	5.806
OspCas12c	4.021	7.595	5.314	4.073	3.471	6.104	7.567	7.436
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	6.591	5.418	6.436	5.503	6.078	3.74	4.064	4.064
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	5.284	3.692	7.015	7.794	5.063	2.937	3.303	3.303
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	7.097	3.668	6.435	6.984	5.91	4.321	3.988	3.988
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	6.684	3.462	5.92	6.726	6.171	3.181	3.627	3.627

Table 4

	Cas12a_UPI0007B78B7F	Cas12a_UPI000B4235F9	Cast4e.2|rifesplowo2_01_scaffold_81231_curated|976..2217	Cast4e.1|rifesphigho2_01_scaffold_566_curated|11306..11 4313	Cas14e.3|rifesphigho2_01_scaffold_4702_curated|8288 1..8 4230|revcom	CasY4	Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	Cas14h.1|3300005602.a|Ga0070762_1000017407377..9071|revcom

Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	4.519	4.519	5.204	6.039	3.808	6.058	7.333	5.767
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	3.519	3.519	5.391	6.595	5.292	4.651	5.063	7.752
Cas12i2	6.326	6.326	3.425	4.207	4.429	5.598	3.626	4.511
Cas12i1	5.505	5.501	3.51	3.321	3.337	3.922	3.053	4.255
Cas12g1	4.604	4.604	4.439	6.144	4.581	6.556	5.27	6.195
Cas14d.3|RIFCSPLOWO2_01_FULL_ODI_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	5.118	5.118	5.663	4.903	6.917	4.348	6.97	6.031
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	4.828	4.828	5.627	6.19	5.538	3.766	5.952	5.381
CasY5	5.773	5.773	3.501	3.298	2.681	6.522	3.469	4.825
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	4.758	4.758	6.259	5.817	7.083	4.635	7.077	5.875
CasY6	6.443	6.443	3.609	3.6	3.852	9.225	3.424	4.481
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	4.278	4.278	6.964	5.93	6.868	6.672	8.026	7.652
Cas14f.2|3300009991.a|Ga0185042_100140|1624..3348	2.961	2.961	8.233	6.777	6.623	4.25	8.847	7.413
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	3.241	3.241	6.705	7.529	6.936	3.466	8.672	7.333
Cas12a_UPI00094EEDB4	42.07	42.039	2.73	2.886	3.196	5.765	3.248	3.752
Cas12a_UPI000B4235CE	99.771	99.466	3.191	3.183	3.658	6.089	2.877	3.979
Cas12a_UPI000818CC52	99.847	99.389	3.191	3.183	3.658	6.098	2.877	3.979
Cas12a_UPI0007B78B7F		99.542	3.191	3.183	3.658	6.089	2.877	3.979
Cas12a_UPI000B4235F9	99.542		3.191	3.183	3.658	6.089	2.877	3.979
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	3.191	3.191		22.222	23.108	2.723	6.346	5.354
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113969..114313	3.183	3.183	22.222		20.816	2.553	7.57	6.879
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	3.658	3.658	23.108	20.816		2.726	6.168	6.146
CasY4	6.089	6.089	2.723	2.553	2.726		3.48	3.361
Cas14b.3|3300009698.a|Ga0116216_10000905|8005..9504	2.877	2.877	6.346	7.57	6.168	3.48		13.942
Cas14b.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	3.979	3.979	5.354	6.879	6.146	3.361	13.942	
Cas14b.2|3300005921.a|Ga0070766_10011912|384..2081	3.991	3.991	5.448	6.154	7.179	2.773	14.56	65.12

Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20888|revcom	3.104	3.104	8.63	8.443	6.964	2.927	9.589	8.889
Cas12h1	5.205	5.205	5.396	5.383	4.556	3.965	5.166	4.577
CasX1	6.13	6.13	4.041	3.316	4.063	7.065	5.217	4.709
CasX2	5.564	5.49	4.603	3.556	4.316	7.422	5.489	4.044
CasY1	5.688	5.688	3.306	4.033	4.5	6.984	3.908	3.953
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	3.026	3.026	7.579	8.598	7.895	3.495	7.679	6.408
Cas14u.7|3300001256.a|3GI122103113797_10004690|5792..7006	3.007	3.007	8.463	8.609	9.298	4.114	13.546	10.764
Cas14u.8|3300005660.a|Gs0073904_10021651|765..1943	3.075	3.075	8.036	8.869	7.438	3.28	12.749	9.457
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	3.077	3.077	8.15	6.813	5.809	2.521	8.984	7.863
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	4.372	4.372	6.191	7.836	7.076	3.757	7.218	7.445
Cas14e.2|3300001245.a|3GI12048J13642_10201286|4257..5489|revcom	3.351	3.351	7.463	6.438	7.6	3.763	13.112	8.263
CasY3	6.877	6.877	3.198	2.936	3.128	7.777	3.926	3.568
633299_527_protein_locus_of_contig_Scfld15 - Query protein (633299_527) (4)	3.236	3.236	9.888	10.811	10.669	3.788	10.097	9.091
8971_2857_protein_locus_of_contig_OEJQ01000083.1 - Query protein (8971_2857)	3.223	3.223	9.832	8.794	7.586	4.111	12.281	9.594
9265_981_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_981)	3.195	3.195	9.579	8.557	7.399	4.248	12.42	9.946
Cas14u.6|3300006028.a|Gs0070717_10000677|54519..56201|revcom	4.189	4.189	7.611	5.146	5.651	4.23	11.058	12.342
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	2.518	2.518	10.909	10.633	8.457	3.972	12.527	10.584
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.004	5.004	6.285	6.667	6.947	3.333	5.308	4.944
CasY2	6.424	6.424	3.072	2.728	2.647	8.408	3.686	3.431
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	3.909	3.909	7.679	7.527	7.482	5.06	8.6	8.531
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	4.425	4.425	7.076	9.441	8.253	3.98	8.734	7.667
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	4.17	4.17	5.959	8.285	7.678	3.62	8.099	7.258

Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	3.106	3.103	7.356	7.638	6.667	4.488	8.829	7.571
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	3.464	3.462	6.713	6.768	6.04	4.73	8.795	7.166
Cas14a.2|3300002172.a|JGI24736J26740_1002785|496..1605|revcom	2.638	2.638	8.844	8.924	9.013	2.981	10.581	8.289
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	4.5	4.5	7.5	8.007	6.885	5.344	8.543	8.458
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	4.15	4.15	8.185	7.143	7.317	4.713	9.318	8.143
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	4.29	4.29	7.871	8.174	7.813	4.778	9.03	8.224
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	4.29	4.29	7.168	7.292	6.424	4.863	8.543	8.581
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	4.267	4.267	6.914	7.155	6.096	5.518	8.401	7.827
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	3.92	3.92	7.12	6.421	5.696	5.887	8.372	8.359
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	4.099	4.099	8.483	6.874	5.769	5.442	8.703	8.399
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	3.265	3.265	7.305	7.532	7.071	4.388	9.176	8.515
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	6.061	6.061	9.417	10.909	10.502	8.592	14	13.061
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	5.992	5.992	9.434	11.005	10.096	8.416	13.808	12.719
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	3.514	3.511	6.636	7.302	5.521	4.519	7.209	5.968
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	5.174	5.174	6.467	7.165	7.87	5.303	6.957	8.859
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2889|revcom	4.502	4.502	6.049	5.122	5.398	5.229	5.289	5.577
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	5.469	5.469	6.12	5.837	4.967	5.048	6.304	6.361
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	5.097	5.015	8.544	6.552	7.899	5.401	7.553	5.655

Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	4.587	4.431	8.416	5.366	7.084	5.755	7.951	6.212
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	4.82	4.82	9.36	7.553	8.94	5.356	7.034	7.251
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	3.04	3.04	8.12	9.013	8.678	3.168	11.469	7.005
Cas12c1	7.064	7.059	2.875	4.003	3.68	6.734	3.969	3.965
Cas12c2	6.555	6.485	2.421	3.003	2.836	5.498	3.997	3.846
Cas12a_UPI001113398F	5.225	5.225	3.768	3.483	5.239	6.737	4.758	5.206
Cas12b_UPI001113398F	5.225	5.225	3.768	3.483	5.239	6.737	4.758	5.206
Cas12b_tr|A0A1I7F1U9|A0A1I7F1U9_9BACL	5.252	5.252	3.772	3.388	5.133	6.546	4.633	5.306
Cas12a_UPI00083514A7	5.44	5.512	3.846	3.822	4.388	5.998	4.112	4.749
Cas12b_UPI00083514A7	5.44	5.512	3.846	3.822	4.388	5.998	4.112	4.749
Cas12a_UPI00097159F1	5.874	5.946	4.03	3.825	5.717	5.998	4.093	5.225
Cas12b_UPI00097159F1	5.874	5.946	4.03	3.825	5.717	5.998	4.093	5.225
Cas12b_sp|T0D7A2|CS12B_ALIAG	5.874	5.946	4.03	3.825	5.717	5.998	4.122	5.225
Cas12a_UPI0009715A14	5.874	5.946	4.03	3.825	5.717	6.074	4.122	5.225
Cas12b_UPI0009715A14	5.874	5.946	4.03	3.825	5.717	6.074	4.122	5.225
Cas12a_UPI00097159CF	5.874	5.946	4.03	3.825	5.717	6.074	4.122	5.225
Cas12b_UPI00097159CF	5.874	5.946	4.03	3.825	5.717	6.074	4.122	5.225
Cas12a_UPI000832F6D2	5.727	5.798	4.213	3.918	5.524	6.226	3.939	5.316
Cas12b_UPI000832F6D2	5.727	5.798	4.213	3.918	5.524	6.226	3.939	5.316
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	5.798	5.87	4.213	3.731	5.337	6.302	4.029	5.316
OspCas12c	7.426	7.567	2.922	3.084	3.328	5.58	3.325	4.133
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	4.064	4.064	4.154	6.37	6.038	5.068	6.96	9.531
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	3.303	3.303	5.096	5.949	5.512	4.017	6.192	7.657
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	3.988	3.988	4.416	6.19	4.212	4.693	7.099	8.769
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	3.627	3.627	5.76	6.924	4.944	4.014	8.791	7.351

Table 5

	Cas14b.2|3300005921.a|Ga0070766_1001912|384..2081	Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	Cas12b1	CasX1	CasX2	CasY1	Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18189	6.307	5.696	6.801	7.116	7.033	6.31	7.628	8.531
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	8.258	6.349	6.015	5.52	5.592	4.979	7.483	7.733
Cas12i2	4.444	4.178	5.403	6.421	5.867	7.038	4.688	2.921
Cas12i1	4.089	3.815	5.47	6.225	5.341	4.286	4.377	3.03
Cas12g1	5.457	5.402	6.919	6.724	6.796	6.423	6.883	5.952
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	7.386	6.036	6.586	6.571	6.522	4.376	9.741	5.855
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	5.706	4.654	4.432	5.714	5.28	4.513	9.105	4
CasY5	4.474	3.616	5.237	5.849	6.061	6.407	2.842	2.743
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	5.643	6.472	5.527	5.443	6.279	5.178	7.945	7.448
CasY6	3.633	2.96	6.121	6	7.645	6.381	4.077	2.927
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	7.477	6.95	6.416	5.825	5.859	6.047	7.343	7.542
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	7.362	8.05	5.131	3.887	3.854	3.874	6.518	9.769
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	6.588	8.818	4.61	5.123	6.515	4.736	9.524	8.554
Cas12a_UPI00094EEDB4	3.379	3.414	3.627	5.151	4.716	5.234	3.73	2.846
Cas12a_UPI000B4235CE	3.991	3.104	5.205	6.204	5.564	5.688	3.026	3.007
Cas12a_UPI000818CC52	3.991	3.104	5.066	6.213	5.572	5.626	3.026	3.007
Cas12a_UPI0007B78B7F	3.991	3.104	5.205	6.13	5.564	5.688	3.026	3.007
Cas12a_UPI000B4235F9	3.991	3.104	5.205	6.13	5.49	5.688	3.026	3.007
Cas14c.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	5.448	8.63	5.396	4.041	4.603	3.306	7.579	8.463

Cas14c.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	6.154	8.443	5.383	3.316	3.556	4.033	8.598	8.609
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	7.179	6.964	4.556	4.063	4.316	4.5	7.895	9.298
CasY4	2.773	2.927	3.965	7.065	7.422	6.984	3.495	4.114
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9564	14.56	9.589	5.166	5.217	5.489	3.908	7.679	13.546
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	65.12	8.889	4.577	4.709	4.044	3.953	6.408	10.764
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081		8.293	4.93	4.5	4.541	4.324	6.229	10.175
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	8.293		4.881	3.969	4.382	4.758	7.705	14.801
Cas12b1	4.93	4.881		5.945	6.267	4.718	4.875	4.745
CasX1	4.5	3.969	5.945		51.406	7.309	5.864	5.664
CasX2	4.541	4.382	6.267	51.406		7.535	5.497	5.411
CasY1	4.324	4.758	4.718	7.309	7.535		5.474	5.249
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	6.229	7.705	4.875	5.864	5.497	5.474		9.145
Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006	10.175	14.801	4.745	5.664	5.411	5.249	9.145	
Cas14u.8|3300005660.a|Ga0073984_10021651|765..1943	9.507	12.5	4.255	6.192	5.521	3.87	10.6	28.261
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	7.958	9.228	4.014	3.905	5.083	3.436	7.171	12.156
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	9.029	7.009	5.156	5.079	5.769	4.424	13.996	7.828
Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|revcom	8.844	12.104	5.041	5.397	5.139	3.953	8.35	18.075
CasY3	3.574	4.225	5.462	9.297	8.394	7.062	3.962	4.17
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	9.705	15.356	5.226	5.041	4.673	4.344	9.486	25.935
8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	10.261	14.228	4.701	6.12	5.96	4.607	10	25.515
9265_901_protein_locus_of_contig_OEFX010000051 - Query protein (9265_901)	10.42	14.712	4.762	6.156	5.889	4.558	9.978	26.316
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	11.774	7.573	5.38	4.67	5.123	3.815	6.436	10.071

466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	12.222	15.464	4.423	5.65	5.92	5.019	9.776	29.563
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.016	5.873	5.012	5.961	5.231	3.597	7.584	7.635
CasY2	3.529	2.977	5.167	7.529	8.089	6.977	4.255	3.442
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	8.065	9.431	6.36	7.611	7.257	5.355	9.206	9.108
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	7.155	8.919	6.683	7.21	7.278	5.119	8.379	10.6
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	7.401	8.136	7.101	7.78	7.749	5.086	8.561	11.637
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	8.833	8.108	5.945	7.07	7.446	5.839	9.508	9.141
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	8.095	8.217	5.813	7.026	7.202	5.641	9.903	9.091
Cas14u.2|3300002172.a|3GI24730326740_1002785|496..1605|revcom	8.496	10.291	4.207	5.981	5.932	3.751	9.919	13.35
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	8.804	8.373	6.413	6.9	6.861	4.666	9.402	10.929
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	8.76	7.813	6.475	6.191	6.78	4.625	10.517	10.83
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	9.349	7.559	6.325	6.263	6.533	4.972	9.879	10.969
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	8.878	6.951	6.205	5.741	6.639	5.249	11.092	10.275
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	8.333	7.562	5.917	6.076	6.757	6.141	10.611	10.247
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	8.217	7.852	6.936	5.906	8.016	7.182	9.365	10.351
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	8.517	7.519	6.746	6.475	8.091	6.9	9.532	11.379
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	8.37	9.534	5.522	5.695	6.032	5.614	11.058	14.481
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	12.863	11.189	8.434	11.905	12.04	9.346	17.593	20.657
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	13.043	10.545	8.202	10.601	10.764	8.633	17.073	20.297

Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	6.696	10.836	6.466	6.97	7.446	5.626	7.903	9.6
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	9.531	7.349	3.913	7.419	7.369	6.806	8.788	7.741
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2993_curated|553..2880|revcom	6.21	6.835	5.509	7.486	6.907	6.643	7.226	7.642
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	6.555	8.087	5.943	6.167	6.997	5.948	8.042	6.762
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	5.891	8.921	6.171	6.612	6.66	6.818	9.176	7.865
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	7.187	8.837	5.977	6.984	7.464	6.828	10.098	10.231
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	8.346	7.965	5.963	7.419	7.906	5.951	9.82	9.091
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	7.951	7.129	3.865	5.456	5.191	4.048	10.331	12.528
Cas12c1	4.196	3.75	5.352	7.083	7.192	7.049	3.92	3.259
Cas12c2	4.01	3.207	5.016	6.63	5.915	5.659	3.172	3.185
Cas12a_UPI001113398F	4.668	3.856	5.598	6.371	6.209	5.166	4.269	3.249
Cas12b_UPI001113398F	4.668	3.856	5.598	6.371	6.209	5.166	4.269	3.249
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	4.852	3.665	5.763	6.31	5.882	5.183	4.269	3.237
Cas12a_UPI00083514A7	4.659	4.087	5.64	6.034	5.705	5.624	3.993	3.584
Cas12b_UPI00083514A7	4.659	4.087	5.64	6.034	5.705	5.624	3.993	3.584
Cas12a_UPI00097159F1	5.133	4.452	6.374	5.916	5.412	4.867	4.457	3.306
Cas12b_UPI00097159F1	5.133	4.452	6.374	5.916	5.412	4.867	4.457	3.306
Cas12b_sp|T0D7A2|CS12B_ALIAG	5.133	4.452	6.374	5.916	5.412	4.867	4.457	3.306
Cas12a_UPI0009715A14	5.133	4.452	6.374	5.916	5.329	4.867	4.457	3.214
Cas12b_UPI0009715A14	5.133	4.452	6.374	5.916	5.329	4.867	4.457	3.214
Cas12a_UPI00097159CF	5.133	4.452	6.374	5.916	5.412	4.867	4.457	3.306
Cas12b_UPI00097159CF	5.133	4.452	6.374	5.916	5.412	4.867	4.457	3.306
Cas12a_UPI000832F6B2	5.225	4.27	5.938	6.076	5.74	5.102	4.731	3.394
Cas12b_UPI000832F6B2	5.225	4.27	5.938	6.076	5.74	5.102	4.731	3.394
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	5.133	4.27	5.766	5.993	5.657	5.102	4.453	3.394
OspCas12c	4.708	3.503	5.263	5.792	6.386	6.691	4.214	3.339
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	8.417	4.032	6.749	6.016	5.731	5.818	6.287	5.589
63461_4106_protein_locus_of_contig_LSK1.01000323 - Query protein (63461_4106) translation (4)	7.055	4.928	6.082	4.187	5.348	3.931	7.981	4.754

58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	7.154	5.24	6.176	5.123	5.184	4.182	6.955	6.139
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	7.87	5.294	6.007	4.266	4.418	4.771	7.442	5.785

Table 6

	Cas14u.8|3300005660.a|Ga0073904_1002165|765..1943	Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	Cas14c.2|3300001245.a|JGI20483136421_10201286|4257..5489|revcom	Cas Y3	633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	8971_2857_protein_locus_of_contig_OEJQ01000083.1 - Query protein (8971_2857)	9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	7.341	6.137	7.444	7.459	5.921	6.853	6.677	6.567
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	5.992	5.615	5.898	7.246	4.781	7.057	6.14	6.043
Cas12i2	3.891	3.783	4.051	3.961	6.715	4.203	5.263	5.203
Cas12i1	3.39	3.491	3.707	4.864	4.958	3.491	2.944	3.012
Cas12g1	5.812	5.797	6.045	6.021	5.753	6.109	5.579	5.493
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	6.317	8.841	11.318	6.156	4.456	5.819	4.866	4.942
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	4.341	3.797	9.486	4.859	3.918	5.28	4.53	4.444
CasY5	2.741	3.527	3.495	3.163	6.795	3.815	3.704	3.759
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	7.26	5.761	6.389	7.191	5.481	6.474	6.922	6.812
CasY6	3.712	2.776	3.772	2.675	8.333	3.323	3.078	3.133
Cas14f.3|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	7.972	4.33	7.412	6.658	5.316	7.832	7.059	6.946
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	8.099	7.317	7.026	5.415	3.772	7.679	8.098	7.934
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	8.704	10.2	11.132	9.312	3.416	9.298	10.478	10.222

Cas12a_UPI00094EEDB4	2.771	3.082	3.991	2.822	5.999	3.009	2.659	2.716
Cas12a_UPI000B4235CE	3.075	3.077	4.372	3.351	6.877	3.236	3.223	3.195
Cas12a_UPI000818CC52	3.075	3.077	4.372	3.351	6.887	3.236	3.223	3.195
Cas12a_UPI0007B78B7F	3.075	3.077	4.372	3.351	6.877	3.236	3.223	3.195
Cas12a_UPI000B4235F9	3.075	3.077	4.372	3.351	6.877	3.236	3.223	3.195
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	8.036	8.15	6.191	7.463	3.198	9.888	9.832	9.579
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	8.869	6.813	7.836	6.438	2.936	10.811	8.794	8.557
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	7.438	5.809	7.076	7.6	3.128	10.669	7.586	7.399
CasY4	3.28	2.521	3.757	3.763	7.777	3.788	4.111	4.248
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	12.749	8.984	7.218	13.112	3.926	10.097	12.281	12.42
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	9.457	7.863	7.445	8.263	3.568	9.091	9.594	9.946
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	9.507	7.958	9.029	8.844	3.574	9.705	10.261	10.42
Cas14e.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	12.5	9.228	7.009	12.104	4.225	15.356	14.228	14.712
Cas12h1	4.255	4.014	5.156	5.041	5.462	5.226	4.701	4.762
CasX1	6.192	3.905	5.079	5.397	9.297	5.041	6.12	6.156
CasX2	5.521	5.083	5.769	5.139	8.394	4.673	5.96	5.889
CasY1	3.87	3.436	4.424	3.953	7.062	4.344	4.607	4.558
Cas14u.3|19ft_2_nophage_unknown_scaffold_0_curated|508188..509648	10.6	7.171	13.996	8.35	3.962	9.486	10	9.978
Cas14u.7|3300001256.a|JGI122103113797_10004690|5792..7006	28.261	12.156	7.828	18.075	4.17	25.935	25.515	26.316
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943		12.121	9.742	15.529	4.174	30.288	33.6	34.456
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	12.121		8.35	11.83	3.416	11.364	14.604	14.217
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	9.742	8.35		6.526	4.352	8.876	8.096	8.12
Cas14c.2|3300001245.a|JGI120483113642_10201286|4257..5489|revcom	15.529	11.83	6.526		5.089	17.29	22.572	21.939
CasY3	4.174	3.416	4.352	5.089		4.437	4.277	4.414
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	30.288	11.364	8.876	17.29	4.437		32.987	33.838

8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	33.6	14.604	8.096	22.572	4.277	32.987		100
9265_901_protein_locus_of_contig_OEFX010000605.1 - Query protein (9265_901)	34.456	14.217	8.12	21.939	4.414	33.838	100	
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	9.769	7.193	7.143	8.448	4.663	8.772	9.851	9.836
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	31.759	13.022	9.562	19.851	4.474	37.047	44.092	44.134
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.056	5.311	7.04	5.263	2.703	6.642	5.394	5.882
CasY2	3.61	4.373	4.195	3.833	8.24	3.987	3.467	3.433
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	9.125	8.939	8.711	11.481	4.613	12.008	8.264	8.283
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	8.73	9.703	9.444	11.637	4.483	12.176	10.067	10.044
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	9.393	10.352	9.444	12.84	4.713	13.189	9.692	9.677
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	10.127	8.288	8.562	9.369	5.077	9.672	10.569	10.537
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	8.913	9.964	9.864	9.414	4.889	10.536	9.827	9.811
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|revcom	14.356	15.115	8.048	12.319	3.279	16.708	14.286	13.874
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	12.044	11.636	9.898	9.222	6.024	9.926	11.858	11.799
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	11.615	11.232	10.881	9.369	6.463	10.766	10.02	10
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	10.806	10.929	10.745	9.42	6.261	9.963	8.946	8.949
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	11.029	11.7	10.727	8.696	5.739	10.37	9.381	9.375
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	9.397	9.894	8.081	10.783	5.786	9.22	10.667	10.634
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	9.901	8.731	8.618	8.483	5.214	9.5	8.955	8.958
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	9.54	10.374	8.483	9.966	7.087	10	8.511	8.523

209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	13.812	12.963	10.448	13.202	4.834	13.536	13.165	13.165
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	18.224	19.725	15.962	17.371	9.487	19.048	17.143	17.143
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	17.241	19.807	14.851	17.327	9.919	18.593	16.08	16.08
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	8.682	8.786	6.38	7.455	7.18	9.179	10.14	9.949
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	8.019	8.805	8.116	8.025	5.766	9.365	7.731	7.921
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	6.162	5.905	7.031	6.282	6.567	6.865	7.173	7.202
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	7.004	6.986	7.833	6.914	6.833	7.672	7.714	7.736
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	8.64	8.28	8.1	7.547	5.056	8.9	9.424	9.589
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	8.553	10.164	7.98	8.347	5.702	8.099	9.386	9.381
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	8.224	8.867	7.516	8.039	5.541	8.609	9.567	9.381
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	14.151	13.122	8.876	10.502	3.643	13.318	13.384	13.022
Cas12c1	3.016	3.41	4.177	3.085	6.218	3.819	3.541	3.509
Cas12c2	3.598	3.434	4.362	3.156	7.863	3.275	3.226	3.283
Cas12a_UPI0011113398F	3.96	3.156	4.779	3.142	5.779	3.258	2.486	2.554
Cas12b_UPI0011113398F	3.96	3.156	4.779	3.142	5.779	3.258	2.486	2.554
Cas12b_tr|A0A1I7F1U9|A0A1I7F1U9_9BACL	3.957	3.055	4.867	3.139	5.807	3.348	2.481	2.55
Cas12a_UPI00083514A7	3.136	3.232	4.487	2.594	6.591	2.599	2.657	2.723
Cas12b_UPI00083514A7	3.136	3.232	4.487	2.594	6.591	2.599	2.657	2.723
Cas12a_UPI00097159F1	2.661	2.663	4.503	3.294	6.298	3.643	2.242	2.314
Cas12b_UPI00097159F1	2.661	2.663	4.503	3.294	6.298	3.643	2.242	2.314
Cas12b_sp|T0D7A2|CS12B_ALIAG	2.661	2.663	4.503	3.294	6.298	3.578	2.242	2.314
Cas12a_UPI0009715A14	2.661	2.663	4.503	3.294	6.298	3.578	2.242	2.314
Cas12b_UPI0009715A14	2.661	2.663	4.503	3.294	6.298	3.578	2.242	2.314
Cas12a_UPI00097159CF	2.661	2.663	4.503	3.294	6.298	3.578	2.242	2.314
Cas12b_UPI00097159CF	2.661	2.663	4.503	3.294	6.298	3.578	2.242	2.314

Cas12a_UPI0000832F6D2	2.75	2.849	4.592	3.294	6.523	3.483	2.045	2.119
Cas12b_UPI0000832F6D2	2.75	2.849	4.592	3.294	6.523	3.483	2.045	2.119
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	2.841	2.755	4.592	3.294	6.37	3.391	2.142	2.216
OspCas12c	3.496	2.685	3.504	3.89	7.179	2.941	3.38	3.519
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	6.938	5.556	5.588	6.577	4.038	5.918	6.988	7.026
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	7.084	5.307	6.907	6.743	3.362	6.988	5.302	5.197
58610_1188_protein_locus_of_contig_LFOD01000093 - Query protein (58610_1188) translation (5)	7.792	4.693	7.121	7.27	3.531	7.143	6.329	6.206
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	6.988	5.473	5.643	7.82	2.431	6.425	5.935	5.82

Table 7

	Cas14u.6|3300006028.a|Ga0070717_10000677|54519..56201|revcom	466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	Cas Y2	Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	Cas14a.3|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18189	7.317	7.007	6.191	5.34	9.517	7.921	7.983	9.986
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	8.101	6.564	4.78	5.364	7.923	7.629	7.422	9.823
Cas12i2	4.094	4.187	3.349	5.168	5.44	5.186	5.442	4.608
Cas12i1	2.993	3.868	5.14	6.993	4.995	4.857	4.447	4.135
Cas12g1	6.806	6.729	4.666	5.294	7.417	8.052	7.403	8.105
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_346_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	6.484	5.271	7.069	5.448	7.339	7.891	6.98	8.739

Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	5.663	6.688	6.923	4.297	5.346	8.1	7.944	5.295
CasY5	3.206	3.439	3.578	5.865	3.767	3.733	3.534	3.826
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	6.292	6.936	5.658	4.878	12.273	12.188	11.523	7.367
CasY6	3.917	2.76	2.441	6.471	4.194	5.436	5.485	3.512
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	9.655	9.272	5.27	4.818	7.65	6.827	6.426	6.711
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	10.224	9.324	4.647	2.85	7.267	7.401	7.049	7.764
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	6.623	10.23	6.549	4.903	17.056	19.342	19.923	8.305
Cas12a_UPI00094EEDB4	2.868	2.679	3.966	6.557	4.855	3.801	3.395	3.807
Cas12a_UPI000B4235CE	4.189	2.518	5.004	6.424	3.909	4.425	4.17	3.106
Cas12a_UPI000818CC52	4.189	2.518	5.008	6.362	3.909	4.425	4.17	3.106
Cas12a_UPI0007B78B7F	4.189	2.518	5.004	6.424	3.909	4.425	4.17	3.106
Cas12a_UPI000B4235F9	4.189	2.518	5.004	6.424	3.909	4.425	4.17	3.103
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	7.611	10.909	6.285	3.072	7.679	7.076	5.959	7.356
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	5.146	10.633	6.667	2.728	7.527	9.441	8.285	7.638
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	5.651	8.457	6.947	2.647	7.482	8.253	7.678	6.667
CasY4	4.23	3.972	3.333	8.408	5.06	3.98	3.62	4.488
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	11.058	12.527	5.308	3.686	8.6	8.734	8.099	8.829
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	12.342	10.584	4.944	3.431	8.531	7.667	7.258	7.571
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	11.774	12.222	5.016	3.529	8.065	7.155	7.401	8.833
Cas14e.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	7.573	15.464	5.873	2.977	9.431	8.919	8.136	8.108
Cas12h1	5.38	4.423	5.012	5.167	6.36	6.683	7.101	5.945
CasX1	4.67	5.65	5.061	7.529	7.611	7.21	7.78	7.07
CasX2	5.123	5.92	5.231	8.089	7.257	7.278	7.749	7.446
CasY1	3.815	5.019	3.597	6.977	5.355	5.119	5.086	5.839
Cas14u.3|19ft_2_nopbage_unknown_scaffold_0_curated|508188..509648	6.436	9.776	7.584	4.255	9.206	8.379	8.561	9.508

Cas14u.7|3300001256.a|JGI122103137 97_1000469 0|5792..7006	10.071	29.563	7.635	3.442	9.108	10.6	11.637	9.141
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	9.769	31.759	5.056	3.61	9.125	8.73	9.393	10.127
Cas14u.4|rifcsp2_39_4_full_scaffold_168_curated|84455..85657	7.193	13.022	5.311	4.373	8.939	9.703	10.352	8.288
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	7.143	9.562	7.04	4.195	8.711	9.444	9.444	8.562
Cas14c.2|3300001245.a|JGI120483136 42_1020128 6|4257..5489|revcom	8.448	19.851	5.263	3.833	11.481	11.637	12.84	9.369
CasY3	4.663	4.474	2.703	8.24	4.613	4.483	4.713	5.077
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	8.772	37.047	6.642	3.987	12.008	12.176	13.189	9.672
8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	9.851	44.092	5.394	3.467	8.264	10.067	9.692	10.569
9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)	9.836	44.134	5.882	3.433	8.283	10.044	9.677	10.537
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom		10.929	3.662	4	8.609	8.013	6.777	7.448
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	10.929		5.469	3.976	8.571	10.883	11.294	9.515
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	3.662	5.469		3.682	9.275	11.607	12.169	7.273
CasY2	4	3.976	3.682		5.665	4.847	5.41	4.588
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	8.609	8.571	9.275	5.665		36.43	35.519	10.697
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	8.013	10.883	11.607	4.847	36.43		81.6	10.788
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	6.777	11.294	12.169	5.41	35.519	81.6		10.103
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	7.448	9.515	7.273	4.588	10.697	10.788	10.103	
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	7.372	9.222	6.656	4.73	11.058	11.185	10.851	42.708
Cas14u.2|3300002172.a|JGI24730326740_1002785|496..1605|revcom	7.881	15.99	6.818	4.34	11.364	10.664	10.913	10.681
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	6.602	10.478	7.967	5.187	11.519	11.356	12.034	16.723

Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	6.897	10.256	8.007	5.326	10.316	9.241	8.911	15.92
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	6.393	10.019	8.02	5.475	12.02	10.248	10.248	16.279
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	6.579	10.575	8.183	5.047	11.39	10.282	9.453	16.5
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	8.401	10.48	8.293	5.963	12.841	11.675	11.675	19.224
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	7.176	8.968	9.37	5.56	11.42	11.22	11.87	19.677
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	8.58	9.343	6.75	5.812	12.324	10.561	10.891	19.569
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	9.015	12.707	8.294	5.468	13.024	13.3	13.547	19.861
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	15.164	17.788	11.814	8.836	22.326	20.183	19.725	30.374
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	14.592	17.259	11.062	8.168	22.549	20.29	19.807	29.557
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	5.832	8.838	5.433	5.241	8.728	8.636	9.242	13.557
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	7.447	10.841	6.871	5.626	9.954	11.145	10.502	11.458
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2889|revcom	5.625	7.171	5.941	6.029	6.804	7.445	7.28	11.14
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	7.098	7.867	5.882	6.426	8.564	8.073	8.29	11.211
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	7.264	9.493	8.722	5.719	11.502	9.969	9.502	13.509
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	10.502	10.94	6.891	5.491	10.129	8.654	9.206	13.744
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	8.976	10.427	7.573	5.008	10.129	9.807	8.931	11.765
Cas14u.1|3300009029.a|Ga0066793_10010091|37..3113|revcom	7.584	13.318	7.707	3.336	9.982	13.069	12.871	8.834
Cas12c1	4.286	3.647	2.584	6.014	4.106	4.466	4.203	4.24
Cas12e2	4.424	4.135	3.878	6.632	5.117	5.518	5.184	4.854

Cas12a_UPI0011133 98F	5.068	2.971	5.103	6.712	5.418	4.288	5.077	4.117
Cas12b_UPI0011133 98F	5.068	2.971	5.103	6.712	5.418	4.288	5.077	4.117
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	5.158	3.058	5.169	6.642	5.142	4.189	4.977	4.026
Cas12a_UPI0008351 4A7	4.599	2.308	4.728	5.927	4.487	4.455	4.517	4.45
Cas12b_UPI0008351 4A7	4.599	2.308	4.728	5.927	4.487	4.455	4.517	4.45
Cas12a_UPI0009715 9F1	4.428	2.844	5.302	6.616	4.69	4.944	5.097	3.911
Cas12b_UPI0009715 9F1	4.428	2.844	5.302	6.616	4.69	4.944	5.097	3.911
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.428	2.844	5.302	6.656	4.69	4.944	5.097	3.911
Cas12a_UPI0009715 A14	4.428	2.844	5.302	6.656	4.69	4.944	5.097	3.911
Cas12b_UPI0009715 A14	4.428	2.844	5.302	6.656	4.69	4.944	5.097	3.911
Cas12a_UPI0009715 9CF	4.428	2.844	5.302	6.656	4.69	4.944	5.097	3.911
Cas12b_UPI0009715 9CF	4.428	2.844	5.302	6.656	4.69	4.944	5.097	3.911
Cas12a_UPI000832F 6D2	4.7	2.746	5.297	6.886	4.592	4.846	4.907	3.814
Cas12b_UPI000832F 6D2	4.7	2.746	5.297	6.886	4.592	4.846	4.907	3.814
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	4.885	2.841	5.205	6.58	4.686	4.939	5.093	3.907
OspCas12e	4.217	3.859	2.885	5.808	4.327	4.302	4.383	4.475
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	8.626	6.991	4.119	4.227	7.225	6.755	6.461	8.346
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	8.15	5.351	5.14	4.503	9.451	6.656	6.815	7.309
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	8.423	6.931	4.695	3.976	6.577	6.211	5.745	5.828
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	7.402	6.187	4.409	4.174	7.553	6.667	7.302	6.202

Table 8

	Cas14b.7|3300013125.a|Ga0172369_10990737|p94..2652|revcom	Cas14u.2|3300002172.a|JGI24730J26740_10827785|496..1605|revcom	Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77378..78983	Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653_	Cas14b.8|3300013125.a|Ga0172369_10010464|8885..2489|revcom	Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_cu

					562 57			rat ed| 663 4..8 466 |rev co m
Cas14g.1|RBG_13_scaffo ld_1401_curated|15949..1 8180	9.655	6.828	9.904	9.218	9.986	10.125	10.028	10.633
Cas14g.2|3300009652.a|G a0123330_1010394|2814.. 5123	8.243	7.084	9.511	9.078	8.071	9.029	8.038	8.311
Cas12i2	5.366	4.02	4.701	5.352	4.931	4.931	4.332	5.604
Cas12i1	4.846	3.425	5.446	4.843	5.104	4.915	5.239	5.365
Cas12g1	6.839	7.723	7.245	7.324	7.029	7.427	8.216	7.97
Cas14d.3|RIFCSPLOW O2_01_FULL_OD1_45_ 34b_rifcsplowo2_01_scaf fold_3495_curated|25656 ..27605|revcom	8.204	5.91	6.619	7.122	7.069	7.806	7.932	7.402
Cas14d.1|RIFCSPHIGH O2_01_FULL_CPR_46_ 36_rifcsphigho2_01_scaff old_646_curated|49808..5 1616|revcom	6.818	5.854	7.362	7.355	7.199	8.764	7.207	6.149
CasY5	4.074	3.209	4.093	4.227	4.029	3.491	5.446	5.013
Cas14a.4|CG10_big_fil_r ev_8_21_14_0.10_scaffol d_20906_curated|649..28 29	8.713	7.022	8.647	10.57	10.497	10.083	8.482	9.707
CasY6	3.816	2.718	3.987	4.19	4.093	3.692	3.92	4.124
Cas14f.1|rifcsp13_1_sub1 0_scaffold_3_curated|389 06..41041	7.662	5.618	8.422	8.56	8.548	7.87	6.937	7.412
Cas14f.2|3300009991.a|G a0105042_100140|1624..3 348	8.75	5.965	7.75	6.615	7.373	7.988	6.202	6.724
Cas14a.6|3300012359.a|G a0137385_10000156|4128 9..42734	8.819	8	10.616	8.848	10.067	9.564	10.282	9.35
Cas12a_UPI00094EEDB 4	4.338	2.644	4.439	4.471	4.766	4.375	3.724	4.08
Cas12a_UPI000B4235CE	3.464	2.638	4.5	4.15	4.29	4.29	4.267	3.92
Cas12a_UPI000818CC52	3.464	2.638	4.507	4.15	4.29	4.29	4.267	3.926
Cas12a_UPI0007B78B7F	3.464	2.638	4.5	4.15	4.29	4.29	4.267	3.92
Cas12a_UPI000B4235F9	3.462	2.638	4.5	4.15	4.29	4.29	4.267	3.92
Cas14e.2|rifcsplowo2_01 _scaffold_81231_curated| 976..2217	6.713	8.844	7.5	8.185	7.871	7.168	6.914	7.12
Cas14e.1|rifcsphigho2_01 _scaffold_566_curated|11 3069..114313	6.768	8.924	8.007	7.143	8.174	7.292	7.155	6.421
Cas14e.3|rifcsphigho2_01 _scaffold_4702_curated|8 2881..84230|revcom	6.04	9.013	6.885	7.317	7.813	6.424	6.096	5.696

CasY4	4.73	2.981	5.344	4.713	4.778	4.863	5.518	5.887
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	8.795	10.581	8.543	9.318	9.03	8.543	8.401	8.372
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	7.166	8.289	8.458	8.143	8.224	8.581	7.827	8.359
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	8.095	8.496	8.804	8.76	9.349	8.878	8.333	8.217
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	8.217	10.291	8.373	7.813	7.559	6.951	7.562	7.852
Cas12h1	5.813	4.207	6.413	6.475	6.325	6.205	5.917	6.936
CasX1	7.026	5.981	6.9	6.191	6.263	5.741	6.076	5.906
CasX2	7.202	5.932	6.861	6.78	6.533	6.639	6.757	8.016
CasY1	5.641	3.751	4.666	4.625	4.972	5.249	6.141	7.182
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	9.903	9.919	9.402	10.517	9.879	11.092	10.611	9.365
Cas14u.7|3300001256.a|JGI122103113797_10004690|5792..7006	9.091	13.35	10.929	10.83	10.969	10.275	10.247	10.351
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	8.913	14.356	12.044	11.615	10.806	11.029	9.397	9.901
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	9.964	13.115	11.636	11.232	10.929	11.7	9.894	8.731
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	9.864	8.048	9.898	10.881	10.745	10.727	8.981	8.618
Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|revcom	9.414	12.319	9.222	9.369	9.42	8.696	10.783	8.483
CasY3	4.889	3.279	6.024	6.463	6.261	5.739	5.786	5.214
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	10.536	16.708	9.926	10.766	9.963	10.37	9.22	9.5
8971_2857_protein_locus_of_contig_OEJQ01000083.1 - Query protein (8971_2857)	9.827	14.286	11.858	10.02	8.946	9.381	10.667	8.955
9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)	9.811	13.874	11.799	10	8.949	9.375	10.634	8.958
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	7.372	7.881	6.602	6.897	6.393	6.579	8.401	7.176
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	9.222	15.99	10.478	10.256	10.019	10.575	10.48	8.968
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	6.656	6.818	7.967	8.007	8.02	8.183	8.293	9.37
CasY2	4.73	4.34	5.187	5.326	5.475	5.047	5.963	5.56

Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	11.058	11.364	11.519	10.316	12.02	11.39	12.841	11.42
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	11.185	10.664	11.356	9.241	10.248	10.282	11.675	11.22
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	10.851	10.913	12.034	8.911	10.248	9.453	11.675	11.87
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	42.708	10.681	16.723	15.92	16.279	16.5	19.224	19.677
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom		10.669	20.27	19.595	21.922	20.405	21.124	20.537
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|revcom	10.669		12.897	13.704	13.133	12.994	12.029	11.933
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	20.27	12.897		54.336	56.15	55.95	23.913	26.108
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	19.595	13.704	54.336		73.743	70.896	23.777	24.165
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	21.922	13.133	56.15	73.743		77.632	24.456	24.921
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	20.405	12.994	55.95	70.896	77.632		23.873	24.132
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	21.124	12.029	23.913	23.777	24.456	23.873		31.111
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	20.537	11.933	26.108	24.165	24.921	24.132	31.111	
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	21.626	10.764	24.463	23.453	25.081	24.032	31.759	42.479
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	19.495	16.427	27.602	26.637	27.765	26.411	32.118	38.636
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	30.841	22.488	45.146	41.063	44.444	42.995	53.241	70.588
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	30.049	22.222	45.128	40.306	44.898	43.367	52.683	69.43
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	13.324	7.792	13.108	13.15	14.574	12.735	11.864	12.624
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	11.51	10.4	13.546	14.353	13.777	13.622	15.152	13.025

Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	12.891	6.649	12.125	13.816	13.203	12.941	12.211	10.553
Cas14b.13|riftesphigho2_01_scaffold_82367_curated|1523..3856|revcom	11.494	7.208	11.765	12.844	12.37	11.979	11.795	11.139
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	13.077	9.431	15.147	15.335	15.848	14.263	15.822	14.074
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	14.6	10.483	15.397	16.066	15.285	15.122	14.33	12.254
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	14.396	10.333	12.711	15.798	15.994	15.024	15.373	12.236
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	9.414	17.115	11.314	11.151	11.84	11.883	10.106	10.114
Cas12c1	4.629	4.157	5.671	4.919	5.221	5.783	4.48	5.242
Cas12c2	4	3.12	3.827	3.782	4.603	4.603	4.841	5.12
Cas12a_UPI0001113398F	4.662	2.74	3.653	3.509	4.136	3.86	4.209	4.039
Cas12b_UPI0001113398F	4.662	2.74	3.653	3.509	4.136	3.86	4.209	4.039
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	4.662	2.742	3.653	3.506	4.132	3.857	4.209	4.032
Cas12a_UPI00083514A7	3.993	3.279	3.822	3.036	3.388	3.663	4.011	4.383
Cas12b_UPI00083514A7	3.993	3.279	3.822	3.036	3.388	3.663	4.011	4.383
Cas12a_UPI00097159F1	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12b_UPI00097159F1	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12a_UPI0009715A14	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12b_UPI0009715A14	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12a_UPI00097159CF	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12b_UPI00097159CF	4.735	2.796	4.186	3.857	4.026	4.588	4.007	3.85
Cas12a_UPI000832F6D2	4.36	2.889	4.089	3.665	3.835	4.303	4.19	4.029
Cas12b_UPI000832F6D2	4.36	2.889	4.089	3.665	3.835	4.303	4.19	4.029
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	4.267	2.889	4.182	3.665	3.742	4.21	4.19	4.304
OspCas12c	4.302	3.358	5.348	4.583	5.134	4.971	5.195	6.667
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	8.453	6.697	6.314	7.544	6.618	7.038	6.877	5.698
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	7.883	7.5	7.74	7.834	7.963	8.129	7.198	7.38
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	7.023	8.007	7.317	6.787	7.681	6.949	6.949	7.887

21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	6.583	8.789	5.376	7.492	7.187	6.585	7.309	6.994

Table 9

	Cas14b.9|3300013127.a|Ga0172365_1000442|633..2366|revcom	209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (2096 58_13971) (2)	209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	Cas14b.14|gwc3_scaffold_8732_curated|2705..4537	Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	Cas14b.13|rifcsphigho2_01_scaffold_8236_curated|1523..3856|revcom
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18189	10.852	11.434	21.344	20.661	8.04	8.09	8.391	8.545
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	9.041	8.289	13.074	13.91	7.412	8.85	7.859	9.06
Cas12i2	5.408	5.632	9.571	9.31	4.074	4.356	4.906	4.72
Cas12i1	5.07	4.11	5.621	5.288	4.384	4.093	6.029	5.326
Cas12g1	8.503	5.732	12.261	12.295	7.067	8.864	7.915	7.65
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	8.146	8.818	16.216	16.588	6.771	7.084	6.409	7.711
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	6.147	6.2	10.046	10.096	5.842	6.723	6.349	6.46
CasY5	4.732	3.591	6.757	6.516	3.704	3.951	4.228	3.887
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	10.174	8.733	13.531	12.329	7.393	7.345	7.078	7.441
CasY6	5.044	3.531	5.979	5.696	3.543	4.282	3.909	3.876
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	8.524	6.709	12.057	12.546	5.728	6.633	5.122	6.034
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	7.364	7.4	10.37	10.811	5.503	4.809	4.492	5.232

Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	9.076	11.616	16.667	16.129	8.423	9.56	6.076	6.378
Cas12a_UPI00094EEDB4	4.405	2.914	5.092	4.792	3.917	4.012	3.474	3.479
Cas12a_UPI000B4235CE	4.099	3.265	6.061	5.992	3.514	5.174	4.502	5.469
Cas12a_UPI000818CC52	4.099	3.265	6.061	5.992	3.514	5.174	4.508	5.477
Cas12a_UPI0007B78B7F	4.099	3.265	6.061	5.992	3.514	5.174	4.502	5.469
Cas12a_UPI000B4235F9	4.099	3.265	6.061	5.992	3.511	5.174	4.502	5.469
Cas14e.2|rifcsphowo2_01_scaffold_81231_curated|976..2217	8.483	7.305	9.417	9.434	6.636	6.467	6.049	6.12
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	6.874	7.532	10.909	11.005	7.302	7.165	5.122	5.837
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	5.769	7.071	10.502	10.096	5.521	7.87	5.398	4.967
CasY4	5.442	4.388	8.592	8.416	4.519	5.303	5.229	5.048
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	8.703	9.176	14	13.808	7.209	6.957	5.289	6.304
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	8.399	8.515	13.061	12.719	5.968	8.859	5.577	6.361
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	8.517	8.37	12.863	13.043	6.696	9.531	6.21	6.555
Cas14e.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	7.519	9.534	11.189	10.545	10.836	7.349	6.835	8.087
Cas12h1	6.746	5.522	8.434	8.202	6.466	3.913	5.509	5.943
CasX1	6.475	5.695	11.905	10.601	6.97	7.419	7.486	6.167
CasX2	8.091	6.032	12.04	10.764	7.446	7.369	6.907	6.997
CasY1	6.9	5.614	9.346	8.633	5.626	6.806	6.643	5.948
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	9.532	11.058	17.593	17.073	7.903	8.788	7.226	8.042
Cas14u.7|3300001256.a|JGI122103137971_10004690|5792..7006	11.379	14.481	20.657	20.297	9.6	7.741	7.642	6.762
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	9.54	13.812	18.224	17.241	8.682	8.019	6.162	7.004
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	10.374	12.963	19.725	19.807	8.786	8.805	5.905	6.986
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	8.483	10.448	15.962	14.851	6.38	8.116	7.031	7.833
Cas14c.2|3300001245.a|JGI12048313642_10201286|4257..5489|revcom	9.966	13.202	17.371	17.327	7.455	8.025	6.282	6.914
CasY3	7.087	4.834	9.487	9.019	7.18	5.766	6.567	6.833

633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	10	13.536	19.048	18.593	9.179	9.365	6.865	7.672
8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	8.511	13.165	17.143	16.08	10.14	7.731	7.173	7.714
9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)	8.523	13.165	17.143	16.08	9.949	7.921	7.202	7.736
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	8.58	9.015	15.164	14.592	5.832	7.447	5.625	7.098
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	9.343	12.707	17.788	17.259	8.838	10.841	7.171	7.867
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	6.75	8.294	11.814	11.062	5.433	6.871	5.941	5.882
CasY2	5.812	5.468	8.836	8.168	5.241	5.626	6.029	6.426
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	12.324	13.024	22.326	22.549	8.728	9.954	6.804	8.564
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	10.561	13.3	20.183	20.29	8.636	11.145	7.445	8.073
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	10.891	13.547	19.735	19.807	9.242	10.502	7.28	8.29
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	19.569	19.861	30.374	29.557	13.557	11.458	11.14	11.211
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	21.626	19.495	30.841	30.049	13.324	11.51	12.891	11.494
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|revcom	10.764	16.427	22.488	22.222	7.792	10.4	6.649	7.208
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	24.463	27.602	45.146	45.128	13.108	13.546	12.125	11.765
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	23.453	26.637	41.063	40.306	13.15	14.353	13.816	12.844
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	25.081	27.765	44.444	44.898	14.574	13.777	13.203	12.37
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	24.032	26.411	42.995	43.367	12.735	13.622	12.941	11.979
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	31.759	32.118	53.241	52.683	11.864	15.152	12.211	11.795
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	42.479	38.636	70.588	69.43	12.624	13.025	10.553	11.139

Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom		40.941	67.317	66.495	13	13.343	12.272	11.454
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	40.941		100	100	13.993	14.286	12.871	13.531
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	67.317	100		100	18.272	24.242	18.927	18.927
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	66.495	100	100		17.931	22.831	18.301	18.301
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	13	13.993	18.272	17.931		16.712	27.394	23.047
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	13.343	14.286	24.242	22.831	16.712		14.951	18.385
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	12.272	12.871	18.927	18.301	27.394	14.951		40.772
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	11.454	13.531	18.927	18.301	23.047	18.385	40.772	
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	14.286	16.364	26.126	25.592	18.759	21.333	19.549	20.411
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	15.123	15.565	24.554	23.944	18.091	23.263	19.798	19.898
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	14.701	14.468	24.554	23.944	17.236	22.87	19.75	21.673
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	10.152	12.983	19.005	19.048	7.932	8.73	6.727	6.43
Cas12c1	5.293	4.287	8.495	8.608	5.141	5.008	5.988	5.478
Cas12c2	4.519	4.063	8.753	8.611	3.878	3.897	5.064	5.263
Cas12a_UPI0011133 98F	4.479	3.345	6.516	5.605	5.328	5.481	4.476	5.171
Cas12b_UPI0011133 98F	4.479	3.345	6.516	5.605	5.328	5.481	4.476	5.171
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	4.388	3.341	6.497	5.588	5.236	5.476	4.476	5.254
Cas12a_UPI00083514A7	3.731	3.1	7.102	6.805	4.522	5.112	4.614	5.329
Cas12b_UPI00083514A7	3.731	3.1	7.102	6.805	4.522	5.112	4.614	5.329
Cas12a_UPI00097159F1	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12b_UPI00097159F1	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12a_UPI0009715A14	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12b_UPI0009715A14	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344

Cas12a_UPI00097159CF	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12b_UPI00097159CF	4.66	2.966	5.698	5.935	4.626	5.316	4.46	5.344
Cas12a_UPI000832F6D2	5.028	2.962	5.966	6.213	4.8	5.128	4.799	5.254
Cas12b_UPI000832F6D2	5.028	2.962	5.966	6.213	4.8	5.128	4.799	5.254
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	5.307	2.962	5.966	6.213	4.711	4.945	4.713	5.508
OspCas12c	5.537	4.028	7.71	7.477	4.309	5.263	5.016	4.71
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	8.213	5.056	9.412	10.084	5.078	6.46	5.788	4.436
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	8.756	6.681	9.449	9.877	4.762	5.532	5.388	4.326
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	5.615	5.749	8.365	8.13	5.321	6.601	4.316	5.179
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	6.175	6.098	8.812	8.8	6.241	6.268	5.604	5.062

Table 10

	Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	Cas14b.10|CG08_1and_8_20_14_0.20_scaffold_1609_curated|6134..7975	Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	Cas14u.3|3300000902.a|Ga0066793_10010091|37..1113|revcom	Cas12c1	Cas12e2	Cas12a_UPI00111339BF	Cas12b_UPI011133 98F
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	8.607	8.969	9.151	7.801	3.749	5.609	4.949	4.949
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	6.86	9.031	7.513	6.658	5.389	5.178	5.412	5.412
Cas12i2	5.529	4.981	4.803	2.761	5.444	5.988	7.131	7.131
Cas12i1	5.909	6.187	5.097	3.636	5.339	4.403	5.547	5.547
Cas12g1	9.554	8.217	8.805	6.992	5.582	5.954	5.649	5.649
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifesplowo2_01_scaffold_3495_curated|25656..27605|revcom	8.604	7.255	7.714	6.535	4.362	4.676	5.709	5.709

Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	8.247	6.647	7.829	7.085	3.803	4.073	5.372	5.372
CasY5	3.53	4.974	4.01	2.599	5.334	5.778	7.105	7.105
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	7.294	8.621	6.974	7.865	3.943	4.396	3.91	3.91
CasY6	4.444	4.167	4.567	2.972	7.076	6.856	7.015	7.015
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	8.161	7.412	7.263	6.276	5.155	4.448	6.356	6.356
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	7.123	7.613	6.589	7.279	3.681	3.598	4.2	4.2
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	9.385	8.661	9.291	8.884	3.421	4.153	2.899	2.899
Cas12a_UPI00094EEDB4	5.104	4.224	4.228	2.422	7.387	5.411	5.679	5.679
Cas12a_UPI000B4235CE	5.097	4.587	4.82	3.04	7.064	6.555	5.297	5.297
Cas12a_UPI000818CC52	5.104	4.671	4.904	3.04	7.074	6.564	5.233	5.233
Cas12a_UPI0007B78B7F	5.097	4.587	4.82	3.04	7.064	6.555	5.225	5.225
Cas12a_UPI000B4235F9	5.015	4.431	4.82	3.04	7.059	6.485	5.225	5.225
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	8.544	8.416	9.36	8.12	2.875	2.421	3.768	3.768
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	6.552	5.366	7.553	9.013	4.003	3.003	3.483	3.483
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	7.899	7.084	8.94	8.678	3.68	2.836	5.239	5.239
CasY4	5.401	5.755	5.356	3.168	6.734	5.498	6.737	6.737
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	7.553	7.951	7.034	11.469	3.969	3.997	4.758	4.758
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	5.655	6.212	7.251	7.005	3.965	3.846	5.206	5.206
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	5.891	7.187	8.346	7.951	4.196	4.01	4.668	4.668
Cas14e.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	8.921	8.837	7.965	7.129	3.75	3.207	3.856	3.856
Cas12h1	6.171	5.977	5.963	3.865	5.352	5.016	5.598	5.598
CasX1	6.612	6.984	7.419	5.456	7.083	6.63	6.371	6.371
CasX2	6.66	7.464	7.096	5.191	7.192	5.915	6.209	6.209
CasY1	6.818	6.828	5.951	4.048	7.049	5.659	5.166	5.166
Cas14u.3|19ft_2_nopbage_unknown_scaffold_0_curated|508188..509648	9.176	10.098	9.82	10.331	3.92	3.172	4.269	4.269

Cas14u.7|3300001256.a|JGI12210313797_1000469 0|5792..7006	7.865	10.231	9.091	12.528	3.259	3.185	3.249	3.249
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	8.64	8.553	8.224	14.151	3.016	3.598	3.96	3.96
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	8.28	10.164	8.867	13.122	3.41	3.434	3.156	3.156
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	8.1	7.98	7.516	8.876	4.177	4.362	4.779	4.779
Cas14c.2|3300001245.a|JGI12048313642_1020128 6|4257..5489|revcom	7.547	8.347	8.039	10.502	3.085	3.156	3.142	3.142
CasY3	5.056	5.702	5.541	3.643	6.218	7.863	5.779	5.779
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	8.9	8.099	8.609	13.318	3.819	3.275	3.258	3.258
8971_2857_protein_locus_of_contig_OEJQ010000 83.1 - Query protein (8971_2857)	9.424	9.386	9.567	13.384	3.541	3.226	2.486	2.486
9265_901_protein_locus_of_contig_OEFX0100000 5.1 - Query protein (9265_901)	9.589	9.381	9.381	13.022	3.509	3.283	2.554	2.554
Cas14u.6|3300006028.a|Ga0070717_10000077|5451 9..56201|revcom	7.264	10.502	8.976	7.584	4.286	4.424	5.068	5.068
466065_250_protein_locus_of_contig_SFKR01000 004.1 - Query protein (466065_250)	9.493	10.94	10.427	13.318	3.647	4.135	2.971	2.971
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	8.722	6.891	7.573	7.707	2.584	3.878	5.103	5.103
CasY2	5.719	5.491	5.008	3.336	6.014	6.632	6.712	6.712
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	11.502	10.129	10.129	9.982	4.106	5.117	5.418	5.418
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	9.969	8.654	9.807	13.069	4.466	5.518	4.288	4.288
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	9.502	9.206	8.931	12.871	4.203	5.184	5.077	5.077
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	13.509	13.744	11.765	8.834	4.24	4.854	4.117	4.117
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	13.077	14.6	14.396	9.414	4.629	4	4.662	4.662
Cas14u.2|3300002172.a|JGI24730326740_1002785|496..1605|revcom	9.431	10.483	10.333	17.115	4.157	3.12	2.74	2.74
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	15.147	15.397	12.711	11.314	5.671	3.827	3.653	3.653

Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	15.335	16.066	15.798	11.151	4.919	3.782	3.509	3.509
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	15.848	15.285	15.994	11.84	5.221	4.603	4.136	4.136
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	14.263	15.122	15.024	11.883	5.783	4.603	3.86	3.86
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	15.822	14.33	15.373	10.106	4.48	4.841	4.209	4.209
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	14.074	12.254	12.236	10.114	5.242	5.12	4.039	4.039
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	14.286	15.123	14.701	10.152	5.293	4.519	4.479	4.479
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	16.364	15.565	14.468	12.983	4.287	4.063	3.345	3.345
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	26.126	24.554	24.554	19.005	8.495	8.753	6.516	6.516
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	25.592	23.944	23.944	19.048	8.608	8.611	5.605	5.605
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	18.759	18.091	17.236	7.932	5.141	3.878	5.328	5.328
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	21.333	23.263	22.87	8.73	5.008	3.897	5.481	5.481
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2889|revcom	19.549	19.798	19.75	6.727	5.988	5.064	4.476	4.476
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	20.411	19.898	21.673	6.43	5.478	5.263	5.171	5.171
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627		30.901	31.394	7.581	4.864	5.033	4.41	4.41
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	30.901		46.582	9	4.265	5.359	4.715	4.715
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	31.394	46.582		7.667	4.657	4.455	4.267	4.267
Cas14u.1|3300009029.a|Ga0066793_10010091|37..3113|revcom	7.581	9	7.667		3.05	3.193	3.768	3.768
Cas12c1	4.864	4.265	4.657	3.05		10.725	6.353	6.353
Cas12e2	5.033	5.359	4.455	3.193	10.725		6.867	6.867

Cas12a_UPI0001133988F	4.41	4.715	4.267	3.768	6.353	6.867		100
Cas12b_UPI0001133988F	4.41	4.715	4.267	3.768	6.353	6.867	100	
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	4.586	4.711	4.085	3.952	6.334	6.809	93.916	93.916
Cas12a_UPI00083514A7	4.301	5.221	5.142	3.571	6.796	6.507	52.754	52.754
Cas12b_UPI00083514A7	4.301	5.221	5.142	3.571	6.796	6.507	52.754	52.754
Cas12a_UPI00097159F1	4.312	4.801	4.265	4.124	6.796	6.274	51.817	51.817
Cas12b_UPI00097159F1	4.312	4.801	4.265	4.124	6.796	6.274	51.817	51.817
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.312	4.801	4.265	4.124	6.796	6.274	51.817	51.817
Cas12a_UPI0009715A14	4.312	4.801	4.265	4.124	6.791	6.274	51.557	51.557
Cas12b_UPI0009715A14	4.312	4.801	4.265	4.124	6.791	6.274	51.557	51.557
Cas12a_UPI00097159CF	4.312	4.801	4.265	4.124	6.791	6.274	51.73	51.73
Cas12b_UPI00097159CF	4.312	4.801	4.265	4.124	6.791	6.274	51.73	51.73
Cas12a_UPI000832F6D2	4.216	4.887	4.533	4.221	6.572	6.042	51.513	51.513
Cas12b_UPI000832F6D2	4.216	4.887	4.533	4.221	6.572	6.042	51.513	51.513
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	4.216	4.615	4.352	4.311	6.497	5.887	51.685	51.685
OspCas12e	4.835	4.75	4.593	3.102	7.138	7.704	5.243	5.243
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	5.501	7.203	7.433	5.706	3.739	4.269	5.596	5.596
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	6.021	6.466	7.292	7.19	3.262	3.621	4.818	4.818
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	6.676	6.686	6.765	6.139	4.344	4.534	4.932	4.932
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	5.333	7.669	6.897	8.086	3.21	4.105	5.105	5.105

Table 11

	Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	Cas12a_UPI00083514A7	Cas12b_UPI00083514A7	Cas12a_UPI00097159F1	Cas12b_UPI00097159F1	Cas12b_sp|T0D7A2|CS12B_ALIAG	Cas12a_UPI0009715A14	Cas12b_UPI0009715A14
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	4.818	5.013	5.013	4.865	4.865	4.865	4.865	4.865

Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	5.541	5.917	5.917	6.396	6.396	6.396	6.396	6.396
Cas12i2	7.248	5.824	5.824	6.03	6.03	6.03	6.03	6.03
Cas12i1	5.708	5.837	5.837	5.934	5.934	5.934	5.934	5.934
Cas12g1	5.434	5.986	5.986	5.845	5.845	5.845	5.935	5.935
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	5.585	5.254	5.254	5.1	5.1	5.1	5.1	5.1
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	5.461	5.085	5.085	5.743	5.743	5.743	5.743	5.743
CasY5	7.186	6.941	6.941	6.921	6.921	6.921	6.838	6.838
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	3.747	4.391	4.391	5.165	5.165	5.165	5.165	5.165
CasY6	6.942	6.428	6.428	6.133	6.133	6.133	6.058	6.058
Cas14f.3|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	6.394	6.014	6.014	6.324	6.324	6.324	6.324	6.324
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	4.259	4.541	4.541	4.558	4.558	4.558	4.649	4.649
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	2.893	4.159	4.159	2.69	2.69	2.69	2.69	2.69
Cas12a_UPI00094EEDB4	5.575	6.026	6.026	6.82	6.82	6.82	6.82	6.82
Cas12a_UPI000B4235CE	5.323	5.583	5.583	6.017	6.017	6.017	6.017	6.017
Cas12a_UPI000818CC52	5.259	5.448	5.448	5.882	5.882	5.882	5.882	5.882
Cas12a_UPI0007B78B7F	5.252	5.44	5.44	5.874	5.874	5.874	5.874	5.874
Cas12a_UPI000B4235F9	5.252	5.512	5.512	5.946	5.946	5.946	5.946	5.946
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	3.772	3.846	3.846	4.03	4.03	4.03	4.03	4.03
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313	3.388	3.822	3.822	3.825	3.825	3.825	3.825	3.825
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	5.133	4.388	4.388	5.717	5.717	5.717	5.717	5.717
CasY4	6.546	5.998	5.998	5.998	5.998	5.998	6.074	6.074
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	4.633	4.112	4.112	4.093	4.093	4.122	4.122	4.122
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	5.306	4.749	4.749	5.225	5.225	5.225	5.225	5.225
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	4.852	4.659	4.659	5.133	5.133	5.133	5.133	5.133

Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20888|revcom	3.665	4.087	4.087	4.452	4.452	4.452	4.452	4.452
Cas12h1	5.763	5.64	5.64	6.374	6.374	6.374	6.374	6.374
CasX1	6.31	6.034	6.034	5.916	5.916	5.916	5.916	5.916
CasX2	5.882	5.705	5.705	5.412	5.412	5.412	5.329	5.329
CasY1	5.183	5.624	5.624	4.867	4.867	4.867	4.867	4.867
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	4.269	3.993	3.993	4.457	4.457	4.457	4.457	4.457
Cas14u.7|3300001256.a|3GI122103113797_10004690|5792..7006	3.237	3.584	3.584	3.306	3.306	3.306	3.214	3.214
Cas14u.8|3300005660.a|Gs0073904_10021651|765..1943	3.957	3.136	3.136	2.661	2.661	2.661	2.661	2.661
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	3.055	3.232	3.232	2.663	2.663	2.663	2.663	2.663
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	4.867	4.487	4.487	4.503	4.503	4.503	4.503	4.503
Cas14e.2|3300001245.a|3GI12048J13642_10201286|4257..5489|revcom	3.139	2.594	2.594	3.294	3.294	3.294	3.294	3.294
CasY3	5.807	6.591	6.591	6.298	6.298	6.298	6.298	6.298
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	3.348	2.599	2.599	3.643	3.643	3.578	3.578	3.578
8971_2857_protein_locus_of_contig_OEJQ01000083.1 - Query protein (8971_2857)	2.481	2.657	2.657	2.242	2.242	2.242	2.242	2.242
9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)	2.55	2.723	2.723	2.314	2.314	2.314	2.314	2.314
Cas14u.6|3300006028.a|Gs0070717_10000677|54519..56201|revcom	5.158	4.599	4.599	4.428	4.428	4.428	4.428	4.428
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	3.058	2.308	2.308	2.844	2.844	2.844	2.844	2.844
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.169	4.728	4.728	5.302	5.302	5.302	5.302	5.302
CasY2	6.642	5.927	5.927	6.616	6.616	6.656	6.656	6.656
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	5.142	4.487	4.487	4.69	4.69	4.69	4.69	4.69
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	4.189	4.455	4.455	4.944	4.944	4.944	4.944	4.944
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	4.977	4.517	4.517	5.097	5.097	5.097	5.097	5.097

Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	4.026	4.45	4.45	3.911	3.911	3.911	3.911	3.911
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	4.662	3.993	3.993	4.735	4.735	4.735	4.735	4.735
Cas14u.2|3300002172.a|JGI24736J26740_1002785|496..1605|revcom	2.742	3.279	3.279	2.796	2.796	2.796	2.796	2.796
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	3.653	3.822	3.822	4.186	4.186	4.186	4.186	4.186
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	3.506	3.036	3.036	3.857	3.857	3.857	3.857	3.857
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	4.132	3.388	3.388	4.026	4.026	4.026	4.026	4.026
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	3.857	3.663	3.663	4.588	4.588	4.588	4.588	4.588
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	4.209	4.011	4.011	4.007	4.007	4.007	4.007	4.007
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	4.032	4.383	4.383	3.85	3.85	3.85	3.85	3.85
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	4.388	3.731	3.731	4.66	4.66	4.66	4.66	4.66
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	3.341	3.1	3.1	2.966	2.966	2.966	2.966	2.966
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	6.497	7.102	7.102	5.698	5.698	5.698	5.698	5.698
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	5.588	6.805	6.805	5.935	5.935	5.935	5.935	5.935
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	5.236	4.522	4.522	4.626	4.626	4.626	4.626	4.626
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	5.476	5.112	5.112	5.316	5.316	5.316	5.316	5.316
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2889|revcom	4.476	4.614	4.614	4.46	4.46	4.46	4.46	4.46
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	5.254	5.329	5.329	5.344	5.344	5.344	5.344	5.344
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	4.586	4.301	4.301	4.312	4.312	4.312	4.312	4.312

Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	4.711	5.221	5.221	4.801	4.801	4.801	4.801	4.801
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	4.085	5.142	5.142	4.265	4.265	4.265	4.265	4.265
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	3.952	3.571	3.571	4.124	4.124	4.124	4.124	4.124
Cas12c1	6.334	6.796	6.796	6.796	6.796	6.796	6.791	6.791
Cas12c2	6.809	6.507	6.507	6.274	6.274	6.274	6.274	6.274
Cas12a_UPI001113398F	93.916	52.754	52.754	51.817	51.817	51.817	51.557	51.557
Cas12b_UPI001113398F	93.916	52.754	52.754	51.817	51.817	51.817	51.557	51.557
Cas12b_tr|A0A1I7F1U9|A0A1I7F1U9_9BACL		50.676	50.676	49.661	49.661	49.661	49.407	49.407
Cas12a_UPI00083514A7	50.676		100	55.45	55.45	55.45	55.19	55.19
Cas12b_UPI00083514A7	50.676	100		55.45	55.45	55.45	55.19	55.19
Cas12a_UPI00097159F1	49.661	55.45	55.45		100	100	99.734	99.734
Cas12b_UPI00097159F1	49.661	55.45	55.45	100		100	99.734	99.734
Cas12b_sp|T0D7A2|CS12B_ALIAG	49.661	55.45	55.45	100	100		99.734	99.734
Cas12a_UPI0009715A14	49.407	55.19	55.19	99.734	99.734	99.734		100
Cas12b_UPI0009715A14	49.407	55.19	55.19	99.734	99.734	99.734	100	
Cas12a_UPI00097159CF	49.576	55.363	55.363	99.911	99.911	99.911	99.823	99.823
Cas12b_UPI00097159CF	49.576	55.363	55.363	99.911	99.911	99.911	99.823	99.823
Cas12a_UPI000832F6D2	49.619	55.796	55.796	93.546	93.546	93.546	93.28	93.28
Cas12b_UPI000832F6D2	49.619	55.796	55.796	93.546	93.546	93.546	93.28	93.28
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	49.619	55.969	55.969	92.838	92.838	92.838	92.573	92.573
OspCas12c	5.283	6.169	6.169	5.864	5.864	5.864	5.864	5.864
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	5.42	5.121	5.121	5.796	5.796	5.796	5.796	5.796
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	4.914	4.163	4.163	5.005	5.005	5.005	5.097	5.097
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	5.027	4.277	4.277	4.753	4.753	4.753	4.753	4.753
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	5.1	4.628	4.628	3.993	3.993	3.993	3.9	3.9

Table 12

	Cas12a_UPI0009715 9CF	Cas12b_UPI0009715 9CF	Cas12a_UPI0009832F6D2	Cas12b_UPI0009832F6D2	Cas12b_tr|A0A5I2CSX2|A0A5I2CSX2_9BACL	OspCas12c	Cas14a.5|3300012532.a|Ga0137373_1000316|3286..5286	63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18189	4.865	4.865	4.861	4.861	5.122	5.082	6.658	5.931
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	6.396	6.396	6.218	6.218	5.959	6.075	8.752	7.333
Cas12i2	6.03	6.03	6.114	6.114	5.946	5.914	4.39	3.933
Cas12i1	5.934	5.934	6.008	6.008	5.692	5.588	4.128	2.982
Cas12g1	5.935	5.935	5.75	5.75	5.93	5.657	9.103	6.91
Cas14d.3|RIFCSPLOWO2_01_FULL_ODI_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	5.1	5.1	5.369	5.369	5.096	5.251	8.21	7.211
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	5.743	5.743	6.011	6.011	6.011	3.54	7.283	6.686
CasY5	6.915	6.915	6.843	6.843	7.076	4.853	5.804	4.204
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829	5.165	5.165	5.33	5.33	5.161	4.021	6.591	5.284
CasY6	6.133	6.133	6.502	6.502	6.353	7.595	5.418	3.692
Cas14f.3|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	6.324	6.324	6.416	6.416	6.416	5.314	6.436	7.015
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	4.558	4.558	4.831	4.831	4.649	4.073	5.503	7.794
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	2.69	2.69	2.966	2.966	2.966	3.471	6.078	5.063
Cas12a_UPI00094EEDB4	6.82	6.82	6.671	6.671	6.671	6.104	3.74	2.937
Cas12a_UPI000B4235CE	6.017	6.017	5.87	5.87	5.941	7.567	4.064	3.303
Cas12a_UPI000818CC52	5.882	5.882	5.735	5.735	5.806	7.436	4.064	3.303
Cas12a_UPI0007B78B7F	5.874	5.874	5.727	5.727	5.798	7.426	4.064	3.303
Cas12a_UPI000B4235F9	5.946	5.946	5.798	5.798	5.87	7.567	4.064	3.303
Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	4.03	4.03	4.213	4.213	4.213	2.922	4.154	5.096
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113969..114313	3.825	3.825	3.918	3.918	3.731	3.084	6.37	5.949
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	5.717	5.717	5.524	5.524	5.337	3.328	6.038	5.512

CasY4	6.074	6.074	6.226	6.226	6.302	5.58	5.068	4.017
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	4.122	4.122	3.939	3.939	4.029	3.325	6.96	6.192
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	5.225	5.225	5.316	5.316	5.316	4.133	9.531	7.657
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	5.133	5.133	5.225	5.225	5.133	4.708	8.417	7.055
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	4.452	4.452	4.27	4.27	4.27	3.503	4.032	4.928
Cas12h1	6.374	6.374	5.938	5.938	5.766	5.263	6.749	6.082
CasX1	5.916	5.916	6.076	6.076	5.993	5.792	6.016	4.187
CasX2	5.412	5.412	5.74	5.74	5.657	6.386	5.731	5.348
CasY1	4.867	4.867	5.102	5.102	5.102	6.691	5.818	3.931
Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188..509648	4.457	4.457	4.731	4.731	4.453	4.214	6.287	7.981
Cas14u.7|3300001256.a|JGI122103113797_10004690|5792..7006	3.306	3.306	3.394	3.394	3.394	3.339	5.589	4.754
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	2.661	2.661	2.75	2.75	2.841	3.496	6.938	7.084
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	2.663	2.663	2.849	2.849	2.755	2.685	5.556	5.307
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	4.503	4.503	4.592	4.592	4.592	3.504	5.588	6.907
Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|revcom	3.294	3.294	3.294	3.294	3.294	3.89	6.577	6.743
CasY3	6.298	6.298	6.523	6.523	6.37	7.179	4.038	3.362
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	3.578	3.578	3.483	3.483	3.391	2.941	5.918	6.988
8971_2857_protein_locus_of_contig_OEJQ01000083.1 - Query protein (8971_2857)	2.242	2.242	2.045	2.045	2.142	3.38	6.988	5.302
9265_901_protein_locus_of_contig_OEFX01000005.1 - Query protein (9265_901)	2.314	2.314	2.119	2.119	2.216	3.519	7.026	5.197
Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	4.428	4.428	4.7	4.7	4.885	4.217	8.626	8.15
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	2.844	2.844	2.746	2.746	2.841	3.859	6.991	5.351
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	5.302	5.302	5.297	5.297	5.205	2.885	4.119	5.14
CasY2	6.656	6.656	6.886	6.886	6.58	5.808	4.227	4.503

Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	4.69	4.69	4.592	4.592	4.686	4.327	7.225	9.451
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	4.944	4.944	4.846	4.846	4.939	4.302	6.755	6.656
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	5.097	5.097	4.907	4.907	5.093	4.383	6.461	6.815
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	3.911	3.911	3.814	3.814	3.907	4.475	8.346	7.309
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	4.735	4.735	4.36	4.36	4.267	4.302	8.453	7.883
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|revcom	2.796	2.796	2.889	2.889	2.889	3.358	6.697	7.5
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	4.186	4.186	4.089	4.089	4.182	5.348	6.314	7.74
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	3.857	3.857	3.665	3.665	3.665	4.583	7.544	7.834
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	4.026	4.026	3.835	3.835	3.742	5.134	6.618	7.963
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	4.588	4.588	4.303	4.303	4.21	4.971	7.038	8.129
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	4.007	4.007	4.19	4.19	4.19	5.195	6.877	7.198
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	3.85	3.85	4.029	4.029	4.304	6.667	5.698	7.38
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	4.66	4.66	5.028	5.028	5.307	5.537	8.213	8.756
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	2.966	2.966	2.962	2.962	2.962	4.028	5.056	6.681
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	5.698	5.698	5.966	5.966	5.966	7.71	9.412	9.449
209660_51257_protein_locus_of_contig_Ga0190335_1015156 - Query protein (209660_51257) (2)	5.935	5.935	6.213	6.213	6.213	7.477	10.084	9.877
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	4.626	4.626	4.8	4.8	4.711	4.309	5.078	4.762
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	5.316	5.316	5.128	5.128	4.945	5.263	6.46	5.532

Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom	4.46	4.46	4.799	4.799	4.713	5.016	5.788	5.388
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	5.344	5.344	5.254	5.254	5.508	4.71	4.436	4.326
Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627	4.312	4.312	4.216	4.216	4.216	4.835	5.501	6.021
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	4.801	4.801	4.887	4.887	4.615	4.75	7.203	6.466
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	4.265	4.265	4.533	4.533	4.352	4.593	7.433	7.292
Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom	4.124	4.124	4.221	4.221	4.311	3.102	5.706	7.19
Cas12c1	6.791	6.791	6.572	6.572	6.497	7.138	3.739	3.262
Cas12c2	6.274	6.274	6.042	6.042	5.887	7.704	4.269	3.621
Cas12a_UPI0001113398F	51.73	51.73	51.513	51.513	51.685	5.243	5.596	4.818
Cas12b_UPI0001113398F	51.73	51.73	51.513	51.513	51.685	5.243	5.596	4.818
Cas12b_tr|A0A117F1U9|A0A117F1U9_9BACL	49.576	49.576	49.619	49.619	49.619	5.283	5.42	4.914
Cas12a_UPI00083514A7	55.363	55.363	55.796	55.796	55.969	6.169	5.121	4.163
Cas12b_UPI00083514A7	55.363	55.363	55.796	55.796	55.969	6.169	5.121	4.163
Cas12a_UPI00097159F1	99.911	99.911	93.546	93.546	92.838	5.864	5.796	5.005
Cas12b_UPI00097159F1	99.911	99.911	93.546	93.546	92.838	5.864	5.796	5.005
Cas12b_sp|T0D7A2|CS12B_ALIAG	99.911	99.911	93.546	93.546	92.838	5.864	5.796	5.005
Cas12a_UPI0009715A14	99.823	99.823	93.28	93.28	92.573	5.864	5.796	5.097
Cas12b_UPI0009715A14	99.823	99.823	93.28	93.28	92.573	5.864	5.796	5.097
Cas12a_UPI00097159CF		100	93.457	93.457	92.75	5.864	5.796	5.097
Cas12b_UPI00097159CF	100		93.457	93.457	92.75	5.864	5.796	5.097
Cas12a_UPI000832F6D2	93.457	93.457		100	95.664	5.941	5.974	4.727
Cas12b_UPI000832F6D2	93.457	93.457	100		95.664	5.941	5.974	4.727
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	92.75	92.75	95.664	95.664		5.788	5.79	4.912
OspCas12c	5.864	5.864	5.941	5.941	5.788		3.769	3.395
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	5.796	5.796	5.974	5.974	5.79	3.769		21.912
63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	5.097	5.097	4.727	4.727	4.912	3.395	21.912	
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	4.753	4.753	4.66	4.66	4.753	3.325	21.358	38.208

21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	3.9	3.9	3.993	3.993	4.085	4.065	23.547	36.783

Table 13

	58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)	21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)
Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180	6.989	6.465
Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123	8.614	7.995
Cas12i2	3.599	3.937
Cas12i1	3.458	3.451
Cas12g1	6.914	8.56
Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom	7.487	6.098
Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom	7.55	6.676
CasY5	4.856	4.668
Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20986_curated|649..2829	7.097	6.684
CasY6	3.668	3.462
Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041	6.435	5.92
Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348	6.984	6.726
Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734	5.91	6.171
Cas12a_UPI00094EEDB4	4.321	3.181
Cas12a_UPI000B4235CE	3.988	3.627
Cas12a_UPI000818CC52	3.988	3.627
Cas12a_UPI0007B78B7F	3.988	3.627
Cas12a_UPI000B4235F9	3.988	3.627

Cas14c.2|rifcsplowo2_01_scaffold_81231_curated|976..2217	4.416	5.76
Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113969..114313	6.19	6.924
Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom	4.212	4.944
CasY4	4.693	4.014
Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504	7.099	8.791
Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom	8.769	7.351
Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081	7.154	7.87
Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curated|19327..20880|revcom	5.24	5.294
Cas12h1	6.176	6.007
CasX1	5.123	4.266
CasX2	5.184	4.418
CasY1	4.182	4.771
Cas14u.3|19ft_2_nophage_nokmown_scaffold_0_curated|508188..509648	6.955	7.442
Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006	6.139	5.785
Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943	7.792	6.988
Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..85657	4.693	5.473
Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom	7.121	5.643
Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|revcom	7.27	7.82
CasY3	3.531	2.431
633299_527_protein_locus_of_contig_Scffd15 - Query protein (633299_527) (4)	7.143	6.425
8971_2857_protein_locus_of_contig_OEJQ010000083.1 - Query protein (8971_2857)	6.329	5.935
9265_901_protein_locus_of_contig_OEFX010000065.1 - Query protein (9265_901)	6.206	5.82

Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom	8.423	7.402
466065_250_protein_locus_of_contig_SFKR01000004.1 - Query protein (466065_250)	6.931	6.187
Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521	4.695	4.409
CasY2	3.976	4.174
Cas14a.3|gwa1_scaffold_1795_curated|25635..27224|revcom	6.577	7.553
Cas14a.1|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom	6.211	6.667
Cas14a.2|gwa2_scaffold_18027_curated|7105..8628	5.745	7.302
Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155	5.828	6.202
Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom	7.023	6.583
Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|revcom	8.007	8.789
Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217	7.317	5.376
Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983	6.787	7.492
Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257	7.681	7.187
Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom	6.949	6.585
Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598	6.949	7.309
Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634..8466|revcom	7.887	6.994
Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom	5.615	6.175
209658_13971_protein_locus_of_contig_Ga0190333_1001561 - Query protein (209658_13971) (2)	5.749	6.098
209657_57738_protein_locus_of_contig_Ga0190332_1015597 - Query protein (209657_57738) (2)	8.365	8.812

209660_51257_protein_locus_of_contig_Ga0196335_1015156 - Query protein (209660_51257) (2)	8.13	8.8
Cas14b.14|gwc1_scaffold_8732_curated|2705..4537	5.321	6.241
Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032	6.601	6.268
Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2886|revcom	4.316	5.604
Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom	5.179	5.062
Cas14b.16|3300005573.a|Ga0078972_1001015a|33756..35627	6.676	5.333
Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975	6.686	7.669
Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219	6.765	6.897
Cas14u.1|3300009029.a|Ga0066793_10010091|37..3113|revcom	6.139	8.086
Cas12c1	4.344	3.21
Cas12c2	4.534	4.105
Cas12a_UPI0001113398F	4.932	5.105
Cas12b_UPI0001113398F	4.932	5.105
Cas12b_tr|A0A1I7F1U9|A0A1I7F1U9_9BACL	5.027	5.1
Cas12a_UPI00083514A7	4.277	4.628
Cas12b_UPI00083514A7	4.277	4.628
Cas12a_UPI00097159F1	4.753	3.993
Cas12b_UPI00097159F1	4.753	3.993
Cas12b_sp|T0D7A2|CS12B_ALIAG	4.753	3.993
Cas12a_UPI0009715A14	4.753	3.9
Cas12b_UPI0009715A14	4.753	3.9
Cas12a_UPI00097159CF	4.753	3.9
Cas12b_UPI00097159CF	4.753	3.9
Cas12a_UPI000832F6D2	4.66	3.993
Cas12b_UPI000832F6D2	4.66	3.993
Cas12b_tr|A0A512CSX2|A0A512CSX2_9BACL	4.753	4.085
OspCas12c	3.325	4.065
Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286	21.358	23.547

63461_4106_protein_locus_of_contig_LSKL01000323 - Query protein (63461_4106) translation (4)	38.208	36.783
58610_1188_protein_locus_of_contig_LFOD01000003 - Query protein (58610_1188) translation (5)		31.115
21566_3969_protein_locus_of_contig_BAFB01000202 - Query protein (21566_3969) translation (4)	31.115	

Table 14

5' modification	
SEQ ID NO: 145 5pr_trunc_4	GTTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 146 5pr_trunc_5	GTATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 147 5pr_trunc_6	GATGCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 148 5pr_trunc_7	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SL1 modification	
SEQ ID NO: 149 SL1_modification_1	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

SEQ ID NO: 150 SL1_modification_2	GCTCCACTTTACTAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 151 SL1_modification_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 152 SL1_modification_4	GCTCCACTTTAATAAGTGGAGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 153 SL1_modification_5	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 154 SL1_modification_6	GTGCTCCACTTTAATAAGTGGTGCATTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 155 SL1_modification_7	GCTCCACTTGTAATCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 156 SL1_modification_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

SEQ ID NO: 157 SL1_modification_9	GCTCCACTTGGCTAATGCCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 158 SL1_modification_10	GCTCCACTTGGCATAATTGCCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 159 SL1_MS2_hp	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SL2 modification	
SEQ ID NO: 160 SL2_modification_1	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTAATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 161 SL2_modification_2	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTAAATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 162 SL2_modification_3	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCCTATATGGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 163 SL2_modification_4	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

SEQ ID NO: 164 SL2_modification_5	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 165 SL2_modification_6	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTTATATAGCAGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 166 SL2_modification_7	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTGTATATCAGCAGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 167 SL2_MS2_hp	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCACATGAGGATCACCCATGTGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SL3 modification	
SEQ ID NO: 168 increase_interaction_w_crRNA_13	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGATTGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 169 increase_interaction_w_crRNA_14	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCACGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAGTGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 170	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGACTGCAACTGGTTGCCCACCCTAGTCATTG

increase_interaction_w_crRNA_15	
SEQ ID NO: 171 increase_interaction_w_crRNA_16	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG TTGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTC AAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 172 increase_interaction_w_crRNA_17	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG TTCGATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTC AAGGAATCGAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 173 increase_interaction_w_crRNA_18	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG TTGAGTGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTC AAGGAACTCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 174 increase_interaction_w_crRNA_19	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG TTGCGTGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTC AAGGAACGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 175 increase_interaction_w_crRNA_20	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG TTGTATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTC AAGGAATACAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 176	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTG CCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGT CAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG

increase_interaction_w_crRNA_21	
SEQ ID NO: 177 increase_interaction_w_crRNA_22	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 178 increase_interaction_w_crRNA_23	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCGGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAACGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 179 increase_interaction_w_crRNA_24	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGTAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATACAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 180 increase_interaction_w_crRNA_25	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 181 increase_interaction_w_crRNA_26	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SL4 modification	

SEQ ID NO: 182 increase_interaction_of_SL4_3	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACGCTAGACGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 183 increase_interaction_of_SL4_4	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACTGCTAGACAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 184 increase_interaction_of_SL4_5	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTGCTAGACAGGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 185 increase_interaction_of_SL4_6	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACGCTCAGACGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 186 increase_interaction_of_SL4_7	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACTGCTCAGACAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
increase_interaction_of_SL4_8	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTGCTCAGACAGGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 187	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACGCTGCTCAGACAGCGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

increase_interaction_of_SL4_9	
SEQ ID NO: 188 increase_interaction_of_SL4_10	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACTGCTGCTCAGACAGCAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 189 SL3_MS2_bp	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACACATGAGGATCACCCATGTGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SL5 modification	
SEQ ID NO: 190 increase_interaction_of_SL5_4	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 191 increase_interaction_of_SL5_5	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAGCTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 192 increase_interaction_of_SL5_6	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGCTAAAAGAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 193	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAAGCATAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

increase_interaction_of_SL5_7	
SEQ ID NO: 194 increase_interaction_of_SL5_8	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGCTGAAAAGCAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 195 increase_interaction_of_SL5_9	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGCTGAAAAGCAGCTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 196 increase_interaction_of_SL5_10	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGCTGAAAAGCAGCATAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 197 SL4_MS2_hp	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTACATGAGGATCACCCATGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 198 sgRNA version3.2	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG

Table 15

PsaCas12f construct name	Location of N-termini (amino acid position)
cpPsaCas12f_1	I77

cpPsaCas12f_2	N104
cpPsaCas12f_3	P146
cpPsaCas12f_4	E224
cpPsaCas12f_5	N266
cpPsaCas12f_6	D375
cpPsaCas12f_7	K349
cpPsaCas12f_8	K55
cpPsaCas12f_9	537K
cpPsaCas12f_10	A407
cpPsaCas12f_11	R216
cpPsaCas12f_12	N520

Table 16

SEQ ID NO: 199 5pr_trunc_7-B12 (= increase_interaction_w_crRNA_21)	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 200 SL1_modification_1 + increase_interaction_w_crRNA_21	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 201 SL1_modification_3 + increase_interaction_w_crRNA_21	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 202 SL1_modification_5 + increase_interaction_w_crRNA_21	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 203 SL1_modification_8 + increase_interaction_w_crRNA_21_sgRNA 3.1	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 204	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG

SL1_MS2_hp + increase_interaction_ w_crRNA_21	
SEQ ID NO: 205 5pr_trunc_7 + increase_interaction_ w_crRNA_22	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAA AAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 206 SL1_modification_1 + increase_interaction_ w_crRNA_22	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAA AAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 207 SL1_modification_3 + increase_interaction_ w_crRNA_22	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAA AAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SL1_modification_5 + increase_interaction_ w_crRNA_22	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGG ATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTAT TGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 208 SL1_modification_8 + increase_interaction_ w_crRNA_22	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTATAT GCTGAGGGAGGATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGT ATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACC CTAGTCATTG
SEQ ID NO: 209 SL1_MS2_hp + increase_interaction_ w_crRNA_22	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGG CGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 210 5pr_trunc_7 + increase_interaction_ w_crRNA_25	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 211 SL1_modification_1 + increase_interaction_ w_crRNA_25	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 212 SL1_modification_3 + increase_interaction_ w_crRNA_25	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 213	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGG ATGGGCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG

SL1_modification_5 + increase_interaction_ w_crRNA_25	
SEQ ID NO: 214 SL1_modification_8 + increase_interaction_ w_crRNA_25	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTATAT GCTGAGGGAGGATGGGCGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTA TCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCC TAGTCATTG
SEQ ID NO: 215 SL1_MS2_hp + increase_interaction_ w_crRNA_25	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGG CGCTGCCGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAA GTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 216 5pr_trunc_7 + increase_interaction_ w_crRNA_26	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 217 SL1_modification_1 + increase_interaction_ w_crRNA_26	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 218 SL1_modification_3 + increase_interaction_ w_crRNA_26	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 219 SL1_modification_5 + increase_interaction_ w_crRNA_26	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGG ATGGGCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATT GAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 220 SL1_modification_8 + increase_interaction_ w_crRNA_26	GCTCCACTTACATGAGGATCACCCATGTAAGTGGTGCCTTCCAAAGCTATAT GCTGAGGGAGGATGGGCGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTA TCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCC TAGTCATTG
SEQ ID NO: 221 SL1_MS2_hp + increase_interaction_ w_crRNA_26	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGG CGCTGCGGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAA GTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 222 best_guide_v2	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGG GCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAA AGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

Table 17

SEQ ID NO: 223 EMX_Cas12fg_2	TCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 224 EMX_Cas12fg_3	GCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 225 EMX1-stagger_25	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTGGAG
SEQ ID NO: 226 EMX1-stagger_24	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTGGA
SEQ ID NO: 227 EMX1-stagger_23	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTGG
SEQ ID NO: 228 EMX1-stagger_22	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATT

NO: 229 EMX1 - stagger 21	
SEQ ID NO: 230 EMX1 - stagger 20	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTG CAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAA TGCAACTGGTTGCCCACCCTAGTCAT
SEQ ID NO: 231 EMX1 - stagger 19	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTG CAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAA TGCAACTGGTTGCCCACCCTAGTCA
SEQ ID NO: 232 EMX1 - stagger 18	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTG CAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAA TGCAACTGGTTGCCCACCCTAGTC

Table 18

SEQ ID NO: 233 Cas12f_intraprotein_NLS_1_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIGSPKKKRKVSGVWLDGVNIFSVSILLVSAWLEFKGFVRAHICKTCY SGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRR NAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLN KAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASI SNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKN FKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLR GDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTY RGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 234 Cas12f_intraprotein_NLS_2_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVRGSPKKKRKVSGAHICKTCY SGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRR NAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLN KAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASI SNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKN FKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLR GDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTY

	RGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 235 Cas12f_intraprotein_NLS_3_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVRAHICKTCYSGVAGNMFIRK QMYPNDKEGWKVSRSYNIKVNAPGSPKKKRKVSGGLTGTEYAMAIRKAISILRSFEKRRR NAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLN KAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASI SNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKN FKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLR GDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTY RGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 236 Cas12f_intraprotein_NLS_4_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVRAHICKTCYSGVAGNMFIRK QMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKS KKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEGSPKKKRKVSGKWQGISLN KAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASI SNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKN FKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLR GDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTY RGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 237 Cas12f_intraprotein_NLS_5_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVRAHICKTCYSGVAGNMFIRK QMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKS KKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEK RIKKLKEWKHPTLNRPYVELHKNGSPKKKRKVSGNVRIVGYETVELKLGNKMYTIHFASI SNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKN FKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLR GDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTY RGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 238 Cas12f_intraprotein_NLS_6_orange	MPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVC DCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVRAHICKTCYSGVAGNMFIRK QMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKS KKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEK RIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQK KKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVN RLAVGCIISKDGSPKKKRKVSGGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKL RGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERT YRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVD ENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK
SEQ ID NO: 239 Cas12f_intraprotein_and_flanking_NLS_1_grey	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIGSPKKKRKVSGVWLDGVNIFSVSILLV SAWLEFKGFVRAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTE YAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRV KRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGY ETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNF FLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKE NRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKE NTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMI IDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEE KLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 240	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVR

Cas12f_intraprotein_and_flanking_NLS_2_grey	GSPKKKRKVSGAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTE YAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRV KRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGY ETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIKRGKNF FLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKE NRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKE NTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMI IDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEE KLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 241 Cas12f_intraprotein_and_flanking_NLS_3_grey	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVR AHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGSPKKKRKVSGGLTGTE YAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRV KRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGY ETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIKRGKNF FLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKE NRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKE NTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMI IDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEE KLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 242 Cas12f_intraprotein_and_flanking_NLS_4_grey	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVR AHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISIL RSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPE GSPKKKRKVSGKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVG YETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKN FFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAK ENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAK ENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPV MIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWE EKLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 243 Cas12f_intraprotein_and_flanking_NLS_5_grey	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVR AHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISIL RSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPE KWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNGSPKKKRKVSGNVRIVG YETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKN FFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAK ENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAK ENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPV MIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWE EKLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 244 Cas12f_intraprotein_and_flanking_NLS_6_grey	MKRTADGSEFESPKKKRKVMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNIVWLDGVNIFSVSILLVSAWLEFKGFVR AHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISIL RSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPE KWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNK MYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTV KVPKLTKNFKAFGIDRGVNRLAVGCIISKDGSPKKKRKVSGGKLTNKNIFFFHGKEAWAK ENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAK ENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPV MIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWE EKLHAYVCSEPDKSGGSKRTADGSEFEPKKKRKV

Table 19

SEQ ID NO: 245	
RNF2_g8_ PsaCas12f_ targeting	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGC GCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATA GGTCAAGGAATGCCGCTATGAGTTACAACGAACACCTC

Table 20

SEQ ID NO: 246	
SL5_4 + cr21 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAA GTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 247	
SL5_4 + cr21 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTAATAG GTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 248	
SL5_4 + cr21 + SL2_4 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTA ATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 249	
SL5_4 + cr21 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTAAT AGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 250	
SL5_4 + cr22 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAA GTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 251	
SL5_4 + cr22 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTAATAG GTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 252	
SL5_4 + cr22 +	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTA ATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG

SL2_4 + SL1_8	
SL5_4 + cr22 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTAAAAGTAAT AGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 253 SL5_5 + cr21 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAA AAGCTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 254 SL5_5 + cr21 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAGCTAA TAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 255 SL5_5 + cr21 + SL2_4 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAG CTAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 256 SL5_5 + cr21 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAGCT AATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 257 SL5_5 + cr22 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAA AAGCTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 258 SL5_5 + cr22 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAGCTAA TAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 259 SL5_5 + cr22 + SL2_4 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAG CTAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG

SEQ ID NO: 260	
SL5_5 + cr22 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGGAAAAGCT AATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 261	
SL5_7 + cr21 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGA AAAGCATAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 262	
SL5_7 + cr21 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAAGCAT AATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 263	
SL5_7 + cr21 + SL2_4 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAA GCATAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 264	
SL5_7 + cr21 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCCGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAAGC ATAATAGGTCAAGGAATGCGGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 265	
SL5_7 + cr22 + SL2_3 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAACCAAAGCCTATATGGCTGAGGGAG GATGGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGA AAAGCATAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 266	
SL5_7 + cr22 + SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAAGCAT AATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 267	
SL5_7 + cr22 +	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAA GCATAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG

SL2_4 + SL1_8	
SEQ ID NO: 268 SL5_7 + cr22 + SL2_5 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGCTATATGCAGCTGAGGGAGGATGG GCGCTGCGGCATGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGTGAAAAGC ATAATAGGTCAAGGAATGCCGCTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 269 SL2_4 + SL1_1	GCTCCGCTTTAATAAGCGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAG GTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 270 SL2_4 + SL1_3	GCACCACTTTAATAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAG GTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 271 SL2_4 + SL1_5	GCTCCACTGTAATCAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGATGGGC GCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAG GTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG
SEQ ID NO: 272 SL2_4 + SL1_8	GCTCCACTTGCTAATGCAAGTGGTGCCTTCCAAAGCTGTATATCAGCTGAGGGAGGAT GGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTA ATAGGTCAAGGAATGCAACTGGTTGCCCACCCTAGTCATTG

Table 21

SEQ ID NO: 273 cpPsaCas12f_1	MKRTADGSEFESPKKKRKVSGGSISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGN MFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKS KKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKL KEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHL LTLALKRNLETYPSHKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLT NKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYL HNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAE EAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFR WEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFI TFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNSGGSKRTADGSEFEPKKKR KV
SEQ ID NO: 274 cpPsaCas12f_2	MKRTADGSEFESPKKKRKVSGGSNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKV NAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEG HQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGY ETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYP VRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDR LYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYL RERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDE NNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSG GSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNE QFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPSGGSKRTADGSEFEPKKKR KV
SEQ ID NO:	MKRTADGSEFESPKKKRKVSGGSPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEY LELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEW

ID NO: 275 cpPsaCas12f_3	KHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTL ALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNK NIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNI SKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAG VPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEE KLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQR AVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYT IKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNASGGSKRTADGSEFEPKKKR KV
SEQ ID NO: 276 cpPsaCas12f_4	MKRTADGSEFESPKKKRKVSGGSEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELH KNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIK RGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAK ENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPT VIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSR KCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDK GGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIR SSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCY SGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAER RIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPSGGSKRTADGSEFEPKK KRKV
SEQ ID NO: 277 cpPsaCas12f_5	MKRTADGSEFESPKKKRKVSGGSNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSI EYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIIS KDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKY FRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMI KYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYE CPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQAL ENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQK DRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYA MAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHK NWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKSGGSKRTADGSEFEPKKKR KV
SEQ ID NO: 278 cpPsaCas12f_6	MKRTADGSEFESPKKKRKVSGGSDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRG DKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRS KKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCL KCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGM PSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGK KEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEG WKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKG KTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYV ELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPS IIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKSGGSKRTADGSEFEPKKKR KV
SEQ ID NO: 279 cpPsaCas12f_7	MKRTADGSEFESPKKKRKVSGGSKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKE AWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAK ENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPR NTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCS EPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRI VDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICK TCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRN AERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSK VKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQ KKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPSGGSKRTADGSEFEPKKKRK V
SEQ	MKRTADGSEFESPKKKRKVSGGSKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYT KDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRK

ID NO: 280 cpPsaCas12f_8	AISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 281 cpPsaCas12f_9	MKRTADGSEFESPKKKRKVSGGSKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 282 cpPsaCas12f_10	MKRTADGSEFESPKKKRKVSGGSAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 283 cpPsaCas12f_11	MKRTADGSEFESPKKKRKVSGGSRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKSGGSKRTADGSEFEPKKKRKV
SEQ ID NO: 284 SEQ ID NO: 285	MKRTADGSEFESPKKKRKVSGGSNTSRKCSKCGYVDENNRKQASFKCLKCGYSLNADLNAAVNIAKAFYECPTFRWEEKLHAYVCSEPDKGGSGGSGGSGGSGGSGGSGGMPSETYITKTLSLKLIPSDEEKQALENYFITFQRAVNFAIDRIVDIRSSFRYLNKNEQFPAVCDCCGKKEKIMYVNISNKTFKFKPSRNQKDRYTKDIYTIKPNAHICKTCYSGVAGNMFIRKQMYPNDKEGWKVSRSYNIKVNAPGLTGTEYAMAIRKAISILRSFEKRRRNAERRIIEYEKSKKEYLELIDDVEKGKTNKIVVLEKEGHQRVKRYKHKNWPEKWQGISLNKAKSKVKDIEKRIKKLKEWKHPTLNRPYVELHKNNVRIVGYETVELKLGNKMYTIHFASISNLRKPFRKQKKKSIEYLKHLLTLALKRNLETYPSIIKRGKNFFLQYPVRVTVKVPKLTKNFKAFGIDRGVNRLAVGCIISKDGKLTNKNIFFFHGKEAWAKENRYKKIRDRLYAMAKKLRGDKTKKIRLYHEIRKKFRHKVKYFRRNYLHNISKQIVEIAKENTPTVIVLEDLRYLRERTYRGKGRSKKAKKTNYKLNTFTYRMLIDMIKYKAEEAGVPVMIIDPRSGGSKRTADGSEFEPKKKRKV

cpPsaCas12f_12	

EXAMPLES

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

Example 1

Computational Discovery of Miniature CRISPR Nucleases

The computational discovery of miniature CRISPR nucleases was performed (**FIGS. 1A-1D**).

Novel miniature CRISPR nucleases from metagenomic samples were identified by computer discovery (**FIG. 1A**). Initial panning for small CRISPR nucleases yielded orthologs, including 30 novel Cas12f orthologs, 20 novel Cas12j orthologs, and 45 novel Cas12m orthologs (**FIG. 1B**). These orthologs comprise a C-terminal RuvC domain indicative of Cas12 systems and CRISPR arrays of 2 or more spacers with direct repeats that fold with an appropriate secondary structure (**FIG. 1E**). The Cas12f and Cas 12m systems have readily identifiable putative tracrRNAs found by a homology search of the DR against the surrounding locus and a secondary structure modeling/prediction to identify the tracrRNA sequence with the best folding energy to the crRNA (**FIG. 1F**). The Cas12js systems do not have any identifiable tracrRNA and the Cas12m systems do have identifiable tracrRNAs. The new subclasses of Cas12s require or do not require tracrRNA.

FIG. 1C shows the size distribution of Cas12a and **FIG. 1D** shows the size distribution of CasM ortholog.

Example 2 – PsaCas12f sgRNA Constructs

PsaCas12f sgRNA constructs were tested in human mammalian cells (**FIG. 4**).

A panel of 24 sgRNA designs against a pUC19 reported plasmid with PsaCas12f was tested. The sgRNA designs are disclosed in **Table 1** and achieved up to about 0.5% editing. The experiments were performed with plasmid expression in HEK293FT for 48-72 hours.

Example 3 – PsaCas12f sgRNA Designs Based On sgRNA Secondary Structure

SgRNA's secondary structure is critical to enabling the specific and effective recognition between Cas9 and the target sequence. To further improve the cleavage efficiency of the PsaCas12f-sgRNA complex, sgRNA variants were designed to comprise genetic mutations which would impact the sgRNA's secondary structure as well as interactions with the sgRNA-protein complex.

The predicted sgRNA secondary structure was obtained through use of *in silico* structure determination. Stem loop 1-3 (SL1-3) were predicted via http://rna.tbi.univie.ac.at/. Stem loop 4 (SL4, interacts with crRNA) and stem loop 5 (SL5) were informed by Takeda *et al.*, Mol Cell, 81(3):558-570 (2021). **FIG. 10A** illustrates the resulting sgRNA secondary structure with SL1-SL3 marked by blue, red, and green boxes, respectively.

Using this predicted sgRNA secondary structure, genetic mutations were engineered into SL1, SL2, SL3, SL4, or SL5. **FIG. 10B** lists and annotates all the sgRNA variants designed (see also sequence listing in **Table 14**). Red denotes nucleobase changes that were introduced, orange denotes nucleobases that form stems, and violet denotes loops that were added to allow recruitment of MS2 coat/proteins.

Subsequently, using an *in vitro* luciferase reporter assay, the sgRNA variants were tested to assess whether secondary structure modifications of SL1-SL5 could impact cleavage efficiency. Briefly, HEK293T cells were seeded and transfected with 25 ng of a luciferase reporter, 100ng of different CRISPR guides annotated above, and 300ng of PsaCas12f-expressing plasmid. Seventy-two hours after transfection, media was harvested from cells and analyzed for luciferase expression.

The corresponding bar graph in **FIG. 10C** shows the results of the reporter assay. Notably, certain genetic modifications to SL1, SL2, SL3, SL4, or SL5 increased the cleavage efficiency over controls (control sgRNA constructs previously optimized using a different strategy, labeled "5pr_trunc4-7" and "best guide v2").

Example 4 – PsaCas12f sgRNA Combination Mutant Stem-loop Constructs

The sgRNA variants in Example 3 each targeted a different stem-loop regions (SL1, SL2, SL3, SL4, or SL5). It was hypothesized that each stem-loop region may impact a variety of

functions (e.g., hairpin stability, transcription efficiency, protein interaction) and that combining the single stem-loop mutant variants designed in Example 3 would further improve cleavage efficiency. Accordingly, sgRNA variants which contained a combination of modifications from the sgRNA variants with single modifications at a particular stem-loop region was designed (also called, "combination constructs"). The aim of the sgRNA combination stem-loop variants was to increase folding and Cas12f interaction (e.g., GC content increase, sgRNA truncation/mismatch correction in stem loops, removal of premature termination signals).

Combination constructs are presented in **Table 16**. **FIG. 11A** shows the resulting performance of the combination constructs relative to controls in the *in vitro* luciferase reporter assay. Surprisingly, certain combinations, such as, the construct labeled, "SL1_modification_1 + increase_interaction_w_crRNA_22," resulted in enhanced cleavage efficiency (about 0.035% RLU cleavage) relative to the single modification construct labeled, "SL1_modification_1," (about 0.025% RLU cleavage), compare FIG 10C to FIG 11A).

Subsequently, combination constructs, either double variants with modifications of stem loop 1 and 2 (labeled, 2X combinations in **FIG. 11B**) or quadruple variants with modifications of stem loop 1, 2, 3, and 5 (labeled 4x combinations in **FIG. 11B**) were interrogated for cleavage efficiency at the EMX1 (empty spiracles-like protein 1) locus.

Briefly to measure cleavage efficiency at the EMX1 locus, 100ng of different CRISPR guides annotated above in Table 16 and 300ng of PsaCas12f-expressing plasmid were transfected into HEK293FT cells. Seventy-two hours after transfection, cells were harvested for their genomic DNA and primers amplifying EMX1 genomic locus were used to amplify the genomic region in the locus. Subsequently, next generation sequencing (NGS) was performed on these amplified gDNA and the insertion/deletion profile caused by Cas12f with the different guides was analyzed with CRISPResso.

FIG. 11B shows the result of the editing efficiencies at the EMX1 locus for the combination constructs noted above. Notably, for the 4x combination constructs tested, the construct labeled, "SL5_4 + cr21 + SL2_4 + SL1_8," had greater editing efficiency at the EMX1 locus than the control constructs with either a single stem-loop modification or no stem-loop modification. It is not entirely obvious why certain combination constructs work better than other combination. For example, compare the EMX1 editing efficiency of the 2x combinations "SL2_4+SL1_1" with "SL2_4+SL1_3." One hypothesis is that certain base-pair combinations do

not provide optimal sgRNA folding/sgRNA-protein interaction and these occurrences are difficult to predict *in silico*.

The best sgRNA combination mutant stem-loop constructs named (1) scaffold "version 2", (2) "version 3.1, SL1_modification_8 + increase_interaction_w_crRNA_21, or SEQ ID NO: 203", and (3) "v. 3.2, SEQ ID NO: 198") from **FIG. 11A and 11B** were subsequently tested with 30 different PsaCas12f mutants relative to controls in the *in vitro* luciferase reporter assay the order to test the robustness of the sgRNA scaffold as shown in **FIG. 11C**. Notably, scaffold "v. 3.2" which includes the modification of mutant combination "SL1_8" and "interaction_w_cRNA_22" performed well across the panel of PsaCas12f mutants tested demonstrating the robustness of the "v.3.2" as a sgRNA scaffold.

Example 5 – Spacer Optimization for sgRNA Scaffold Version 3.2 for PsaCas12f

The sgRNA spacer sequence can impact target specificity and the degree of off-target activity. **FIG. 12A** is a schematic of the sgRNA scaffold version 3.2 which highlights the position of the spacer sequence at the 3' end. This experiment was designed to test the cleavage efficiency of the sgRNA v. 3.2 scaffold from Example 4 by varying the nucleotide length of the sgRNA spacer sequence.

To test spacer length, the version 3.2 sgRNA scaffold was tested in the *in vitro* luciferase reporter assay at spacer sequence lengths of 2, 3, 18, 19, 20, 21, 22, 23, 24, and 25 base pairs relative to controls. **FIG. 12B** shows that using v3.2 sgRNA scaffold for PsaCas12f, the highest cleavage efficiency was achieved using a spacer sequence of 21bp for this specific target. While 22bp, 20bp, 19bp and even 18bp still worked, 21bp showed the highest gene editing. As such, for the PsaCas12f-version3.2 sgRNA 20bp or 21 bp is enough to allow sufficient base-pairing before cleavage.

Example 6 – PsaCas12f with the sgRNA Scaffold Version 3.2 is more efficacious than UnCas12f (Cas14a1)

PsaCas12f with the sgRNA scaffold version 3.2 described in Example 4 was then compared to a different Cas12f protein which is similarly small and has good on-target efficiency called, Un1Cas12f1 (also called Cas14a1) at either the HBB (hemoglobin subunit beta) or the

RNF2 (ring finger protein 2) genomic locus. Un1Cas12f1 is a protein identified from an uncultured archaeon (Un1).

Briefly, 100ng of different CRISPR guides based on scaffold version 2 with different spacer lengths according to their descriptions (e.g., stagger_24 denotes a spacer length of 24 nt) annotated in **Table 17** and 300ng of PsaCas12f-expressing plasmid are transfected into HEK293FT cells. Two spacer sequences targeting either RNF2 or HBB genomic locus were designed with sgRNA v3.2 scaffold. Seventy-two hours after transfection, cells were harvested for their genomic DNA and primers amplifying the corresponding genomic locus were used to amplify the gDNA in the locus. Subsequently, next generation sequencing (NGS) was performed on these amplified gDNA, and insertion/deletion profile caused by Cas12f with different guide was analyzed with CRISPResso.

FIG. 13 shows that PsaCas12f with the sgRNA scaffold version 3.2 outperformed Un1Cas12f1 with the nbt scaffold in terms of indel activity (insertion/deletion formation) at both sites tested in the Hbb locus (g1 and g2) as well as one a site in the RNF locus (g4). As such, PsaCas12f with the sgRNA scaffold version 3.2 allows efficient indel formation and may be a useful tool for broad genome engineering applications.

Example 7 – PsaCas12f NLS Constructs

PsaCas12f Nuclear Localization Signals (NLS) constructs were tested in HEK293FT human mammalian cells (**FIG. 5A-5D**).

A panel of 15 NLS designs fused to PsaCas12f against a pUC19 reported plasmid using the top two guide sequences from Example 2 was tested. The NLS designs are disclosed in **Table 1** and achieve up to about 0.1% editing (**FIG. 5A**). The experiments were performed with plasmid expression in HEK293FT for 48-72 hours. The sequencing traces show bona-fide editing as illustrated in **FIGS. 5B-5E**. Editing with PsaCas12f (NLS14) with sgRNA (**FIG. 5B**) or non-targeting guide (**FIG. 5C**) shows clear deletions (purple) and insertions (red). Editing with PsaCas12f (no NLS) with sgRNA (**FIG. 5D**) or non-targeting target guide (**FIG. 5E**) also shows clear deletion (purple) and insertions (red).

Intra NLS signals could allow better design of proteins delivered via viral-like particles, Banskota *et al.*, Cell, 185(2):250-265 (2022), or enable inducible NLS signals following conformational change, Saleh *et al.*, Exp Cell Res, 260(1):105-115 (2000). As such, an intra-

protein NLS sequence derived from SV40 (simian virus 40) was fused at random positions into PsaCas12f as shown in **FIG. 14** and annotated in Table 18. These constructs were tested for indel activity at the EMX genomic locus.

Briefly, seventy-two hours after transfection, cells were harvested for their genomic DNA and primers amplifying the corresponding EMX genomic locus was used to amplify the gDNA in the locus. Subsequently, next generation sequencing (NGS) is performed on these amplified gDNA, and insertion/deletion profile was analyzed with CRISPResso.

Intra NLS signals, labeled "NLS_2", "NLS_3", "NLS-5", and "NLS_6," had higher indel activity at the EMX locus than wild-type PsaCas12f which was flanked by two NLS sequences on the N- and C- terminus (labeled, "pDF0106")as shown in **FIG. 14**. Therefore, intra NLS signals could provide alternative localization to flanking NLS signals while still maintaining optimal gene editing activity. Intra NLS signals could be advantageous for example, when the N- or C- terminal NLS fusions interfere with protein function.

Example 8 – CRISPR editing with PsaCas12f and guide RNA delivered by adeno-associated virus (AAV)

Adeno associated virus (AAV) is a US Food and Drug administration approved safe vehicle for gene therapies and for this reason AAV-loadable CRISPR tools are advantageous. AAV has a limited payload size of <4.7 kb which hampers clinical applications of most CRISPR tools. Therefore, this Example validates AAV delivery of PsaCas12f-sgRNA.

Briefly, PsaCas12f with the best NLS configuration (flanking SV40NLS) was cloned into AAV ITR along with a guide targeting RUNX1 (runt-related transcription factor 1) genomic locus. Subsequently, the plasmid was transfected into HEK293FT cells with AAV helper plasmid to make AAV particles. AAV particles in the media from the producer cell line was collected and subsequently added to HEK293FT cells. Four days after transduction, the indel profile at the RUNX1 locus was analyzed with NGS.

As shown in **FIG. 15,** the AAV-loaded with PsaCas12f plus guide had indel frequencies of about 10-14% at the RUNX1 genomic locus increasing commensurately with the amount transduced into HEK293 cells (1, 5, or 25 µl). This experiment demonstrates that PsaCas12f can

be effectively expressed from AAV particles while maintaining the ability to induce cleavage at a genomic target.

Example 9 – PsaCas12f with Guide CrRNA/TracrRNA

PsaCas12f with CrRNA/tracrRNA guide was screened at different free-energy local minima (FIG. 6).

Results from PsaCas12f show that many crRNA/tracrRNA designs must be screened at a variety of free-energy local minima to find optimal combinations for activity in bacterial or mammalian protein lysate. A 20-nt DR and 90-nt tracrRNA were found to provide optimal activity for dsDNA cleavage and that they can be combined for a sgRNA. These designs showed that the computational and experimental RNA screening can yield optimal designs and that sgRNA has a significant effect on activity.

Example 10 – Genome Editing by Cas12f Family Members

Cas12f family members were tested for genome editing (**FIG. 7**). These tests from Cas12f family members for indel generation at EMX1 result in editing efficiencies above background.

Example 11 – Screening of a Panel of 12 Cas12f Orthologs

A panel of 12 novel Cas12f orthologs ranging in size between 400-800 amino acids was screened. In order to maintain the correct small RNA species from these orthologs, non-coding regions from the surrounding loci along with the Cas12f genes were cloned (**FIG. 8A**). Purification of lysate from these samples enabled testing of *in vitro* cleavage on degenerate PAM libraries, where cleaved fragments can be enriched to determine the PAM. Of all 12 proteins, one of the orthologs, the Cas12f from *Pseudomonas aeruginosa (g-proteobacteria)* (PsaCas12f), a 586-residue protein, had substantial cleavage activity determined by this high-throughput PAM screen. PAM characterization had determined the motif of PsaCas12f to be TTR (**FIG. 8B**). Additionally, small RNA sequencing of these purified proteins can determine the mature isoforms of the processed crRNA and tracrRNA (**FIG. 8C**), yielding a natural DR length of 31 nt

and tracrRNA length of 97 nt. Lastly, the PAM of PsaCas12f on fixed sequence targets was validated to demonstrate detectable *in vitro* cleavage by gel readouts (**FIG. 8D**). The characterization of PsaCas12f and the corresponding RNA species, as well as other effectors selected from the high-throughput screening can be optimized for activity by guide RNA engineering.

Example 12 – PsaCas12f Circular Permutation

While Cas nucleases did not evolve to function as a modular DNA-binding scaffold optimizing Cas nucleases by fusion to functional protein domains using linkers may enable controlled nuclease activity and broaden the use of Cas nuclease as a genetic tool. Oakes *et al.* Cell, 176(2): 254-267 (2019). One way to change the CRISPR architecture to enable fusion to other protein domains is by protein circular permutation (CP). *Id.* CP is the topological rearrangement of a protein's primary sequence, connecting its N- and C-terminus with a peptide linker, while concurrently splitting its sequence at a different position to create new, adjacent N and C termini. Yu and Lutz, Trends Biotechnol, 28: 18-25 (2011).

To test whether PsaCas12f proteins as described above could undergo circular permutation without impacting functional activity, the PsaCas12f sequence was split at different positions to create new adjacent N- and C- termini using a $(GGS)_6$ peptide linker as shown in **Table 15** (see also, bottom schematic in **FIG. 16A**).

Circular permutation constructs listed in **Table 21** were then tested for editing efficiency either using the *in vitro* luciferase reporter assay described above or by testing indel formation at the RUNX1 genomic locus as shown in **FIG. 16A** and **FIG. 16B**, respectively.

Briefly, for the *in vitro* luciferase reporter assay 25ng of Gluc reporter, 100ng of the CRISPR guide, and 300ng of either regular PsaCas12f-expressing plasmid (control, labeled pDF0106) or different circular permutation of the protein encoding plasmids were transfected into HEK293FT cells. Seventy-two hours after transfection, media is harvested from cells and analyzed for luciferase expression. For assessment of indel formation at the RUNX1 genomic locus, the same panel of circular permutations of PsaCas12f proteins were tested with guides targeting genomic RUNX1 locus. Cell transfection conditions were the same as for the in vitro luciferase, PCR was used to amplify the genomic locus at RUNX1 and indel efficiency estimated by CRISPResso.

Notably, some circular permutations of PsaCas12f are functional and allow for different positioned N- and C-termini. Interestingly, the editing efficiency changes depending on the guide that is used (compare editing efficiencies from **FIG. 16A** and **FIG. 16B**).

Example 13 – PsaCas12f Sequence Optimization via Machine Learning

The wild-type PsaCas12f sequences was sent to a machine learning model (Facebook Evolutionary Scale Modeling (ESM), https://github.com/facebookresearch/esm) for prediction of point mutations on the protein that could result in higher editing efficiencies. Namely, the original WT sequence was used as input in the ESM model. The output of the ESM model was a single vector (1x1280), and this vector was subsequently used as an input in a linear regression model to predict the output which is the indel formation rate. New mutations made on the protein were sent through the model in a similar fashion to predict the indel and subsequently tested *in vitro*.

Forty-eight different point mutations were compared with one unifying best guide, v3.2 scaffold described above and a spacer targeting RNF2 (tatgagttacaacgaacacctc) (*see* Table 18) targeting the genomic RNF2 locus. Seventy-two hours after transfection of the panel of PsaCas12f varaints containing a single point mutation (plus the sgRNA), genomic locus at RNF2 was PCR amplified and subjected to NGS. Indel profile is quantified by CRISPResso for all the mutants.

Of the panel of point mutations tested, the point mutation at position 333 of PsaCas12f to Valine from Lysine dramatically increased the cleavage efficacy of PsaCas12f as shown in **FIG. 17**.

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.

CLAIMS

What is claimed is:

1. A composition comprising:

(a) a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19; and

(b) a guide RNA (gRNA)

wherein a target comprises a DNA target.

2. The composition of claim 1, wherein the DNA target is a single stranded DNA.

3. The composition of claim 1, wherein the DNA target is a double stranded DNA.

4. The composition of claim 1, wherein the target specific nuclease has a length less than about 1000 amino acids.

5. The composition of claim 4, wherein the target specific nuclease has a length less than about 900 amino acids.

6. The composition of claim 5, wherein the target specific nuclease has a length less than about 800 amino acids.

7. The composition of claim 1, wherein the amino acid sequence is SEQ ID NO: 1.

8. The composition of claim 1 wherein the target specific nuclease comprises an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO: 1.

9. The composition of claim 1, wherein the target specific nuclease comprises an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO: 1.

10. The composition of claim 1, wherein the target specific nuclease comprises an amino acid sequence 98% identical to the amino acid sequence of SEQ ID NO: 1.

11. The composition of claim 1, wherein the target specific nuclease comprises an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO: 1.

12. The composition of claim 1, wherein the nuclease is the amino acid sequence of SEQ ID NO: 1.

13. The composition of any one of the previous claims, wherein the target specific nuclease is selected from the group consisting of Cas12f, Cas12m, and any variants thereof; and optionally wherein the target specific nuclease is PsaCas12f.

14. The composition of any one of the previous claims, wherein the gRNA is a single guide RNA (sgRNA) or a dual guide (dgRNA).

15. The composition of any one of the previous claims, wherein the gRNA is a sgRNA comprising a nucleic acid sequence 70% identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43, 61-79, 145-198.

16. The composition of anyone one of the previous claims, wherein the gRNA has a spacer region with a sequence comprising a length of about 17 to about 53 nucleotides (nt), optionally wherein the sequence comprises a length of about 29 to about 53 nt, optionally wherein the sequence comprises a length of about 40 to about 50 nt; or optionally wherein the sequence comprises a length of about 21 to 22 nt.

17. The composition of anyone one of the previous claims, wherein the gRNA has a direct repeat region with a sequence having a length of from about 20 to about 29 nt.

18. The composition of anyone of the previous claims, wherein the gRNA has a tracrRNA region with a sequence having a length of from about 27 to about 35 nt.

19. The composition of anyone one of the previous claims, wherein the target is in a cell.

20. The composition of claim 19, wherein the cell is a prokaryotic cell.

21. The composition of claim 19, wherein the cell is a eukaryotic cell.

22. The composition of claim 21, wherein the eukaryotic cell is a mammalian cell.

23. The composition of claim 22, wherein the mammalian cell is a human cell.

24. The composition of anyone one of the previous claims, wherein the amino acid sequence specifically binds to a protospacer-adjacent motif (PAM).

25. The composition of claim 24, wherein the PAM is selected from the group consisting of NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

26. A nucleic acid molecule encoding the target specific nuclease of any of the preceding claims.

27. A nucleic acid molecule encoding the gRNA of any of the preceding claims.

28. One or more vectors comprising the nucleic acid molecule of claims 26-27.

29. A cell comprising the composition of claims 1-25, the nucleic acid molecule of claims 26-27 or the one or more vectors of claim 28.

30. The cell of claim 29, wherein the cell is a prokaryotic cell.

31. The cell of claim 29, wherein the cell is a eukaryotic cell.

32. The cell of claim 31, wherein the eukaryotic cell is a mammalian cell.

33. The cell of claim 32, wherein the mammalian cell is a human cell.

34. A method of inserting or deleting one or more base pairs in a DNA, the method comprising:

(a) cleaving the DNA at a target site with a target specific nuclease, wherein the cleavage results in overhangs on both DNA ends;

(b) inserting a nucleotide complementary to the overhanging nucleotide on both of the DNA ends, or removing the overhanging nucleotide on both of the DNA ends; and

(c) ligating the DNA ends together, thereby inserting or deleting one or more base pairs in the DNA,

wherein the nuclease comprising an amino acid sequence 70% identical to an amino acid sequence selected from the group consisting of SEQ ID NOs: 1-19, and

wherein the target specificity of the target specific nuclease is provided by a guide RNA (gRNA).

35. The method of claim 34, wherein the target specific nuclease has a length less than about 1000 amino acids.

36. The method of claim 35, wherein the target specific nuclease has a length less than about 900 amino acids.

37. The method of claim 36, wherein the target specific nuclease has a length less than about 800 amino acids.

38. The method of claim 34, wherein the amino acid sequence is SEQ ID NO: 1.

39. The method of claim 38, wherein the target specific nuclease comprises an amino acid sequence 90% identical to the amino acid sequence of SEQ ID NO: 1.

40. The method of claim 38, wherein the target specific nuclease comprises an amino acid sequence 95% identical to the amino acid sequence of SEQ ID NO: 1.

41. The method of claim 38, wherein the target specific nuclease comprises an amino acid sequence 98% identical to the amino acid sequence of SEQ ID NO: 1.

42. The method of claim 38, wherein the target specific nuclease comprises an amino acid sequence 99% identical to the amino acid sequence of SEQ ID NO: 1.

43. The method of claim 34, wherein the nuclease is the amino acid sequence of SEQ ID NO: 1.

44. The method of any one of claims 34-43 wherein the target specific nuclease is selected from the group consisting of Cas12f, Cas12m, and any variants thereof; and optionally wherein the target specific nuclease is PsaCas12f.

45. The composition of any one of claims 34-44, wherein the gRNA is a single guide RNA (sgRNA) or a dual guide RNA (dgRNA).

46. The method of claim 45, wherein the gRNA is a sgRNA comprising a nucleic acid sequence 70% identical to a nucleic acid sequence selected from the group consisting of SEQ ID NOs: 20-43, 61-79, and 145-198.

47. The method of any one of claims 34-46, wherein the gRNA has a spacer region with a sequence having a length of from about 17 to about 30 nucleotides (nt), about 22 nt; or wherein the gRNA has a spacer region with a sequence having a length of from about 20 to about 53 nt, from about 29 to about 53 nt or from about 40 to about 50 nt.

48. The method of any one of claims 34-47, wherein the DNA target is in a cell.

49. The method of claim 48, wherein the cell is a prokaryotic cell.

50. The method of claim 49, wherein the cell is a eukaryotic cell.

51. The method of claim 50, wherein the eukaryotic cell is a mammalian cell.

52. The method of claim 51, wherein the mammalian cell is a human cell.

53. The method of any one of claims 34-52, wherein the amino acid sequence specifically binds to a protospacer-adjacent motif (PAM).

54. The method of claim 53, wherein the PAM is selected from the group consisting of NNNNGATT, NNNNGNNN, NNG, NG, NGAN, NGNG, NGAG, NGCG, NAAG, NGN, NRN, NNGRRN, NNNRRT, TTTN, TTTV, TYCV, TATV, TYCV, TATV, TTN, KYTV, TYCV, TATV, TBN, any variants thereof, and any combinations thereof.

55. A method of detecting a DNA target, the method comprising:

coupling the DNA target with a reporter to form a DNA-reporter complex;

mixing the DNA-reporter complex with a target specific nuclease and a guide RNA (gRNA);

cleaving the DNA-reporter complex; and

measuring a signal from the reporter, thereby detecting the DNA target.

56. The method of claim 55, wherein the target specific nuclease is selected from the group consisting of Cas12f, Cas12m, and any variants thereof; and optionally wherein the target specific nuclease is PsaCas12f.

57. The method of claim 55 wherein the target specific nuclease is complexed with a crRNA.

58. The method of claim 55, wherein the reporter is a fluorescent reporter.

59. A method for activating or inhibiting the expression of a gene, the method comprising mixing the composition of claim 1 with one or more transcription factors, wherein the target specific nuclease lacks endonuclease ability, wherein the target DNA comprises the gene, thereby activating the gene.

60. A method for nucleic acid base editing, the method comprising mixing the composition of claim 1, wherein the target specific nuclease is a nickase or a nuclease coupled to a deaminase, thereby editing the nucleic acid base from the target DNA.

61. A method for activating or inhibiting the expression of a gene, the method comprising mixing the composition of claim 1 with one or more epigenetic modifiers, wherein the target specific nuclease lacks endonuclease activity, wherein the target DNA comprises the gene, and modifying the target DNA or one or more histones associated to the target DNA, thereby activating or inhibiting the gene.

62. The method of claim 68, wherein the epigenetic modifier comprises KRAB, DNMT3a, DNMT1, DNMT3b, DNMT3L, TET1, p300, any variants thereof, or any combinations thereof.

63. The composition of any one of claims 1-25, wherein the gRNA comprises a nucleic acid sequence 70% identical to a nucleic acid sequence from the group consisting of SEQ ID NO: 246-272.

64. The composition of any one of claims 1-25, wherein the target specific nuclease is fused to a nuclear localization signal (NLS).

65. The composition of claim 64, wherein the NLS signal is at the 5' or 3' termini of the target specific nuclease nucleic acid sequence.

66. The composition of claim 64, wherein the NLS signal is in an intra-protein region.

67. The composition of any one of claims 63-65, wherein the NLS is derived from SV40.

68. The composition of any one of claims 63-66, wherein the target specific nuclease comprises a nucleic acid sequence 70% identical to a nucleic acid sequence from the group consisting of SEQ ID NO: 233-244.

69. The composition of any one of claims 1-25 or 63-68, wherein the target specific nuclease and the gRNA are delivered to the cell containing the DNA target in one or more adeno-associated viral (AAV) vectors.

70. The composition of any one of claims 1-25 or 63-69, wherein the target specific nuclease has been circular permutated.

71. The composition of claim 70, wherein the target specific nuclease is PasCas12f.

72. The composition of claim 70 or 71, wherein the target specific nuclease comprises a nucleic acid sequence 70% identical to a nucleic acid sequence from the group consisting of SEQ ID NO: 273-285.

73. The composition of any one of claims 1-25 or 63-72, wherein the target specific nuclease has a point mutation at amino acid position 333 encoding a valine.

74. The composition of claim 73, wherein the point mutation at amino acid position 333 is mutated to a lysine.

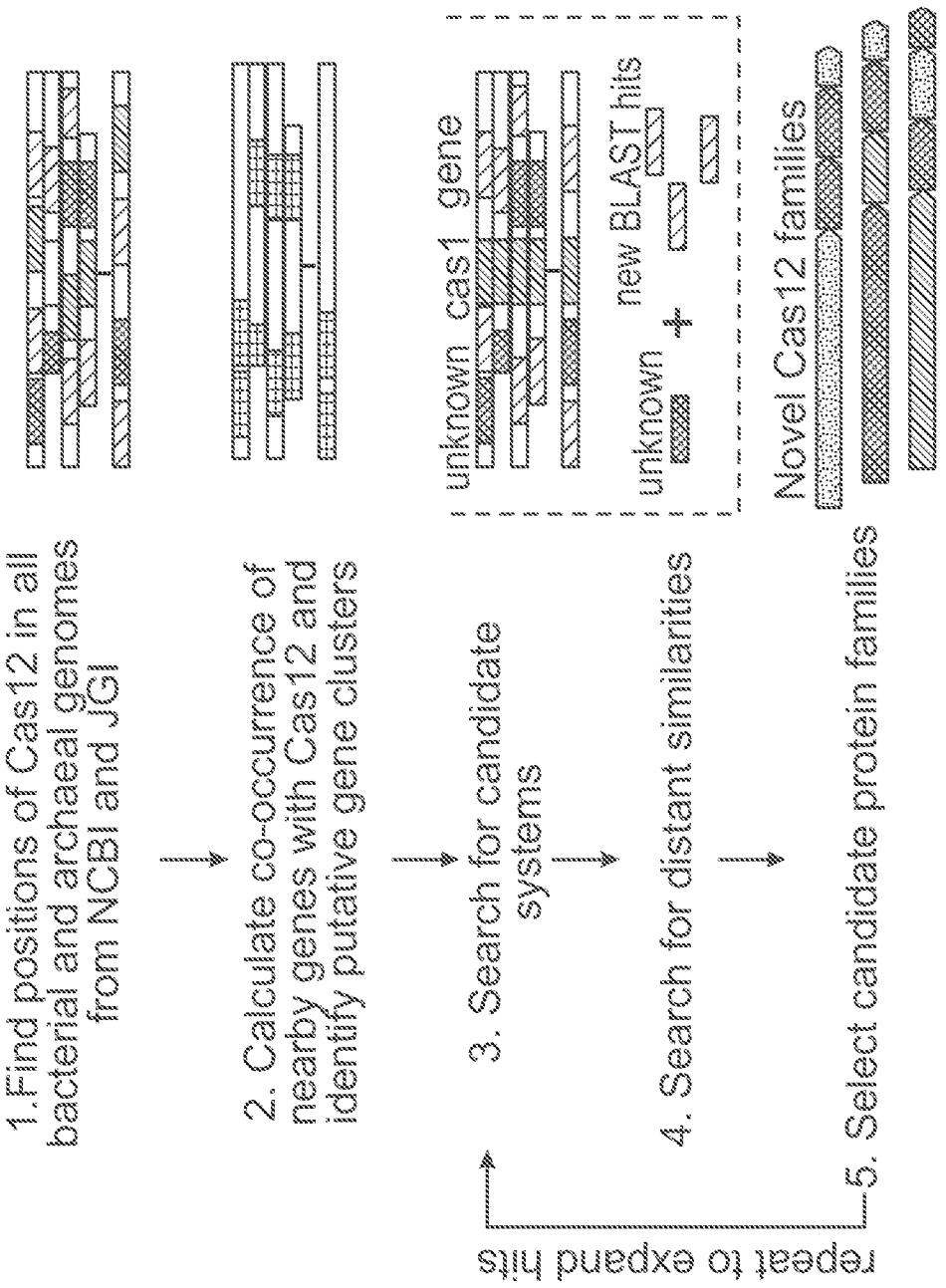
1.Find positions of Cas12 in all bacterial and archaeal genomes from NCBI and JGI
2. Calculate co-occurrence of nearby genes with Cas12 and identify putative gene clusters
3. Search for candidate systems
4. Search for distant similarities
5. Select candidate protein families
repeat to expand hits
repeat for each candidate
unknown cas1 gene
unknown
+
new BLAST hits
Novel Cas12 families

FIG. 1A

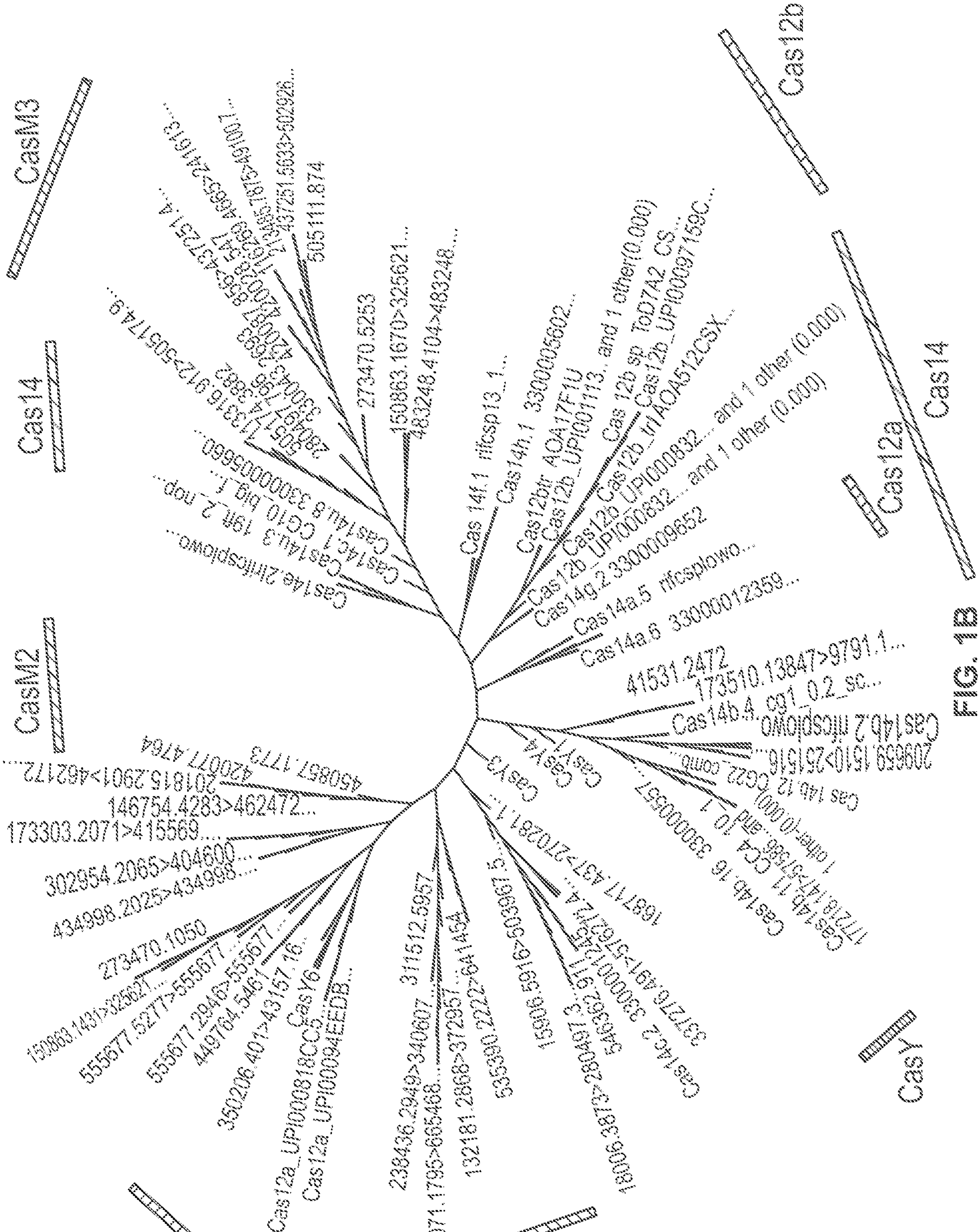
Cas12b
CasM3
Cas14
Cas12a
Cas14
CasM2
CasY
CasM1
Cas12a
505111.874
273470.5253
150863.1670>325621...
483248.4104>483248...
Cas 14l.1 rifcsp13_1...
Cas14h.1 3300005602...
Cas12bir AOA17F1U
Cas12b_UPI001113... and 1 other(0.000)
Cas 12b sp ToD7A2 CS...
Cas12b_UPI00097159C...
Cas12b_tr|AOA512CSX...
Cas12b_UPI00832... and 1 other (0.000)
Cas12b_UPI000832... and 1 other (0.000)
Cas14g.2 3300009652
Cas14a.5 rifcsplowo...
Cas14a.6 33000012359...
41531.2472
173510.13847>9791.1...
Cas14b.4_cg1_0.2_sc...
Cas14b.2 rifcsplowo
209659.1510>251516...
Cas14b.12
Cas14b.16 3300000557
Cas14e 2|rifcsplowo...
Cas14u.8 3300009660...
Cas14u.3 19ft_2_nop...
CG10_big_f...
CasY3
CasY4
CasY1
450857.1773
420077.4764
201815.2901>462172...
146754.4283>462472...
173303.2071>415569...
302954.2065>404600...
434998.2025>434998...
273470.1050
555677.5277>555677...
555677.2946>555677...
449764.5461
350206.401>43157.16...
CasY6
Cas12a_UPI000818CC5...
Cas12a_UPI00094EEDB...
311512.5957
238436.2949>340607...
450071.1795>665468...
132181.2868>372957...
535390.2222>641454
15906.5916>503967.5...
168717.437>270281.1...
337276.491>...
Cas14c.2
18006.3873>280497.3...

FIG. 1B

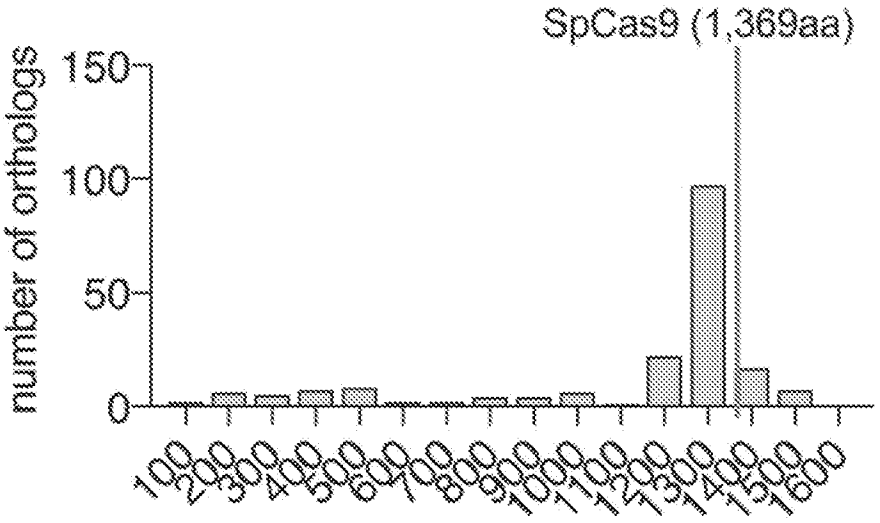
SpCas9 (1,369aa)
number of orthologs
150
100
50
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600

FIG. 1C

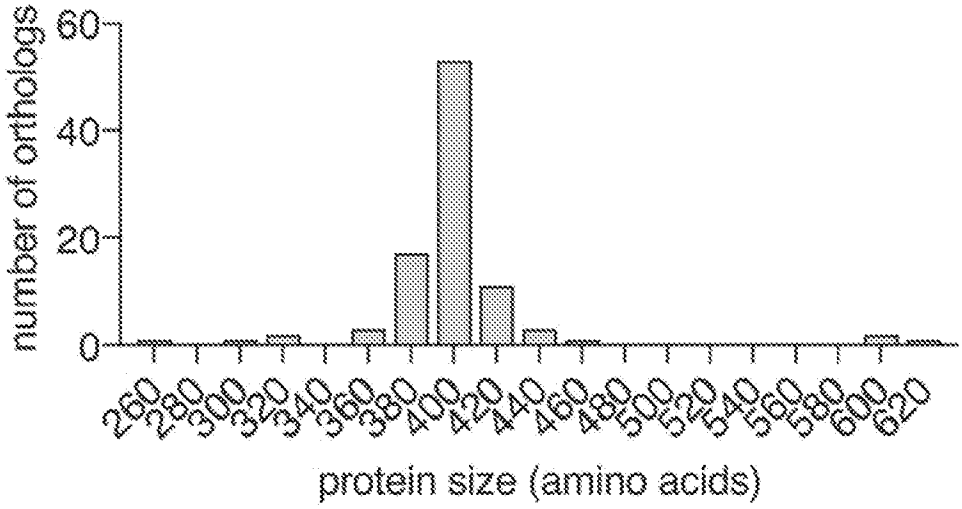
number of orthologs
60
40
20
0
260
280
300
320
340
360
380
400
420
440
460
480
500
520
540
560
580
600
620
protein size (amino acids)

FIG. 1D

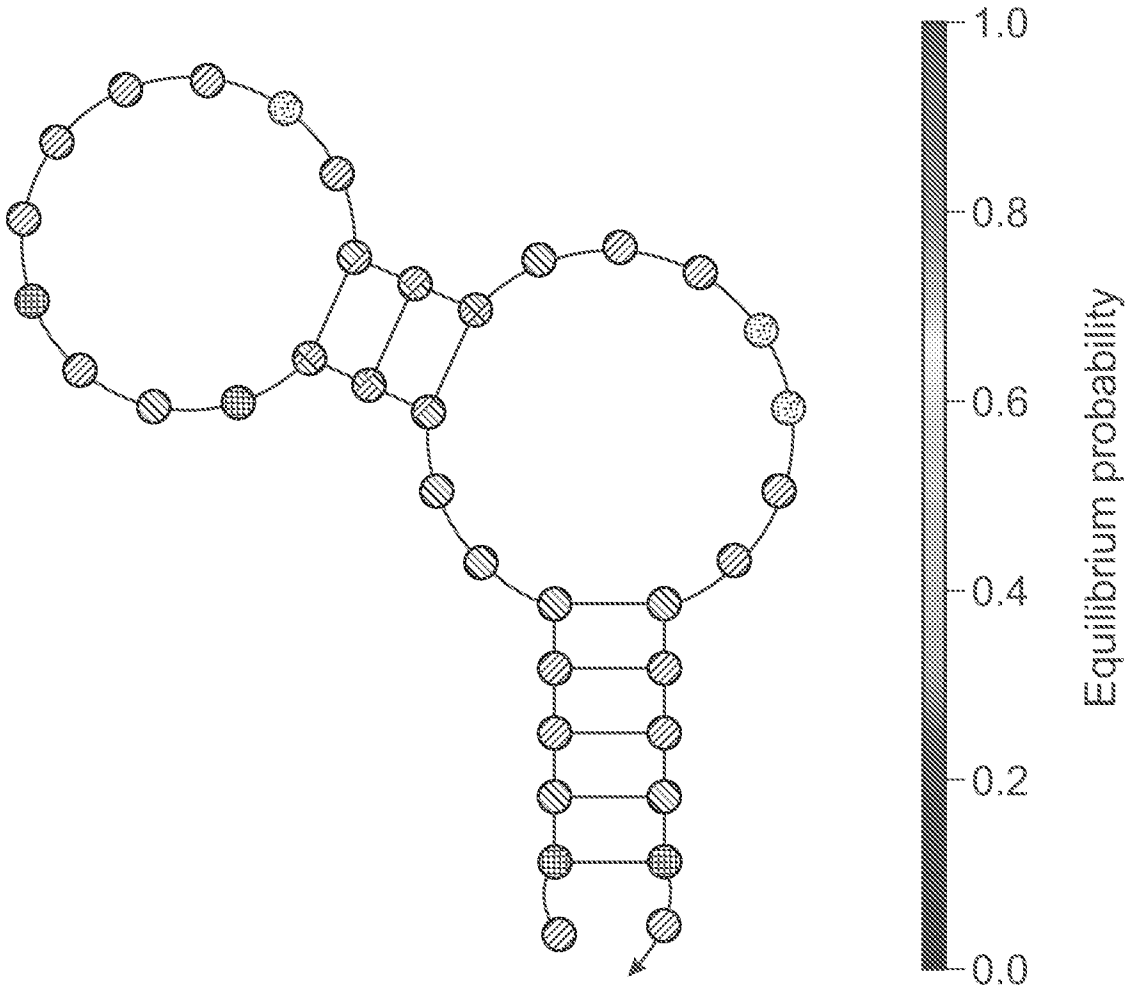

FIG. 1E

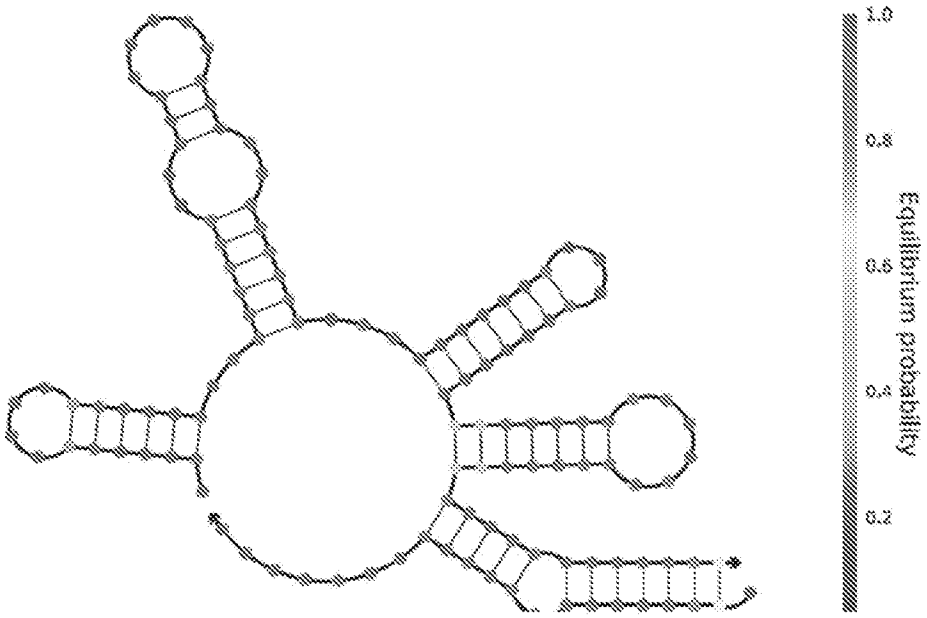

FIG. 1F

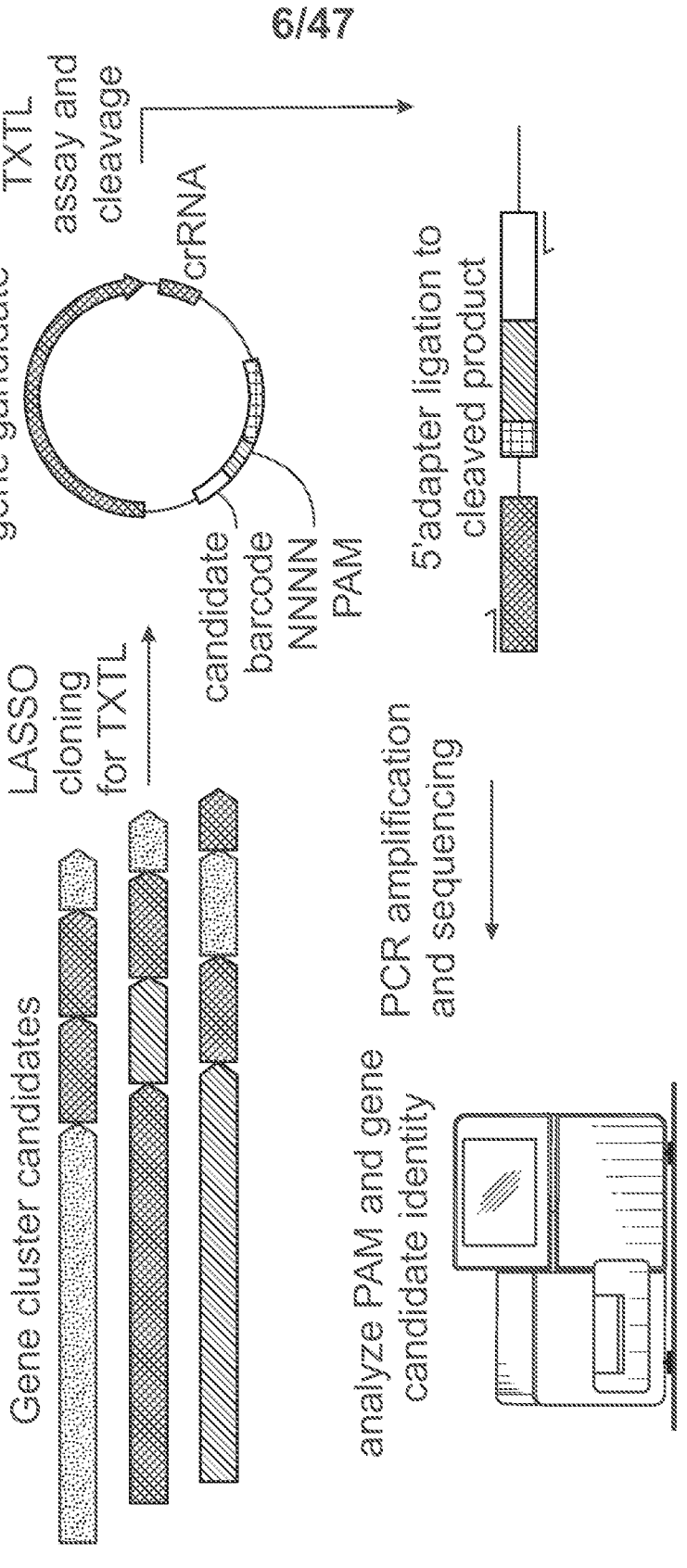
Gene cluster candidates
LASSO cloning for TXTL
gene gandidate
candidate barcode NNNN PAM
crRNA
TXTL assay and cleavage
5'adapter ligation to cleaved product
PCR amplification and sequencing
analyze PAM and gene candidate identity

FIG. 2

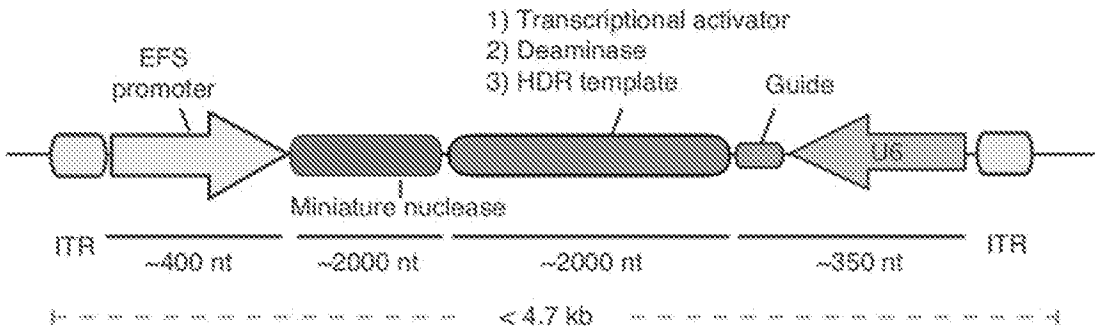
1) Transcriptional activator
2) Deaminase
3) HDR template
EFS promoter
Guide
pA
Miniature nuclease
ITR
~400 nt
~2000 nt
~2000 nt
~350 nt
ITR
< 4.7 kb

FIG. 3A

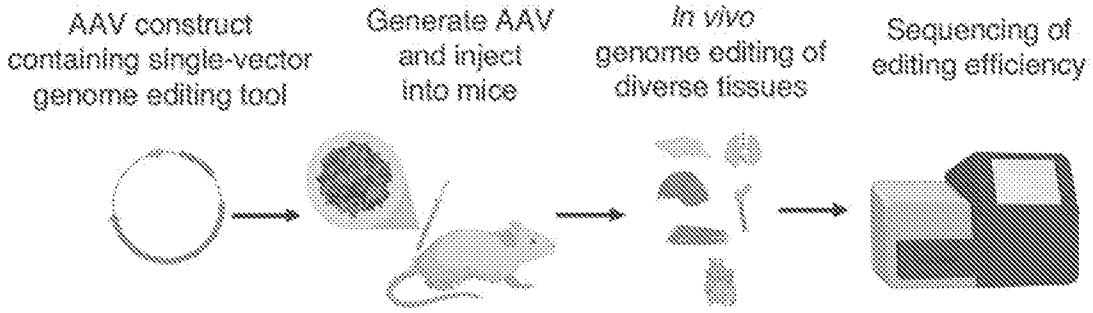
AAV construct containing single-vector genome editing tool
Generate AAV and inject into mice
In vivo genome editing of diverse tissues
Sequencing of editing efficiency

FIG. 3B

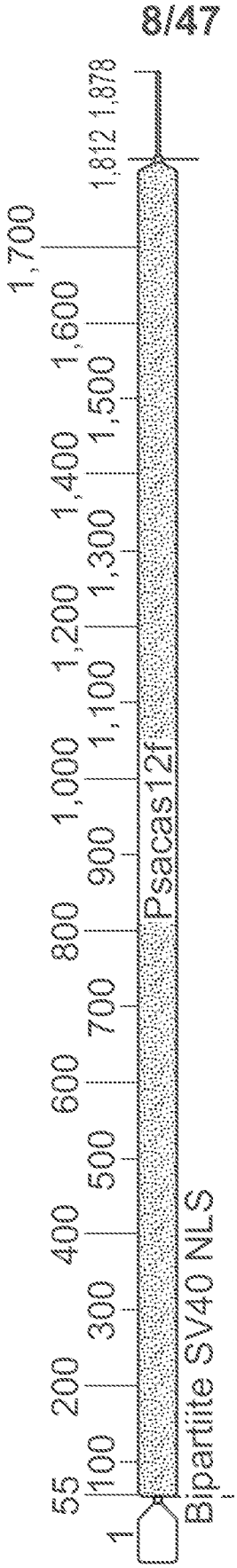
1
55
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,300
1,400
1,500
1,600
1,700
1,812 1,878
Bipartiite SV40 NLS
Psacas12f

FIG. 3C

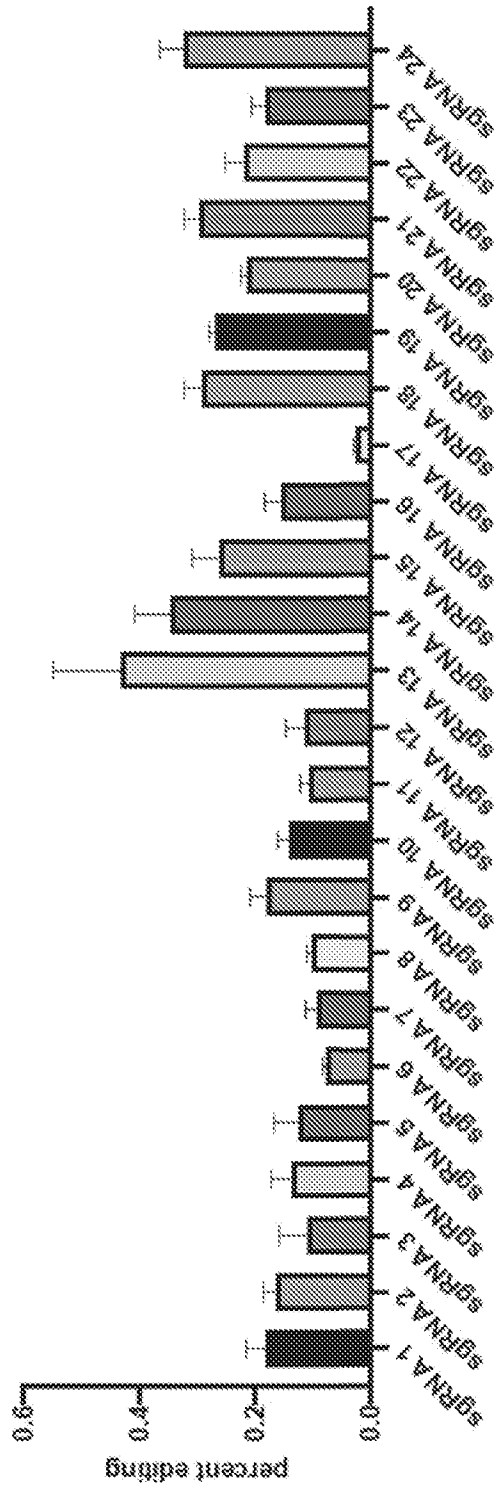
percent editing
0.0
0.2
0.4
0.6
sgRNA 1
sgRNA 2
sgRNA 3
sgRNA 4
sgRNA 5
sgRNA 6
sgRNA 7
sgRNA 8
sgRNA 9
sgRNA 10
sgRNA 11
sgRNA 12
sgRNA 13
sgRNA 14
sgRNA 15
sgRNA 16
sgRNA 17
sgRNA 18
sgRNA 19
sgRNA 20
sgRNA 21
sgRNA 22
sgRNA 23
sgRNA 24

FIG. 4

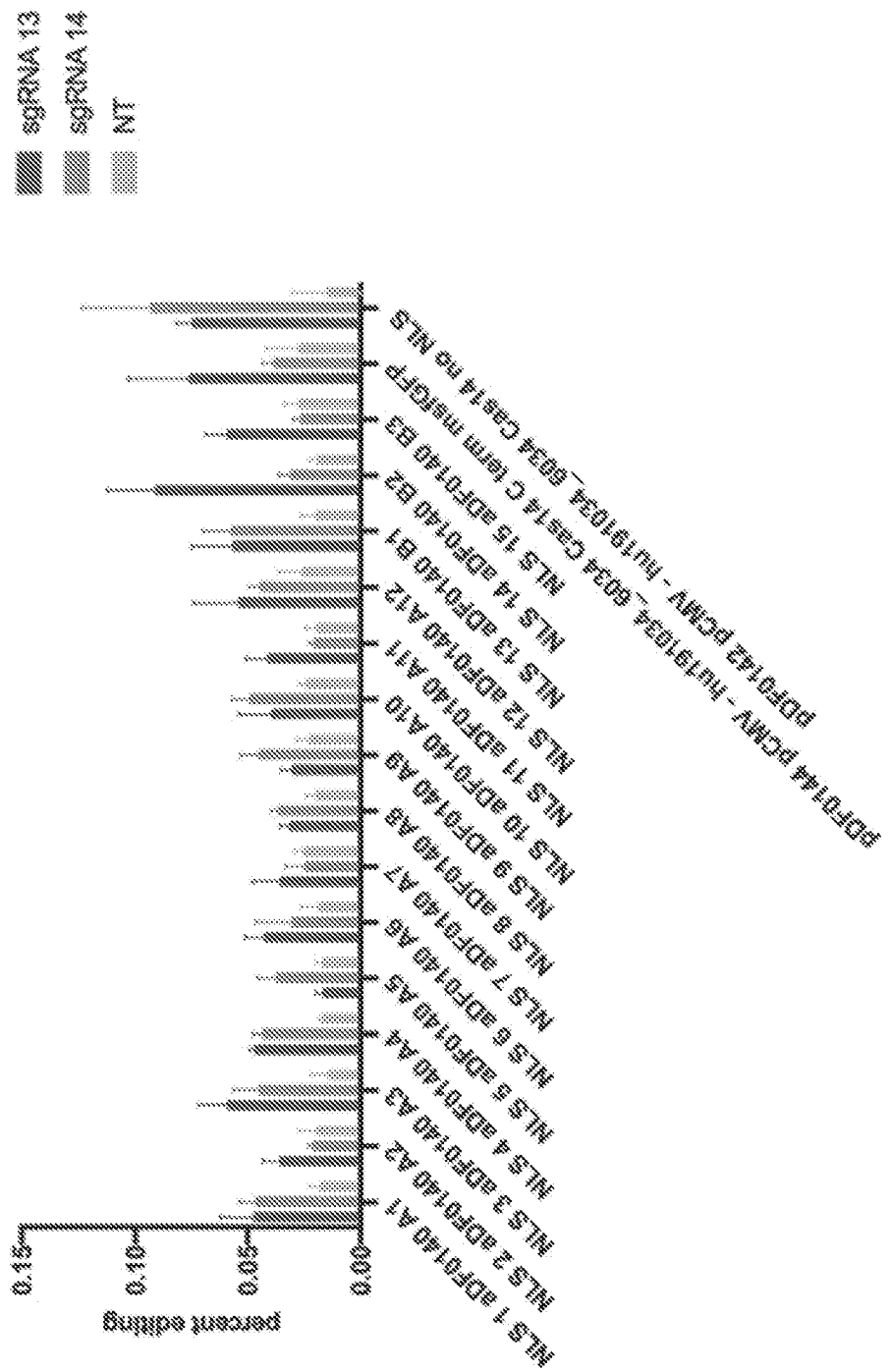
sgRNA 13
sgRNA 14
NT
percent editing
0.15
0.10
0.05
0.00
NLS 1 ADF0140 A1
NLS 2 ADF0140 A2
NLS 3 ADF0140 A3
NLS 4 ADF0140 A4
NLS 5 ADF0140 A5
NLS 6 ADF0140 A6
NLS 7 ADF0140 A7
NLS 8 ADF0140 A8
NLS 9 ADF0140 A9
NLS 10 ADF0140 A10
NLS 11 ADF0140 A11
NLS 12 ADF0140 A12
NLS 13 ADF0140 B1
NLS 14 ADF0140 B2
NLS 15 ADF0140 B3
pDF0144 pCMV - hu19103d_6034 Cas14 C (term mutGFP
pDF0142 pCMV - hu19103d_6034 Cas14 no NLS

FIG. 5A

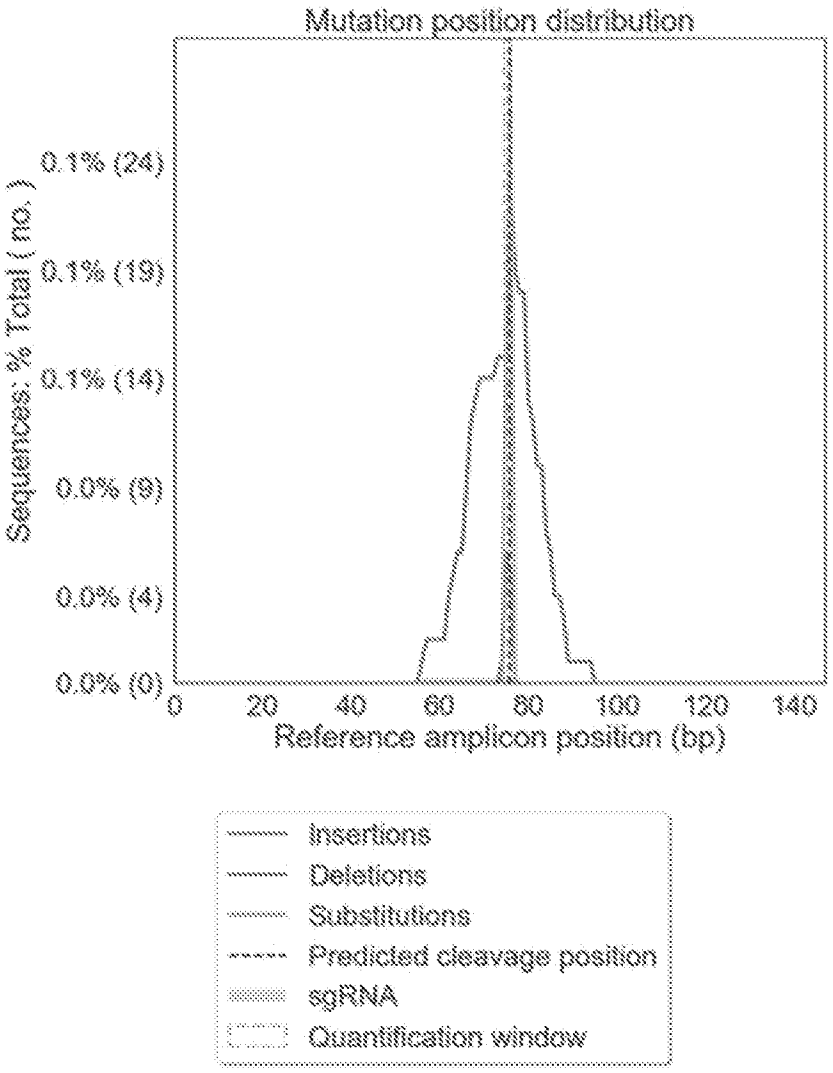
Mutation position distribution
Sequences; % Total (no.)
0.1% (24)
0.1% (19)
0.1% (14)
0.0% (9)
0.0% (4)
0.0% (0)
0
20
40
60
80
100
120
140
Reference amplicon position (bp)
Insertions
Deletions
Substitutions
Predicted cleavage position
sgRNA
Quantification window

FIG. 5B

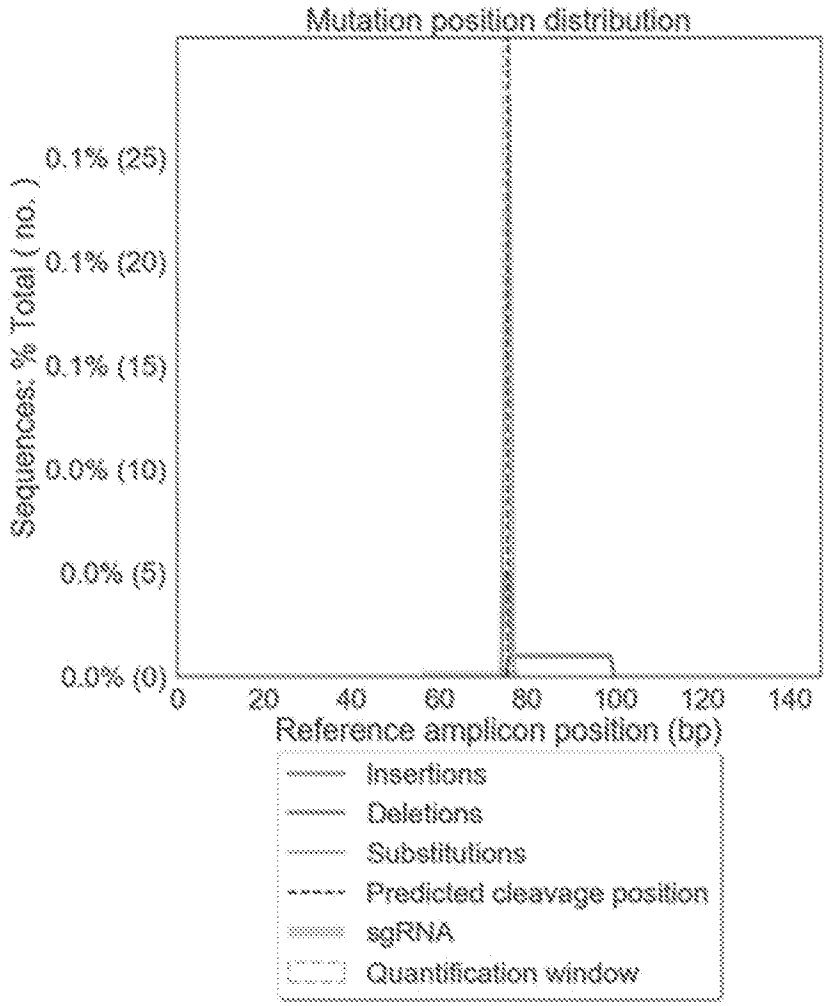
Mutation position distribution
Sequences: % Total (no.)
0.1% (25)
0.1% (20)
0.1% (15)
0.0% (10)
0.0% (5)
0.0% (0)
0
20
40
60
80
100
120
140
Reference amplicon position (bp)
Insertions
Deletions
Substitutions
Predicted cleavage position
sgRNA
Quantification window

FIG. 5C

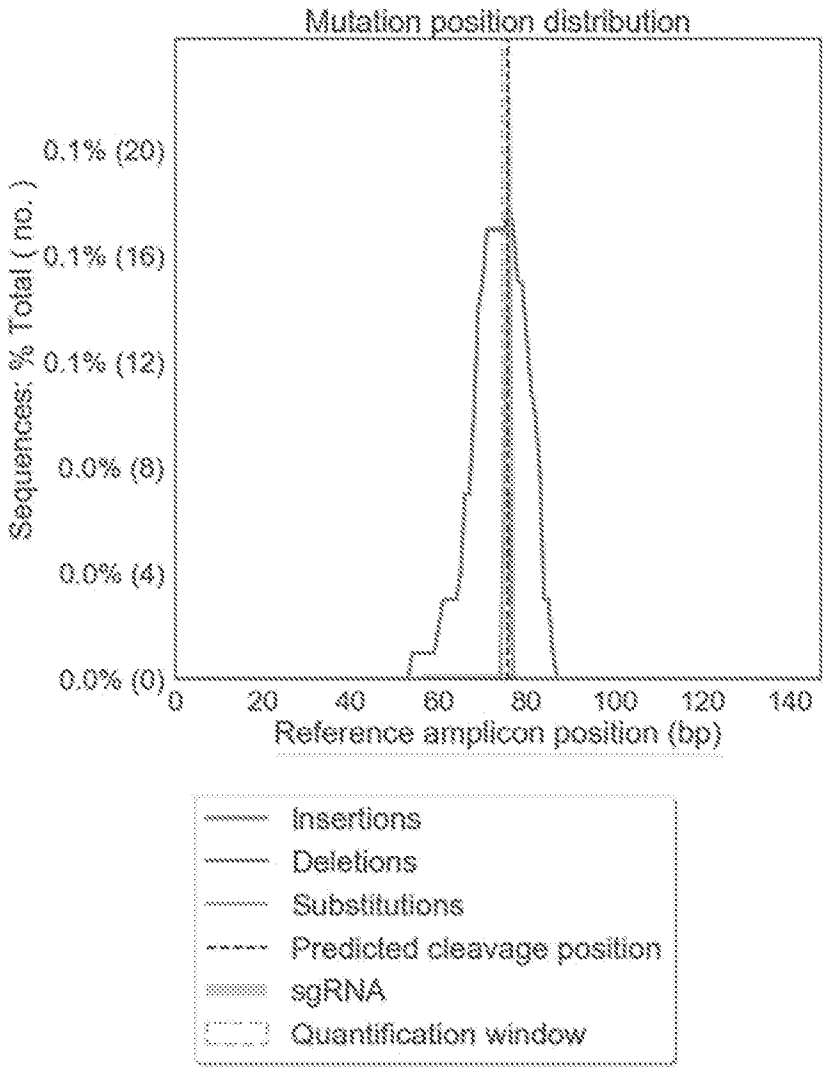
Mutation position distribution
Sequences; % Total (no.)
0.1% (20)
0.1% (16)
0.1% (12)
0.0% (8)
0.0% (4)
0.0% (0)
0
20
40
60
80
100
120
140
Reference amplicon position (bp)
Insertions
Deletions
Substitutions
Predicted cleavage position
sgRNA
Quantification window

FIG. 5D

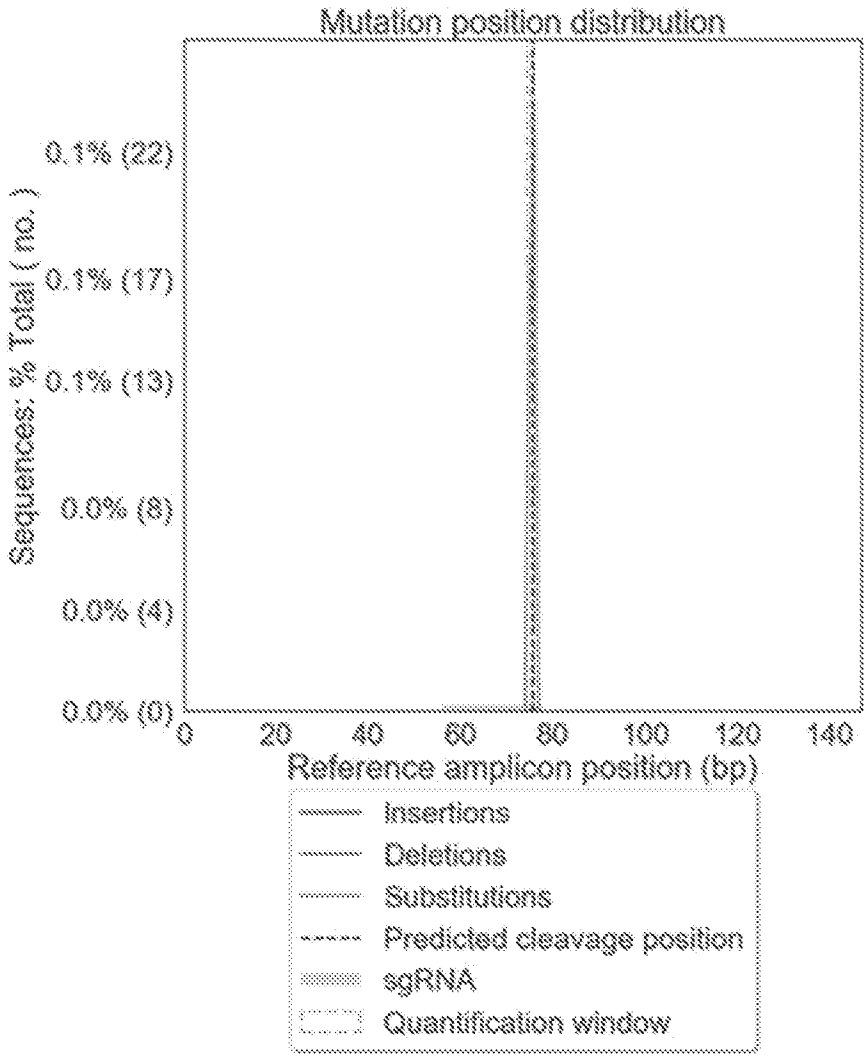
Mutation position distribution
Sequences; % Total (no.)
0.1% (22)
0.1% (17)
0.1% (13)
0.0% (8)
0.0% (4)
0.0% (0)
0
20
40
60
80
100
120
140
Reference amplicon position (bp)
Insertions
Deletions
Substitutions
Predicted cleavage position
sgRNA
Quantification window

FIG. 5E

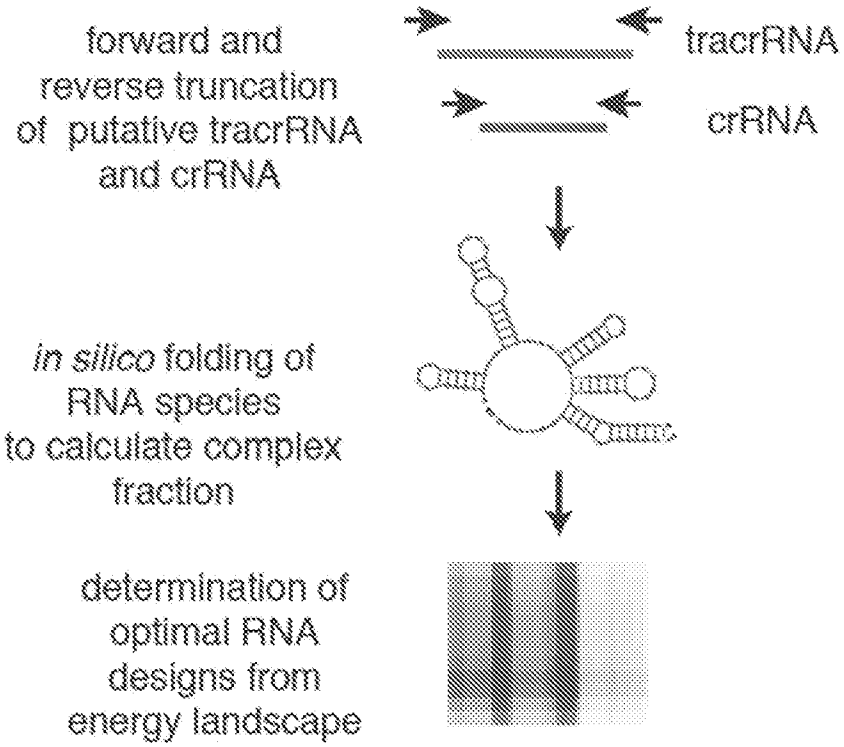
forward and
reverse truncation
of putative tracrRNA
and crRNA
tracrRNA
crRNA
in silico folding of
RNA species
to calculate complex
fraction
determination of
optimal RNA
designs from
energy landscape

FIG. 6A

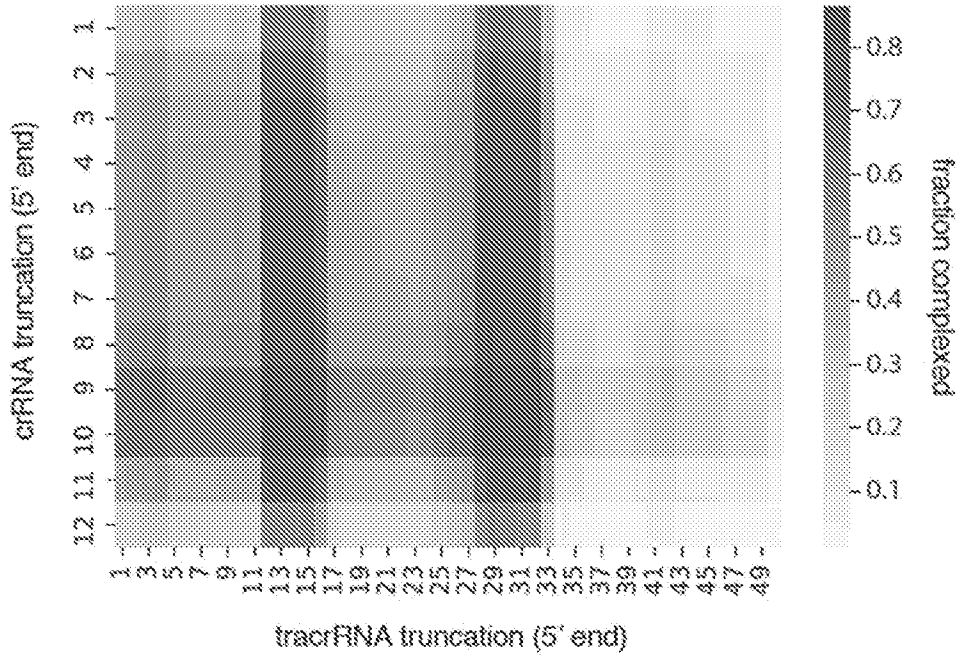
crRNA truncation (5' end)
12 11 10 9 8 7 6 5 4 3 2 1
tracrRNA truncation (5' end)
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49
fraction complexed
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1

FIG. 6B

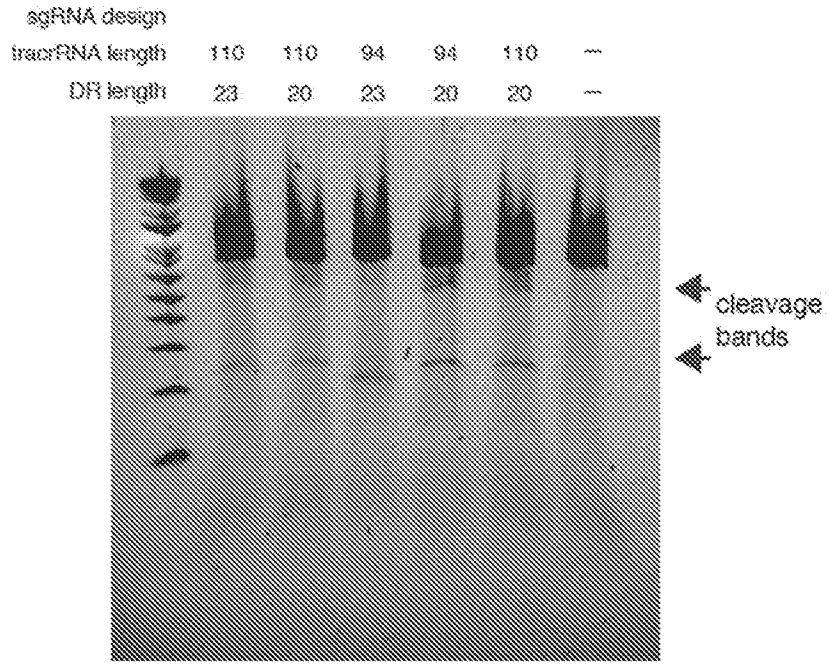
sgRNA design
tracrRNA length 110 110 94 94 110 —
DR length 23 20 23 20 20 —
cleavage bands

FIG. 6C

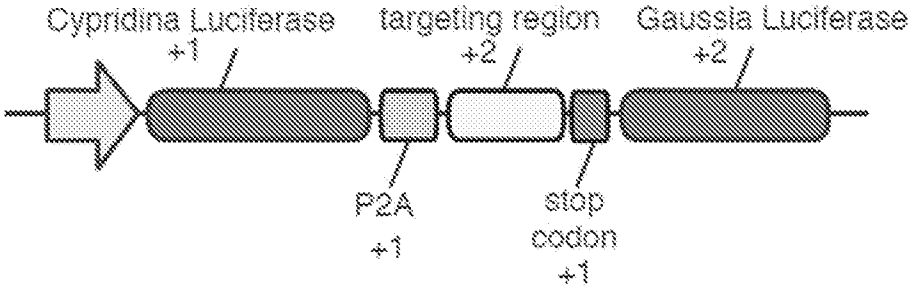
Cypridina Luciferase
+1
targeting region
+2
Gaussia Luciferase
+2
P2A
+1
stop
codon
+1

FIG. 7A

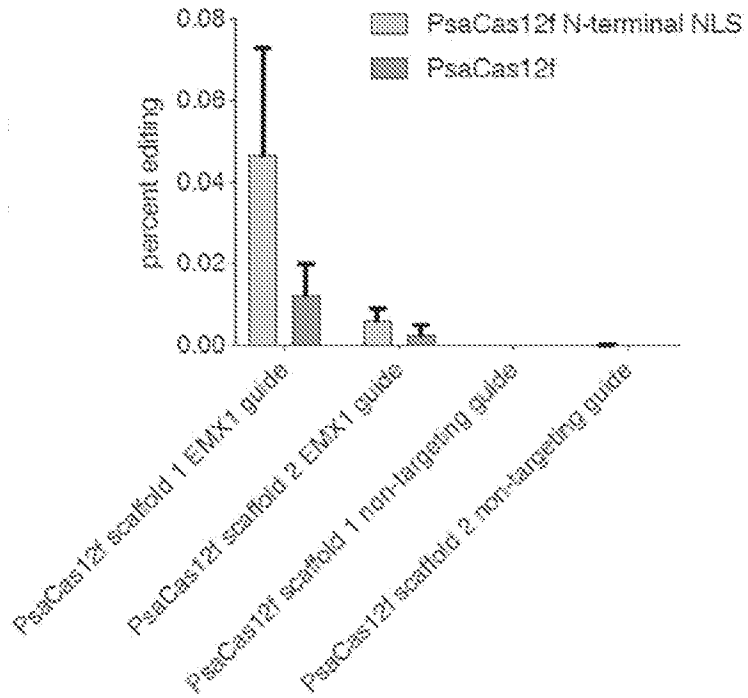
PsaCas12f N-terminal NLS
PsaCas12f
percent editing
0.08
0.06
0.04
0.02
0.00
PsaCas12f scaffold 1 EMX1 guide
PsaCas12f scaffold 2 EMX1 guide
PsaCas12f scaffold 1 non-targeting guide
PsaCas12f scaffold 2 non-targeting guide

FIG. 7B

Consensus
Coverage
127
0
EMX1 reference
89.05% No nucleotide variant(27,806 reads,including 525 with sequencing errors)
0.71% Frame shift 26 bp deletion(223 reads,including 6 with sequencing errors)
0.59% Frame shift 14 bp deletion(183 reads,including 7 with sequencing errors)
0.56% Frame shift 23 bp deletion(176 reads,including 1 with sequencing errors)
0.50% In frame protien varieni 30 bp deletion(156 reads,including 1 with sequenci.
0.38% Frame shift 26 bp deletion(118 reads)
0.31% Frame shift 34 bp deletion(97 reads)
0.29% Frame shift 19 bp deletion(92 reads)
0.24% Frame shift 34 bp deletion(74 reads)
0.23% Frame shift 13 bp deletion(71 reads)
0.21% In Frame stop codon 6 bp deletion(67 reads)
0.20% In Frame protien variant 15 bp deletion(64 reads)
0.20% Frame shift 10 bp deletion(62 reads,including 1 with sequencing errors)
0.19% Frame shift 22 bp deletion(59 reads,including 1 with sequencing errors)
0.14% Frame shift 7 bp deletion(45 reads)
0.12% Frame shift 25 bp deletion(39 reads)
0.12% Frame shift 16 bp deletion(37 reads)
0.11% Frame shift 35 bp deletion(35 reads)
0.11% In frame protien variant 21 bp deletion(34 reads)
0.11% In frame protien varieni 21 bp deletion(33 reads,including 1 with sequencin....
0.11% Frame shift 8 bp deletion(33 reads,including 1 with sequencing errors)
0.10% In Frame stop codon 9 bp deletion(31 reads)
0.09% Frame shift 37 bp deletion(29 reads)
0.09% Frame shift 22 bp deletion(27 reads)

FIG. 7C

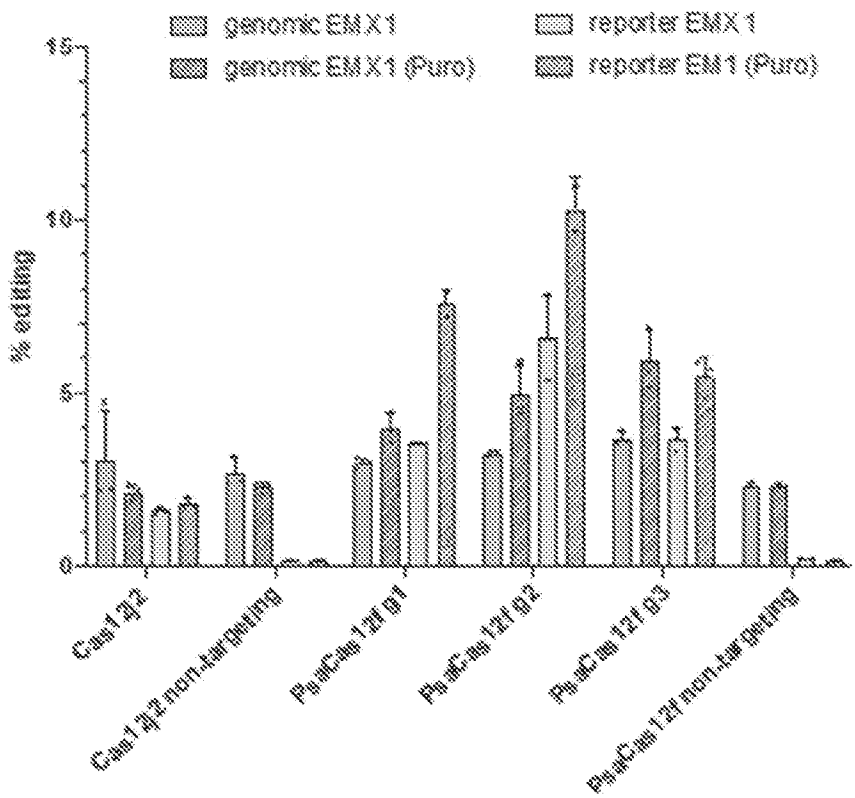
genomic EMX1
genomic EMX1 (Puro)
reporter EMX1
reporter EM1 (Puro)
% editing
15
10
5
0

FIG. 7D

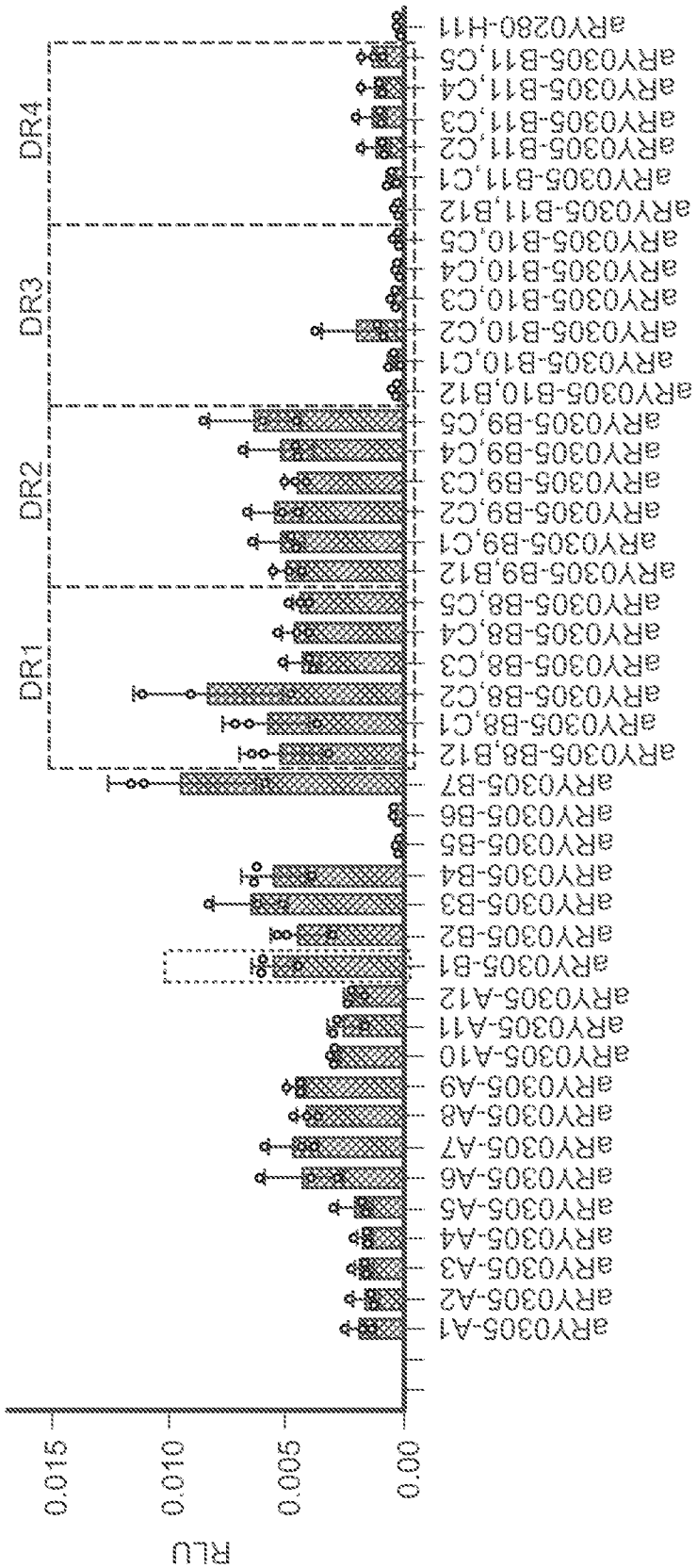
DR1
DR2
DR3
DR4
aRY0305-A1
aRY0305-A2
aRY0305-A3
aRY0305-A4
aRY0305-A5
aRY0305-A6
aRY0305-A7
aRY0305-A8
aRY0305-A9
aRY0305-A10
aRY0305-A11
aRY0305-A12
aRY0305-B1
aRY0305-B2
aRY0305-B3
aRY0305-B4
aRY0305-B5
aRY0305-B6
aRY0305-B7
aRY0305-B8,B12
aRY0305-B8,C1
aRY0305-B8,C2
aRY0305-B8,C3
aRY0305-B8,C4
aRY0305-B8,C5
aRY0305-B9,B12
aRY0305-B9,C1
aRY0305-B9,C2
aRY0305-B9,C3
aRY0305-B9,C4
aRY0305-B9,C5
aRY0305-B10,B12
aRY0305-B10,C1
aRY0305-B10,C2
aRY0305-B10,C3
aRY0305-B10,C4
aRY0305-B10,C5
aRY0305-B11,B12
aRY0305-B11,C1
aRY0305-B11,C2
aRY0305-B11,C3
aRY0305-B11,C4
aRY0305-B11,C5
aRY0280-H11
0.00
0.005
0.010
0.015
RLU

FIG. 7E

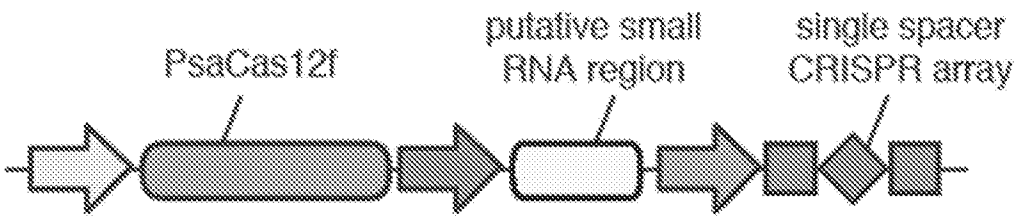
PsaCas12f
putative small RNA region
single spacer CRISPR array

FIG. 8A

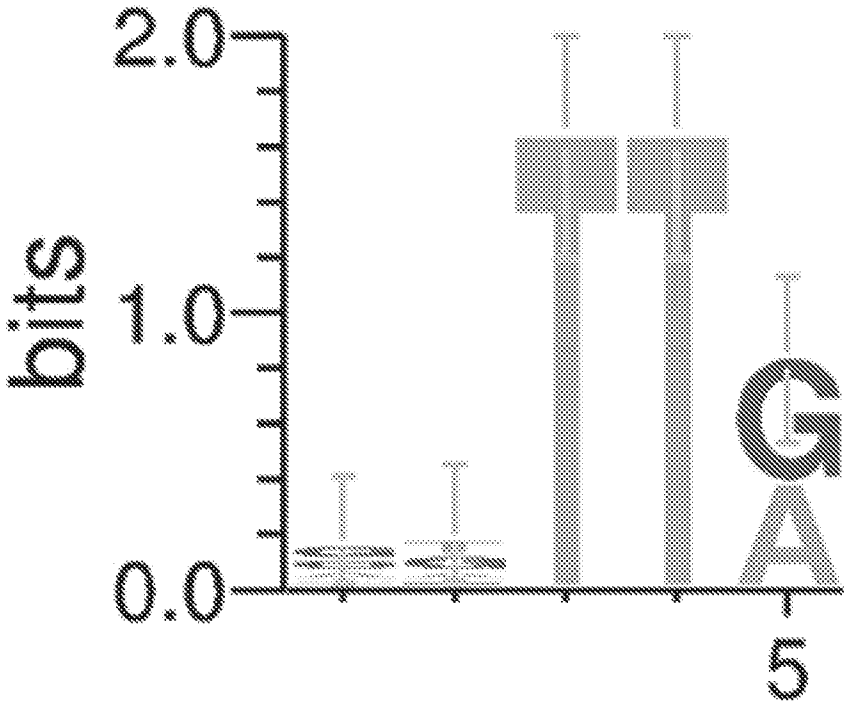

FIG. 8B

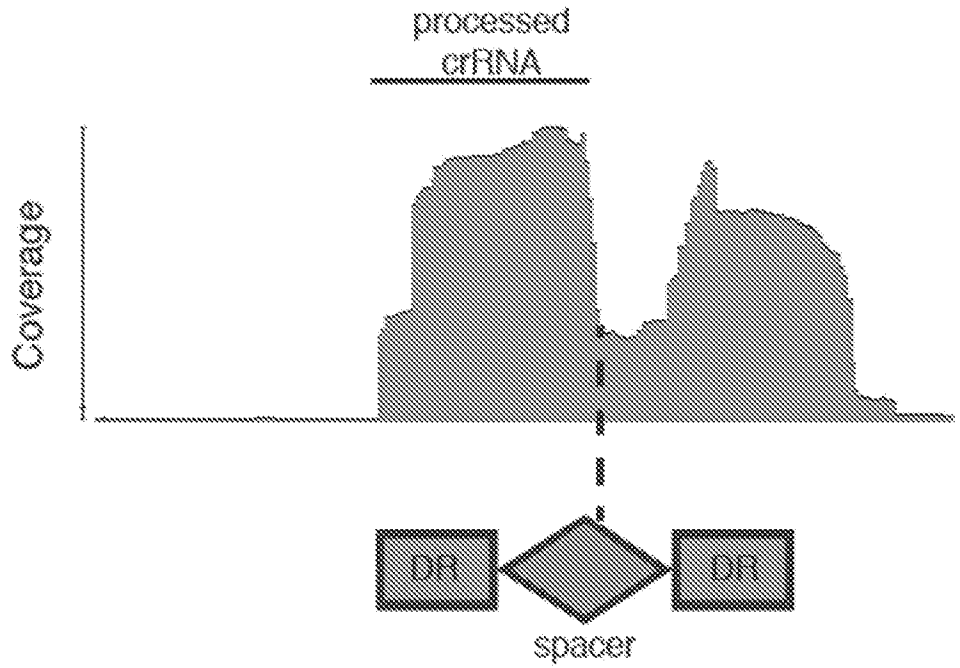

FIG. 8C

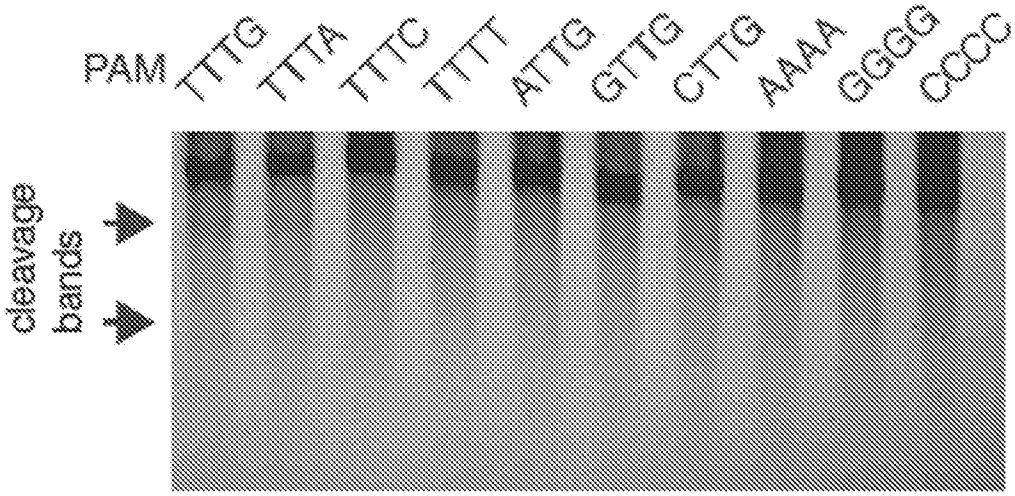
PAM
TTTG
TTTA
TTTC
TTTT
ATTG
GTTG
CTTG
AAAA
GGGG
CCCC
cleavage bands

FIG. 8D

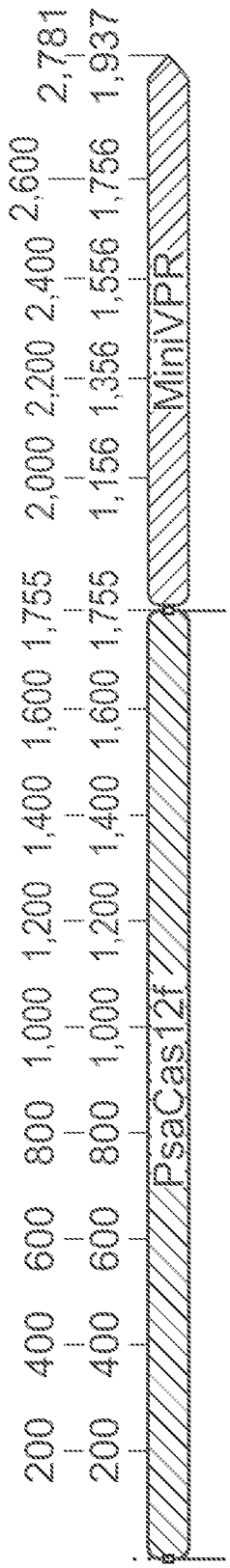
200
400
600
800
1,000
1,200
1,400
1,600
1,755
2,000
2,200
2,400
2,600
2,781
200
400
600
800
1,000
1,200
1,400
1,600
1,755
1,156
1,356
1,556
1,756
1,937
PsaCas12f
MinVPR

FIG. 9A

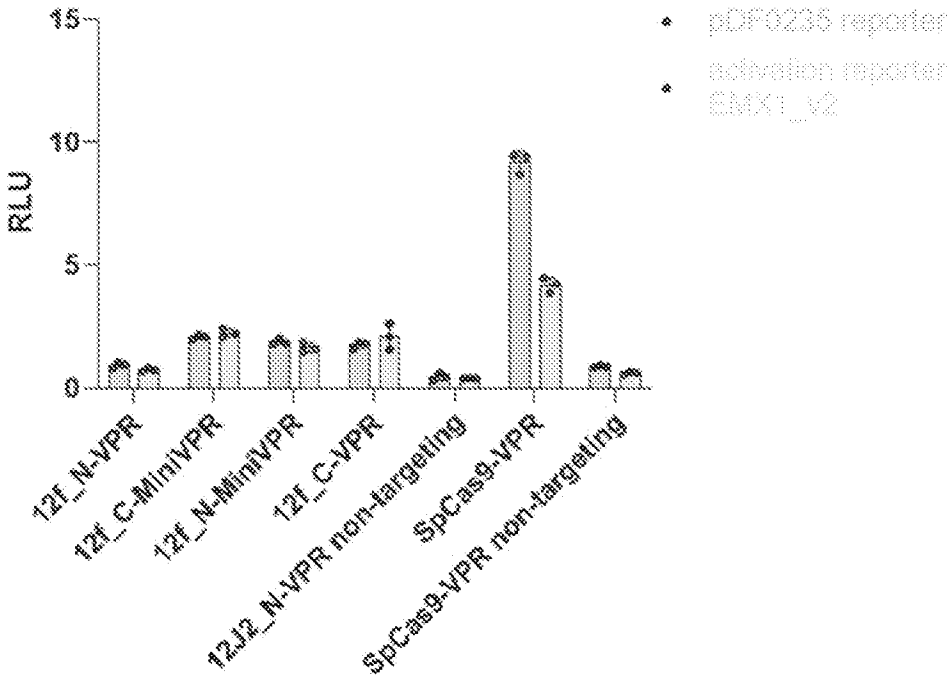
15
10
5
0
RLU
pDF0235 reporter
activation reporter
EMX1_v2
12f_N-VPR
12f_C-MiniVPR
12f_N-MiniVPR
12f_C-VPR
12J2_N-VPR non-targeting
SpCas9-VPR
SpCas9-VPR non-targeting

FIG. 9B

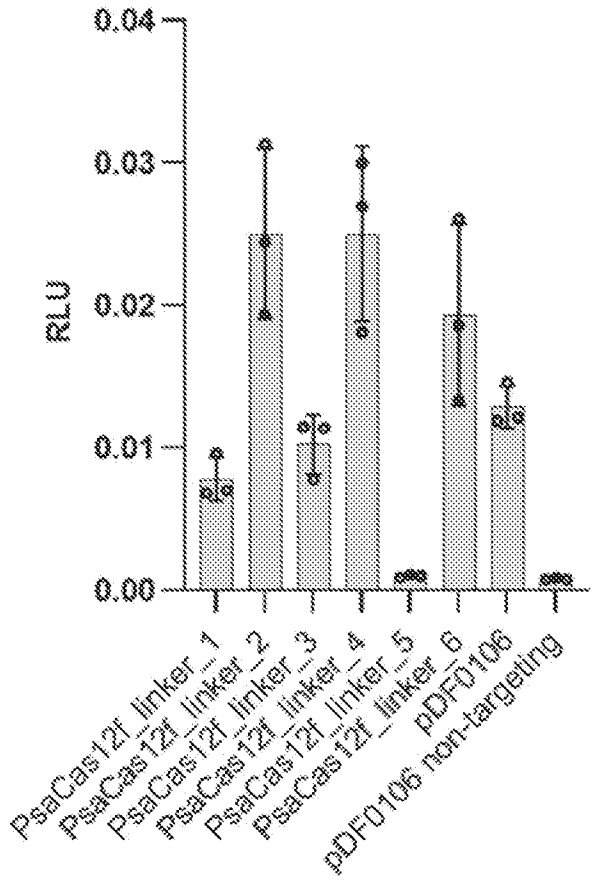
0.04
0.03
0.02
0.01
0.00
RLU
PsaCas12f_linker_1
psaCas12f_linker_2
PsaCas12f_linker_3
psaCas12f_linker_4
PsaCas12f_linker_5
psaCas12f_linker_6
pDF0106
pDF0106 non-targeting

FIG. 9C

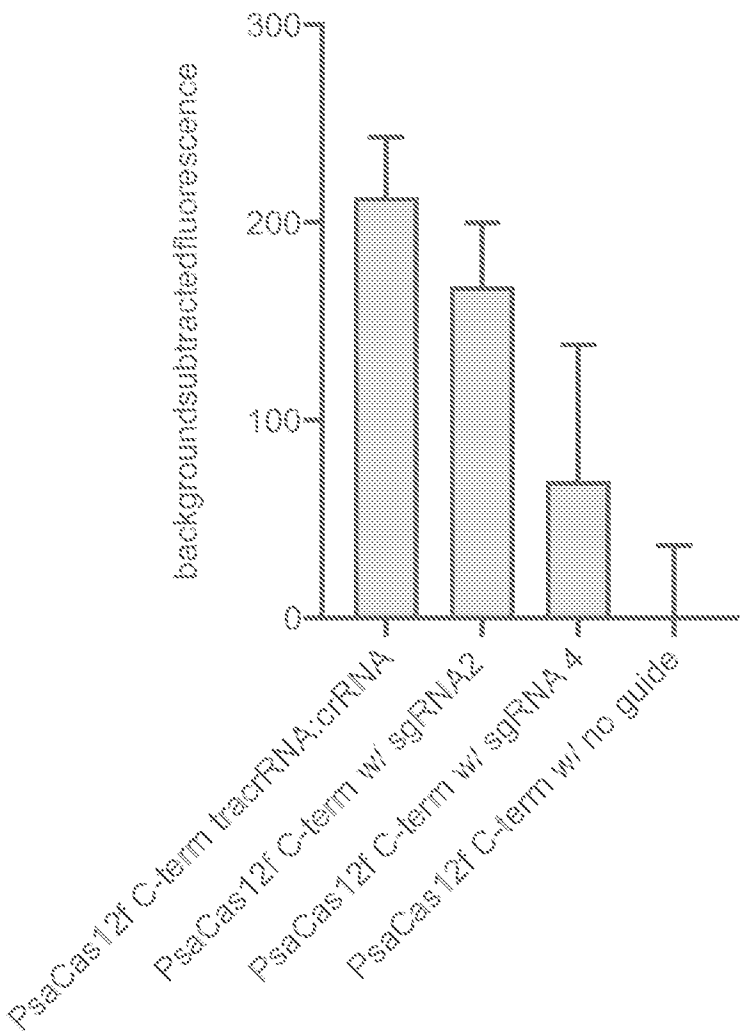
300
200
100
0
backgroundsubtractedfluorescence
PsaCas12f C-term tracrRNA:crRNA
PsaCas12f C-term w/ sgRNA2
PsaCas12f C-term w/ sgRNA 4
PsaCas12f C-term w/ no guide

FIG. 9D

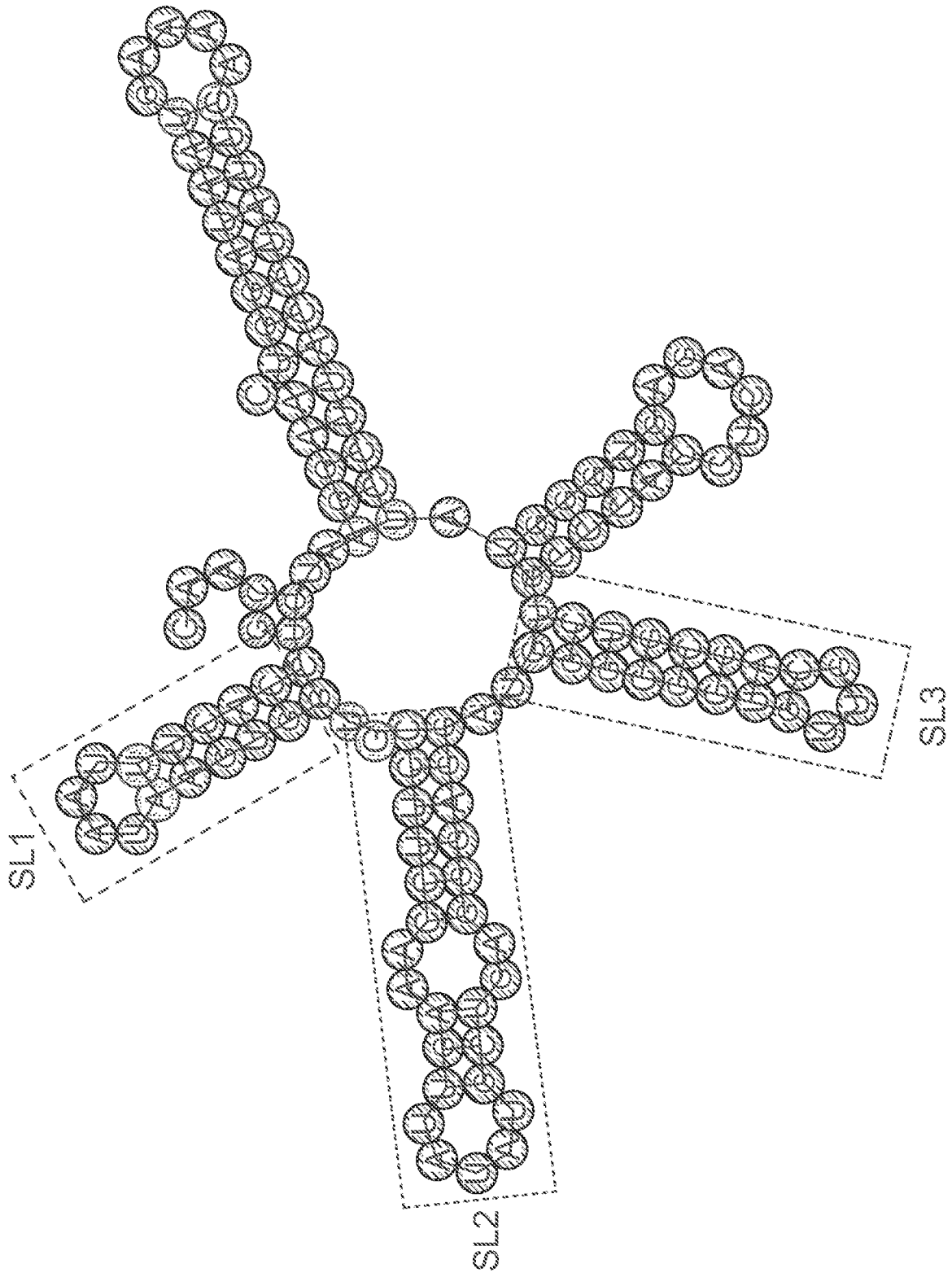

FIG. 10A

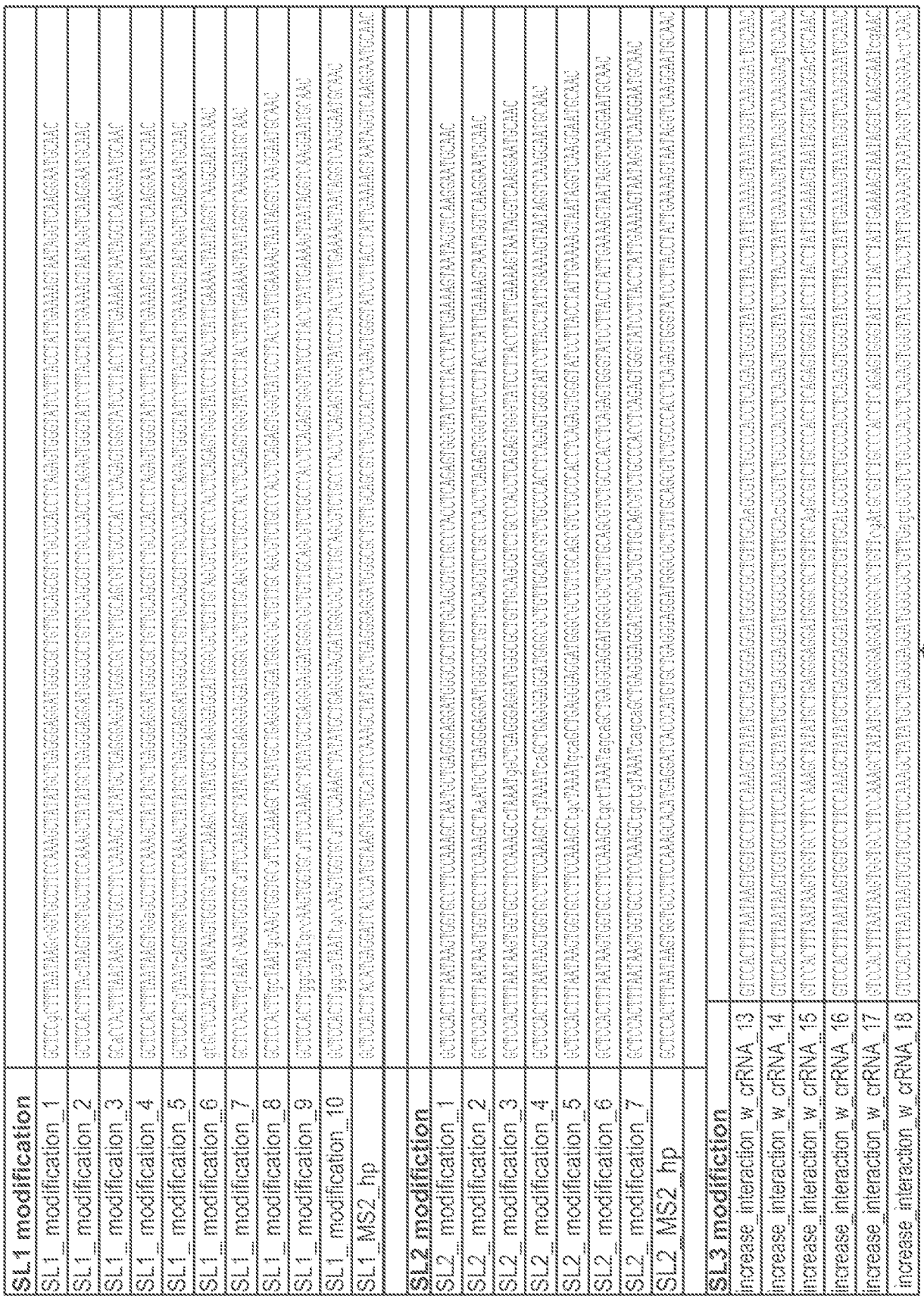
SL1 modification
SL1_modification_1
SL1_modification_2
SL1_modification_3
SL1_modification_4
SL1_modification_5
SL1_modification_6
SL1_modification_7
SL1_modification_8
SL1_modification_9
SL1_modification_10
SL1_MS2_hp
SL2 modification
SL2_modification_1
SL2_modification_2
SL2_modification_3
SL2_modification_4
SL2_modification_5
SL2_modification_6
SL2_modification_7
SL2_MS2_hp
SL3 modification
increase_interaction_w_crRNA_13
increase_interaction_w_crRNA_14
increase_interaction_w_crRNA_15
increase_interaction_w_crRNA_16
increase_interaction_w_crRNA_17
increase_interaction_w_crRNA_18
A

FIG. 10B

Ⓐ

increase_interaction_w_crRNA_19	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCgtGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAAcGCAAC
increase_interaction_w_crRNA_20	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGtAtGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATaCAAC
increase_interaction_w_crRNA_21	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGccGCAtGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCggC
increase_interaction_w_crRNA_22	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGcgGCAtGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCcgC
increase_interaction_w_crRNA_23	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCgGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAAcGCAAC
increase_interaction_w_crRNA_24	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGtAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATaCAAC
increase_interaction_w_crRNA_25	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGccGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCggC
increase_interaction_w_crRNA_26	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGcgGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCcgC
SL4 modifiction	
increase_interaction_of_SL4_3	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACgCTAGAcGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_4	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACtgCTAGAcaGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_5	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACctgCTAGAcagGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_6	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACgCTCAGAcGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_7	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACtgCTCAGAcaGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_8	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACctgCTCAGAcagGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_9	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACgctgCTCAGAcagcGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL4_10	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACtgctgCTCAGAcagcaGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SL3_MS2_hp	GCTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACACATGAGGATCACCCATGTGTGGGTATCCTTACCTATTGAAAAGTAATAGGTCAAGGAATGCAAC
SL5 modifiction	
increase_interaction_of_SL5_4	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTaAAAGTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_5	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGgAAAGcTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_6	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGctAAAGagTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_7	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGtgAAAGcaTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_8	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGctgAAAGcagTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_9	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGgctgAAAGcagcTAATAGGTCAAGGAATGCAAC
increase_interaction_of_SL5_10	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTGtgctgAAAGcagcaTAATAGGTCAAGGAATGCAAC
SL4_MS2_hp	GTCCACTTTAATAAGTGGTGCCTTCCAAAGCTATATGCTGAGGGAGGATGGGCGCTGTTGCAGCGTCTGCCCACCTCAGAGTGGGTATCCTTACCTATTACATGAGGATCACCCATGTAATAGGTCAAGGAATGCAAC

FIG. 10B(Continued)

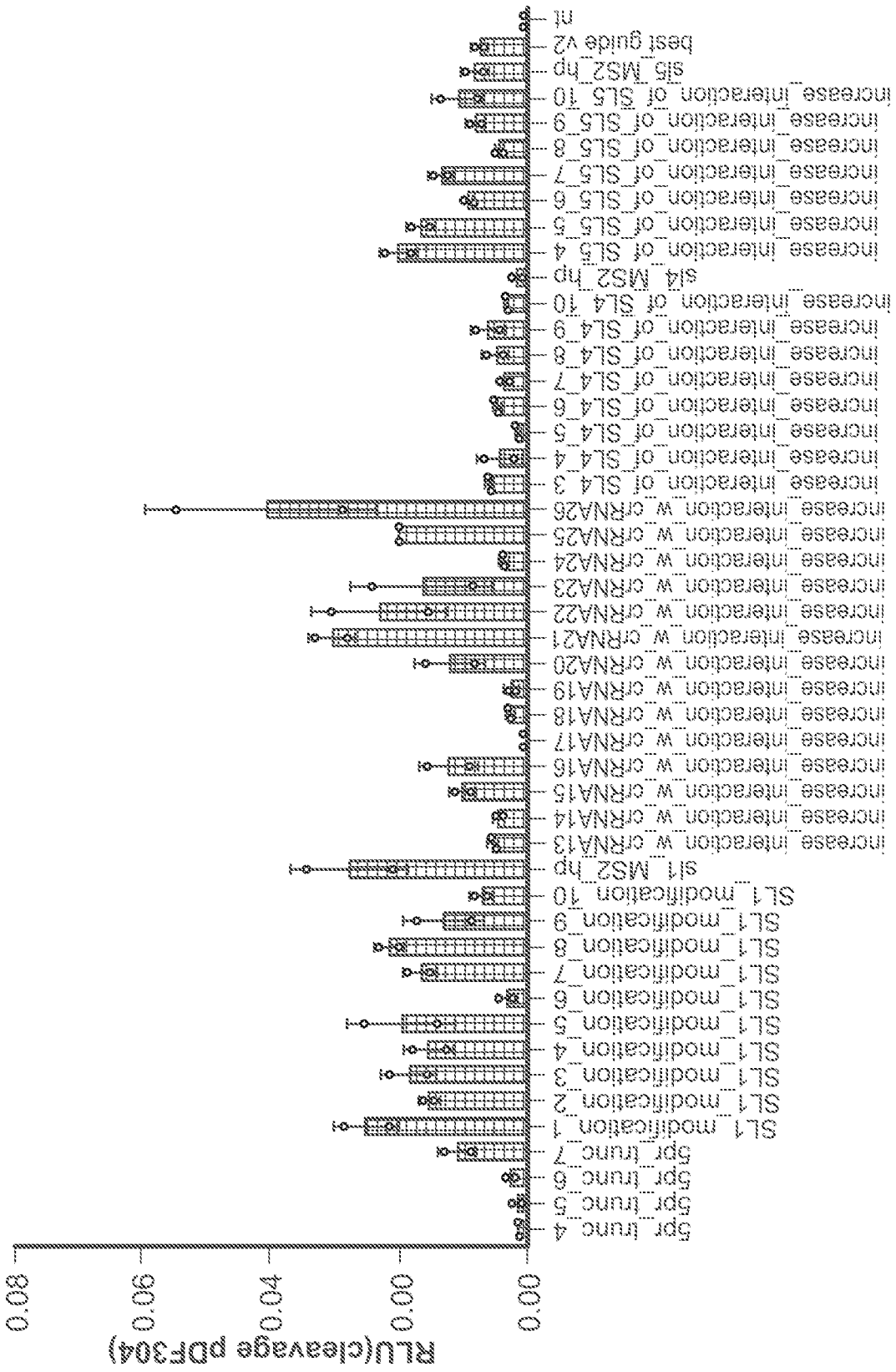
RLU(cleavage pDF304)
0.08
0.06
0.04
0.00
0.00
5pr_trunc_4
5pr_trunc_5
5pr_trunc_6
5pr_trunc_7
SL1_modification_1
SL1_modification_2
SL1_modification_3
SL1_modification_4
SL1_modification_5
SL1_modification_6
SL1_modification_7
SL1_modification_8
SL1_modification_9
SL1_modification_10
sl1_MS2_hp
increase_interaction_w_crRNA13
increase_interaction_w_crRNA14
increase_interaction_w_crRNA15
increase_interaction_w_crRNA16
increase_interaction_w_crRNA17
increase_interaction_w_crRNA18
increase_interaction_w_crRNA19
increase_interaction_w_crRNA20
increase_interaction_w_crRNA21
increase_interaction_w_crRNA22
increase_interaction_w_crRNA23
increase_interaction_w_crRNA24
increase_interaction_w_crRNA25
increase_interaction_w_crRNA26
increase_interaction_of_SL4_3
increase_interaction_of_SL4_4
increase_interaction_of_SL4_5
increase_interaction_of_SL4_6
increase_interaction_of_SL4_7
increase_interaction_of_SL4_8
increase_interaction_of_SL4_9
increase_interaction_of_SL4_10
sl4_MS2_hp
increase_interaction_of_SL5_4
increase_interaction_of_SL5_5
increase_interaction_of_SL5_6
increase_interaction_of_SL5_7
increase_interaction_of_SL5_8
increase_interaction_of_SL5_9
increase_interaction_of_SL5_10
sl5_MS2_hp
best guide v2
nt

FIG. 10C

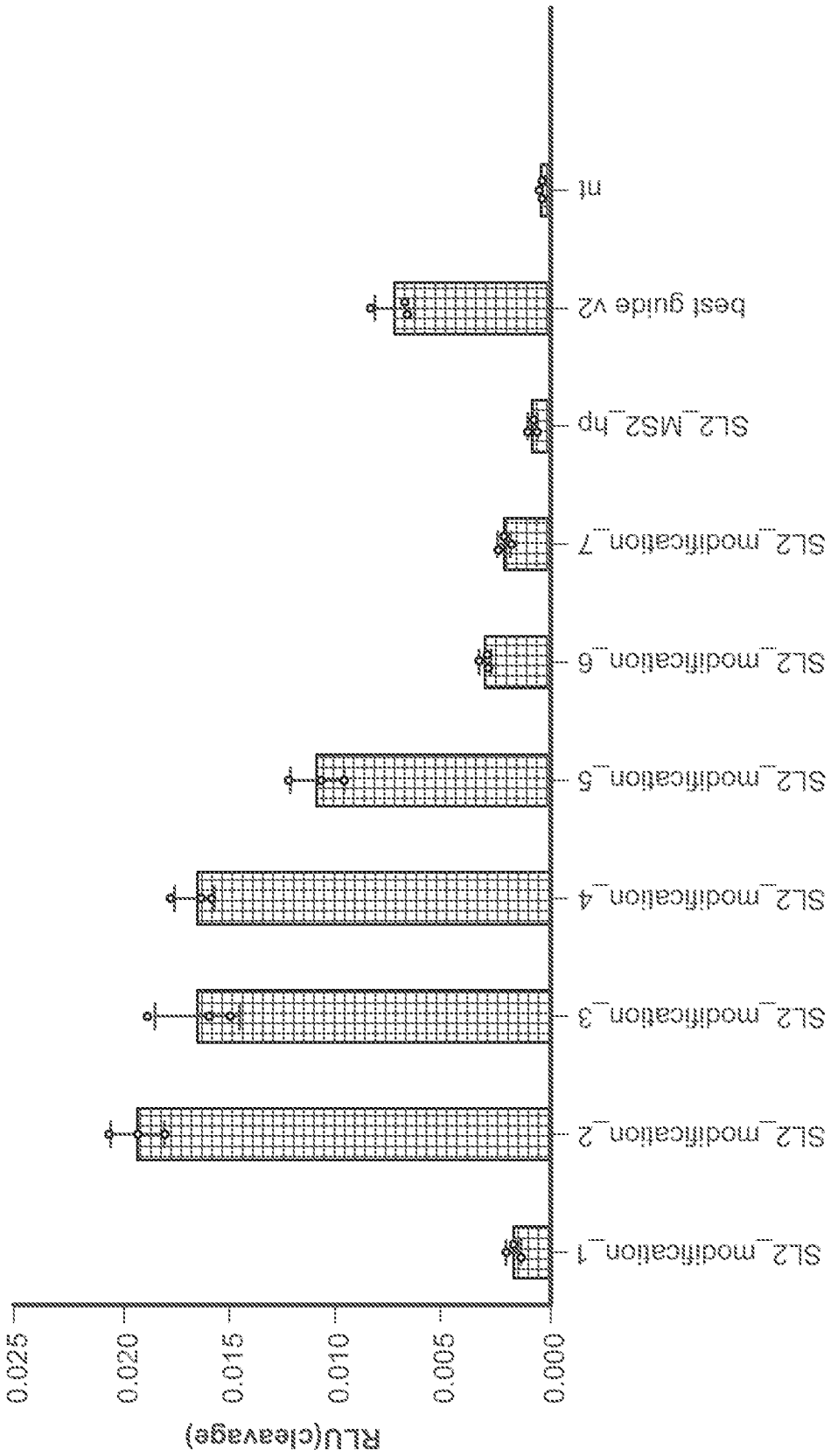

FIG. 10C
(Continued)

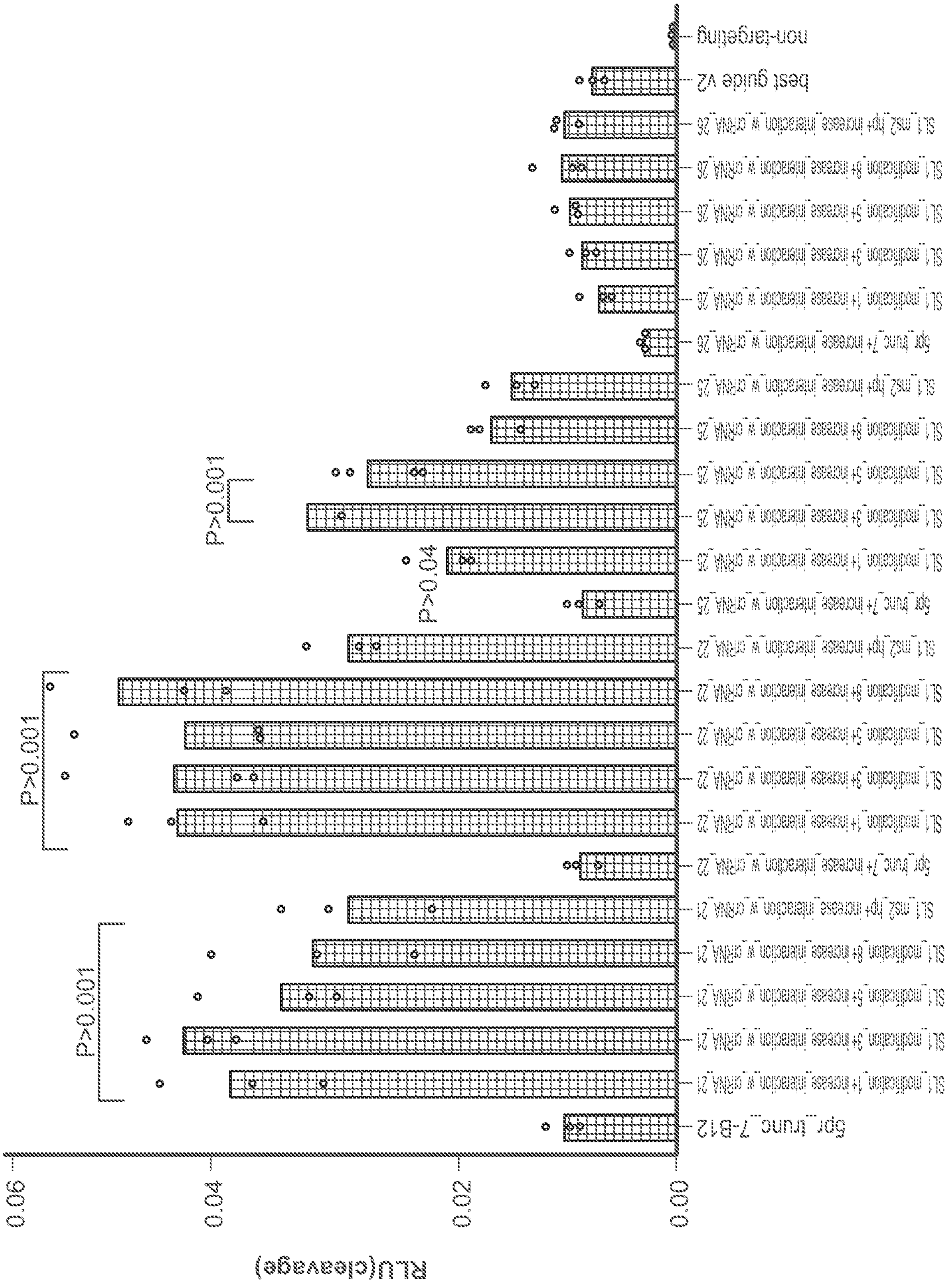
RLU(cleavage)
0.00
0.02
0.04
0.06
P>0.001
P>0.001
P>0.04
P>0.001
5pr_trunc_7-B12
best guide v2
non-targeting

FIG. 11A

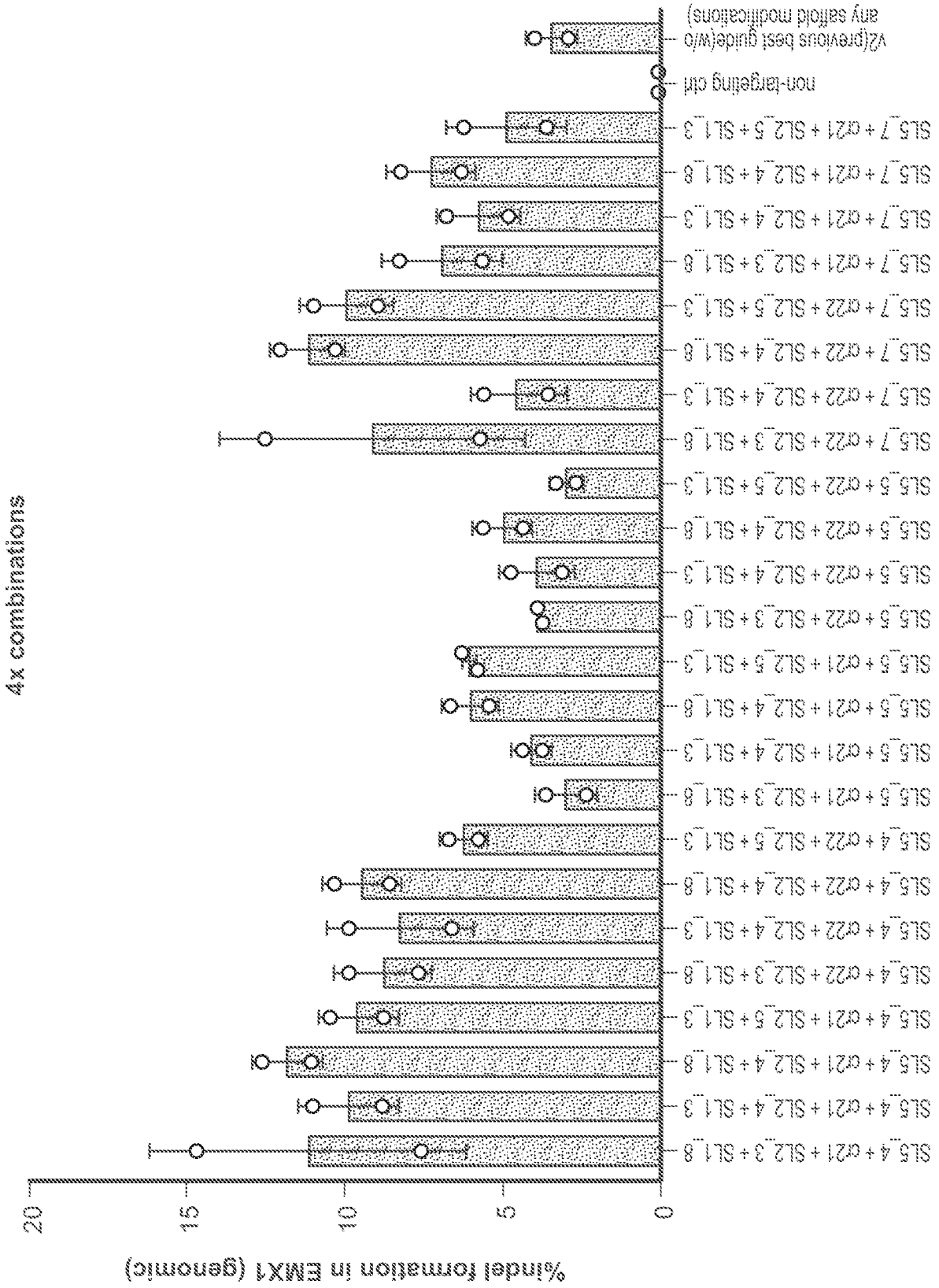
%indel formation in EMX1 (genomic)
0
5
10
15
20
4x combinations
SL5_4 + α21 + SL2_3 + SL1_8
SL5_4 + α21 + SL2_4 + SL1_3
SL5_4 + α21 + SL2_4 + SL1_8
SL5_4 + α21 + SL2_5 + SL1_3
SL5_4 + α22 + SL2_3 + SL1_8
SL5_4 + α22 + SL2_4 + SL1_3
SL5_4 + α22 + SL2_4 + SL1_8
SL5_4 + α22 + SL2_5 + SL1_3
SL5_5 + α21 + SL2_3 + SL1_8
SL5_5 + α21 + SL2_4 + SL1_3
SL5_5 + α21 + SL2_4 + SL1_8
SL5_5 + α21 + SL2_5 + SL1_3
SL5_5 + α22 + SL2_3 + SL1_8
SL5_5 + α22 + SL2_4 + SL1_3
SL5_5 + α22 + SL2_4 + SL1_8
SL5_5 + α22 + SL2_5 + SL1_3
SL5_7 + α22 + SL2_3 + SL1_8
SL5_7 + α22 + SL2_4 + SL1_3
SL5_7 + α22 + SL2_4 + SL1_8
SL5_7 + α22 + SL2_5 + SL1_3
SL5_7 + α21 + SL2_3 + SL1_8
SL5_7 + α21 + SL2_4 + SL1_3
SL5_7 + α21 + SL2_4 + SL1_8
SL5_7 + α21 + SL2_5 + SL1_3
non-targeting crrl
v2(previous best guide(w/o any saRoid modifications)

FIG. 11B

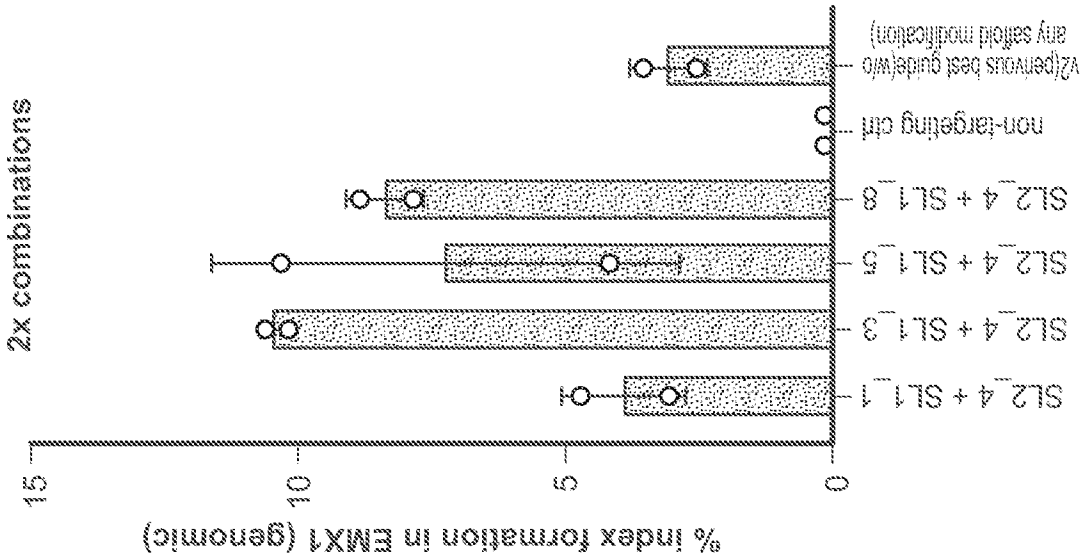

FIG. 11B
(Continued)

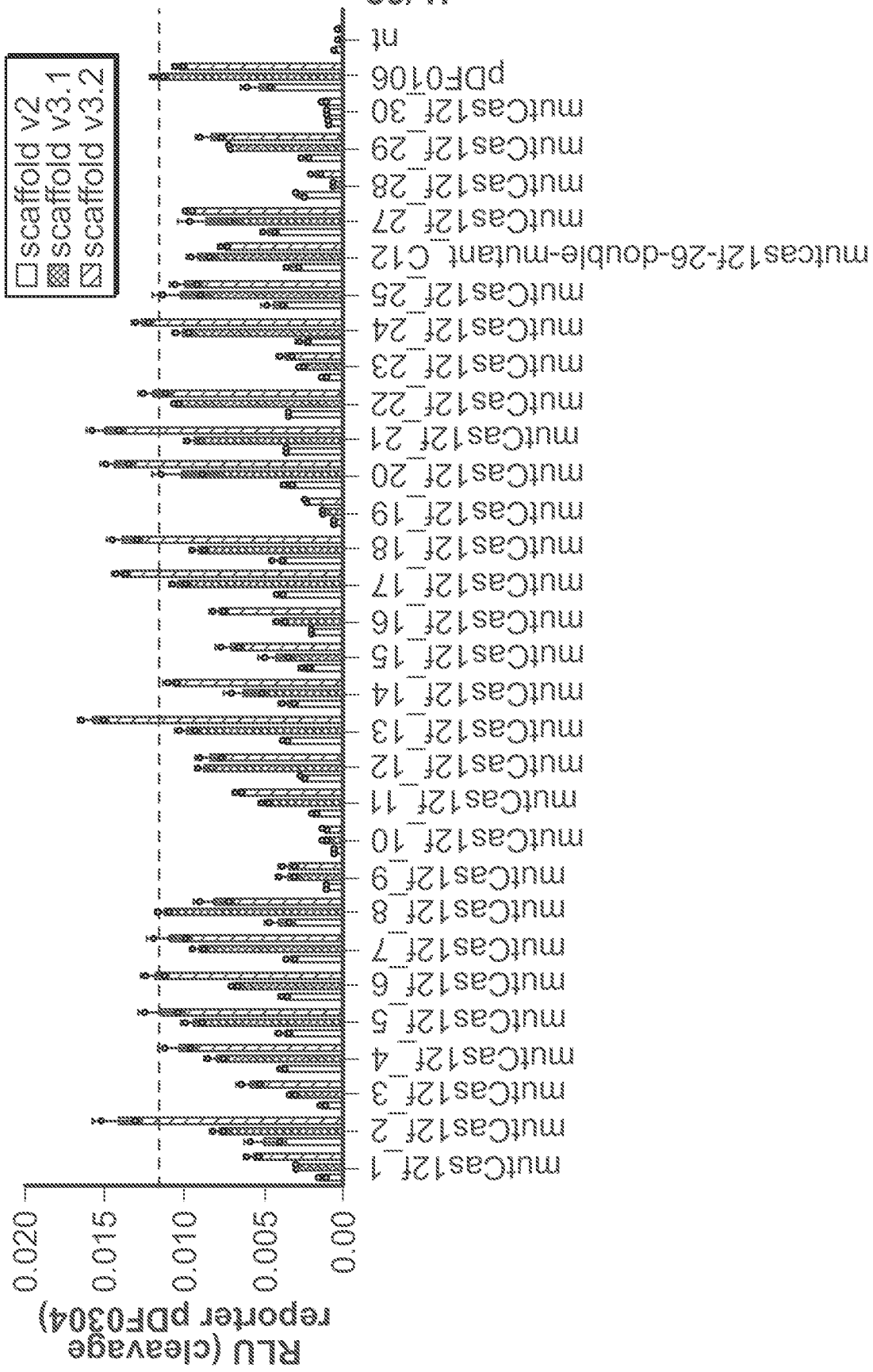
scaffold v2
scaffold v3.1
scaffold v3.2
RLU (cleavage reporter pDF0304)
0.020
0.015
0.010
0.005
0.00
nt
pDF0106
mutCas12f_30
mutCas12f_29
mutCas12f_28
mutCas12f_27
mutcas12f-26-double-mutant_C12
mutCas12f_25
mutCas12f_24
mutCas12f_23
mutCas12f_22
mutCas12f_21
mutCas12f_20
mutCas12f_19
mutCas12f_18
mutCas12f_17
mutCas12f_16
mutCas12f_15
mutCas12f_14
mutCas12f_13
mutCas12f_12
mutCas12f_11
mutCas12f_10
mutCas12f_9
mutCas12f_8
mutCas12f_7
mutCas12f_6
mutCas12f_5
mutCas12f_4
mutCas12f_3
mutCas12f_2
mutCas12f_1

FIG. 11C

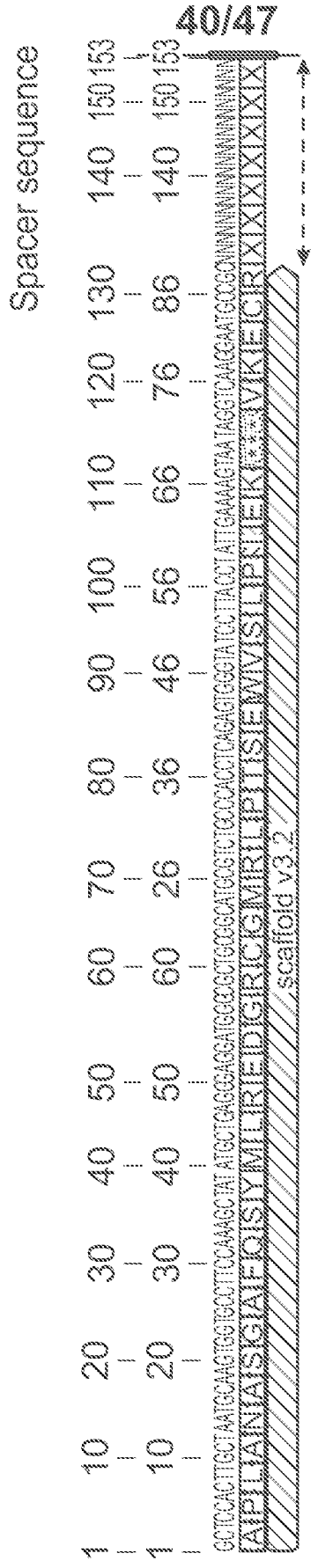
Spacer sequence
1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 153
1 10 20 30 40 50 60 26 36 46 56 66 76 86 140 150 153
scaffold v3.2

FIG. 12A

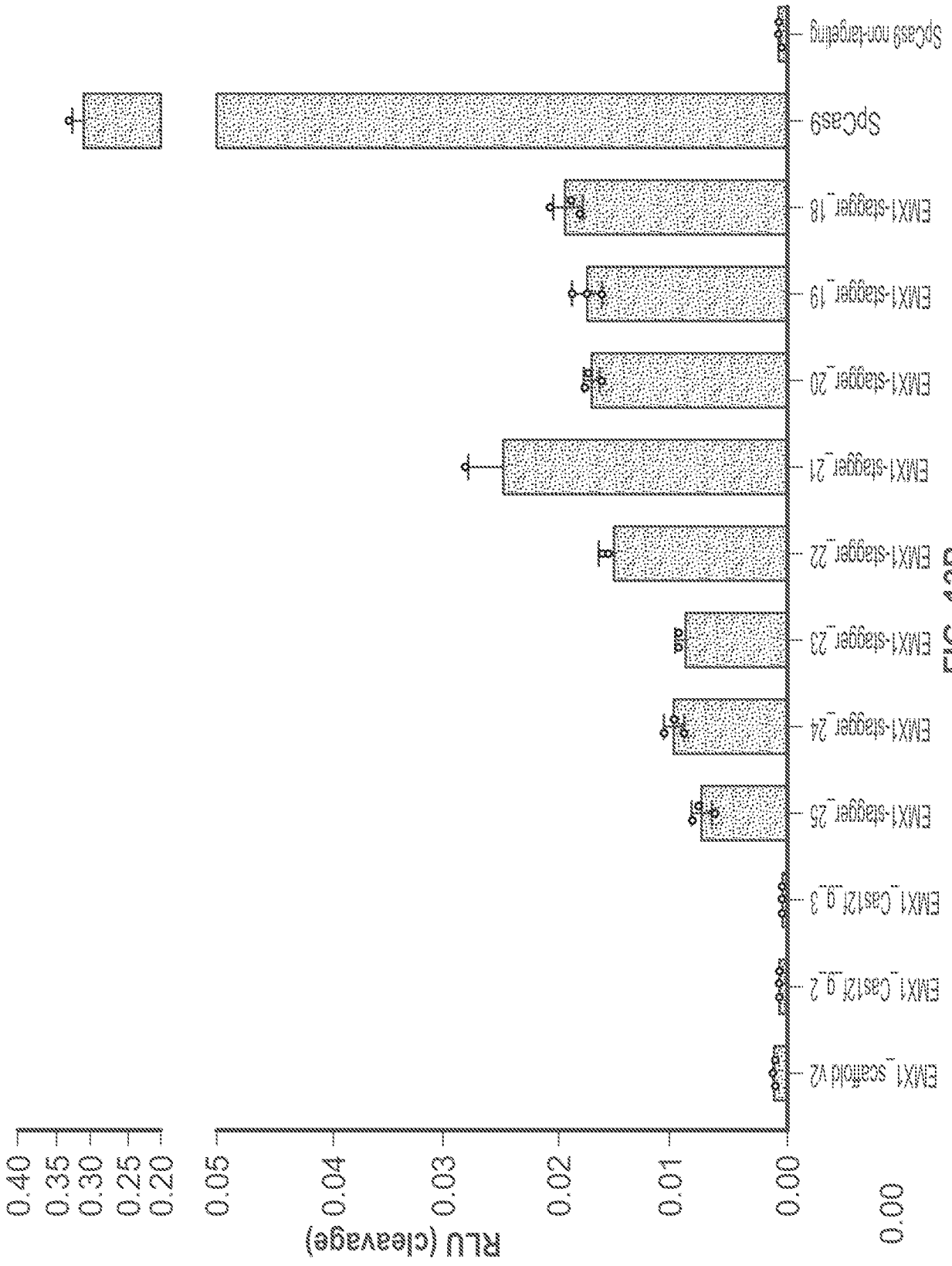
RLU (cleavage)
0.00
0.01
0.02
0.03
0.04
0.05
0.00
0.20
0.25
0.30
0.35
0.40
EMX1_scaffold V2
EMX1_Cas12i_g_2
EMX1_Cas12i_g_3
EMX1-stagger_25
EMX1-stagger_24
EMX1-stagger_23
EMX1-stagger_22
EMX1-stagger_21
EMX1-stagger_20
EMX1-stagger_19
EMX1-stagger_18
SpCas9
SpCas9 non-targeting

FIG. 12B

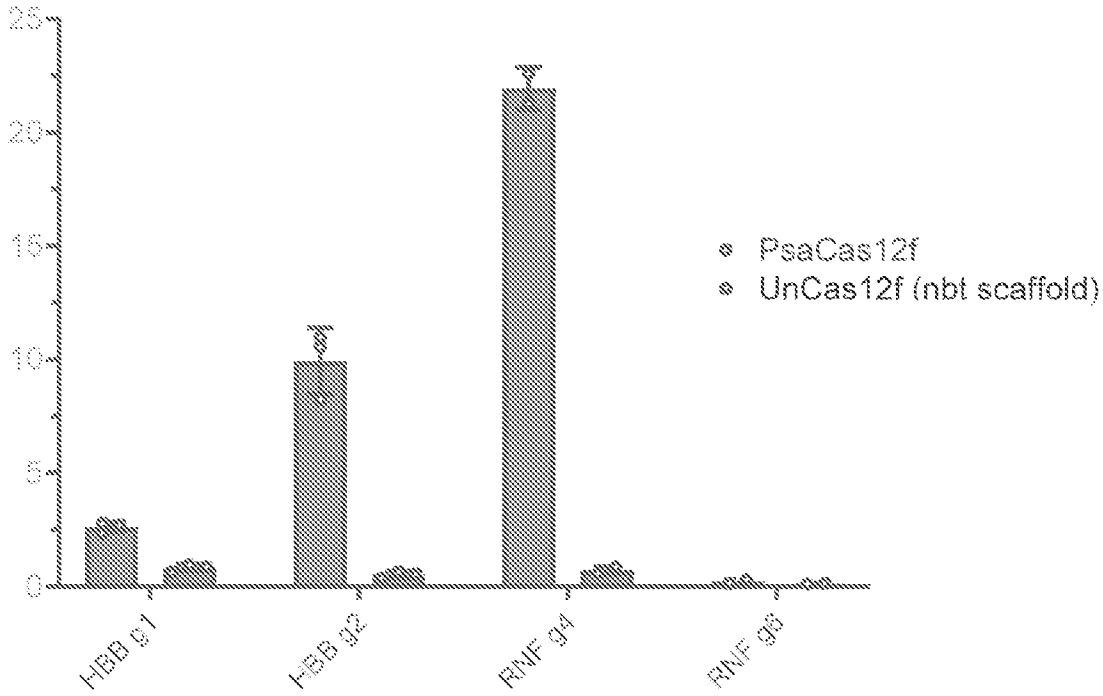
25
20
15
10
5
0
HBB g1
HBB g2
RNF g4
RNF g8
PsaCas12f
UnCas12f (nbt scaffold)

FIG. 13

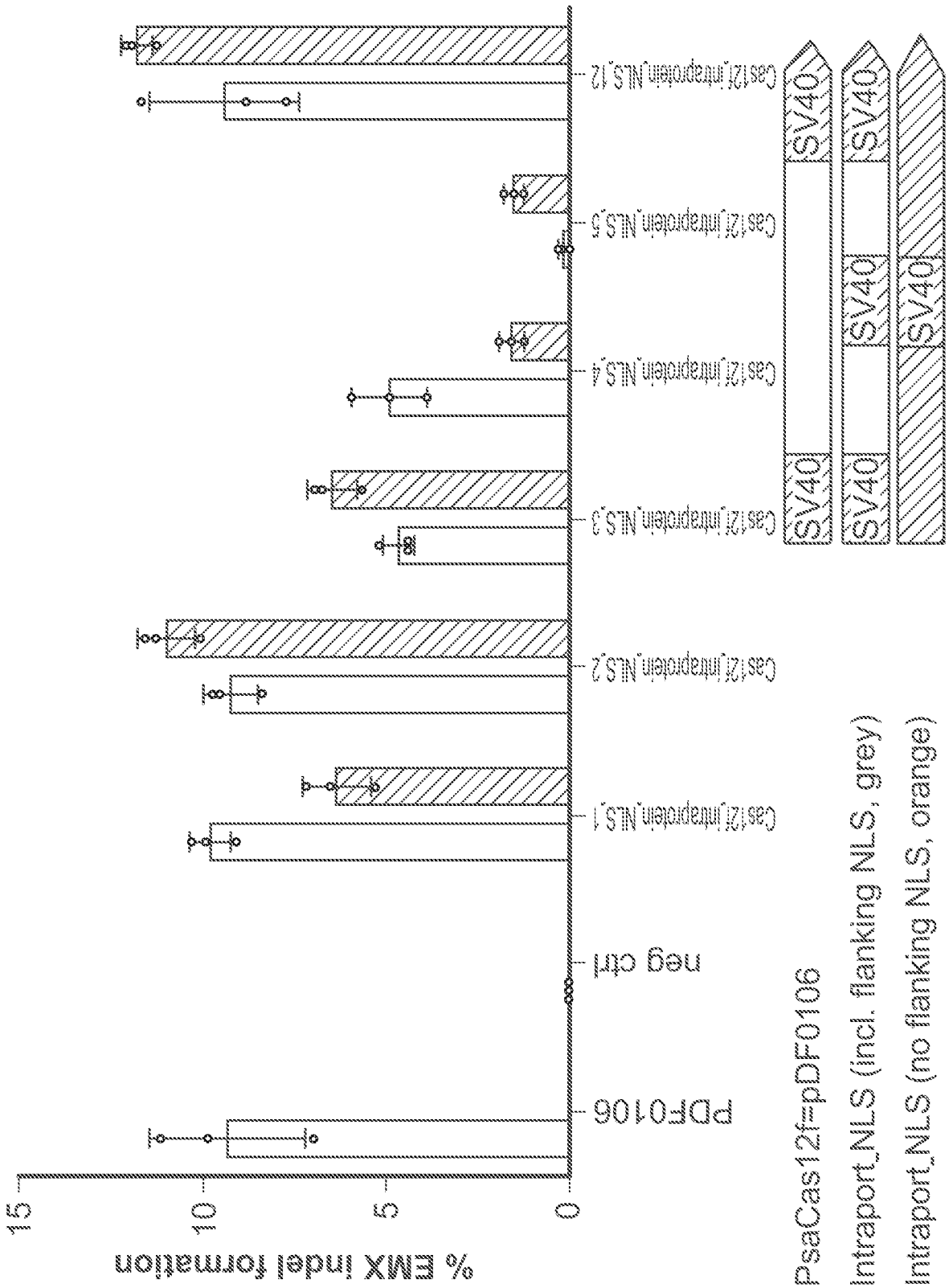
% EMX indel formation
15
10
5
0
PDF0106
neg ctrl
Cas12f_intraprotein_NLS.1
Cas12f_intraprotein_NLS.2
Cas12f_intraprotein_NLS.3
Cas12f_intraprotein_NLS.4
Cas12f_intraprotein_NLS.5
Cas12f_intraprotein_NLS.12
SV40
SV40
SV40
SV40
SV40
SV40
PsaCas12f=pDF0106
Intraport_NLS (incl. flanking NLS, grey)
Intraport_NLS (no flanking NLS, orange)

FIG. 14

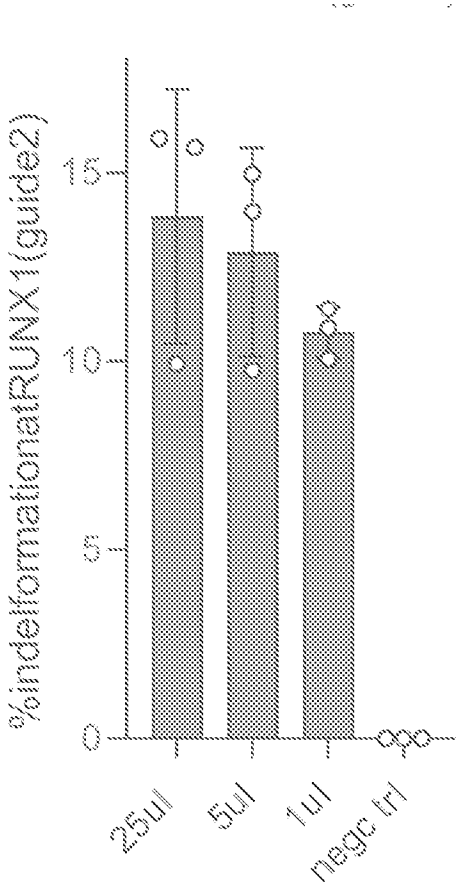
%indelformationatRUNX1(guide2)
0
5
10
15
25ul
5ul
1ul
negc trl

FIG. 15

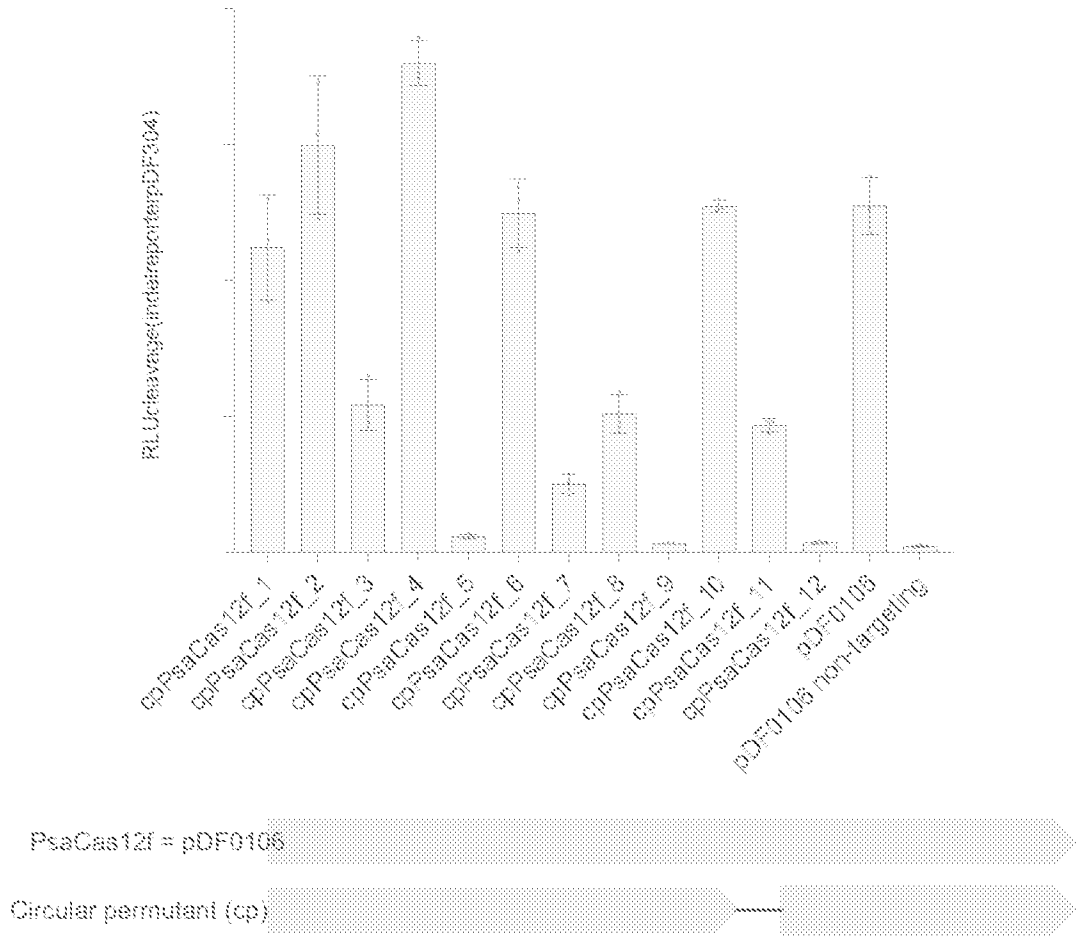
RL:cleavage(mdatreporter:pDF304)
cpPsaCas12f_1
cpPsaCas12f_2
cpPsaCas12f_3
cpPsaCas12f_4
cpPsaCas12f_5
cpPsaCas12f_6
cpPsaCas12f_7
cpPsaCas12f_8
cpPsaCas12f_9
cpPsaCas12f_10
cpPsaCas12f_11
cpPsaCas12f_12
pDF0108
pDF0108 non-targeting
PsaCas12f = pDF0108
Circular permutant (cp)

FIG. 16A

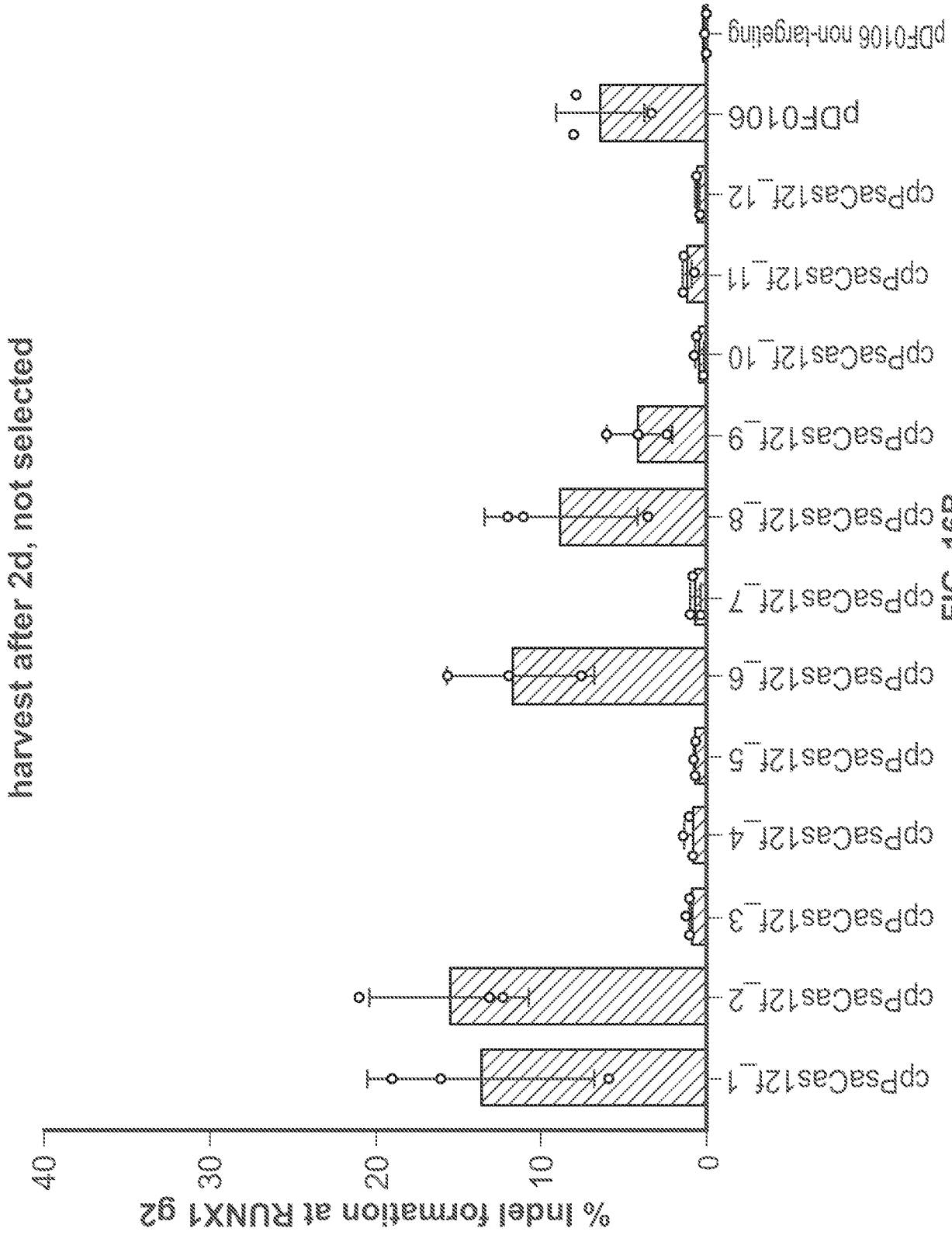
harvest after 2d, not selected
% Indel formation at RUNX1 g2
0
10
20
30
40
cpPsaCas12f_1
cpPsaCas12f_2
cpPsaCas12f_3
cpPsaCas12f_4
cpPsaCas12f_5
cpPsaCas12f_6
cpPsaCas12f_7
cpPsaCas12f_8
cpPsaCas12f_9
cpPsaCas12f_10
cpPsaCas12f_11
cpPsaCas12f_12
pDF0106
pDF0106 non-targeting

FIG. 16B

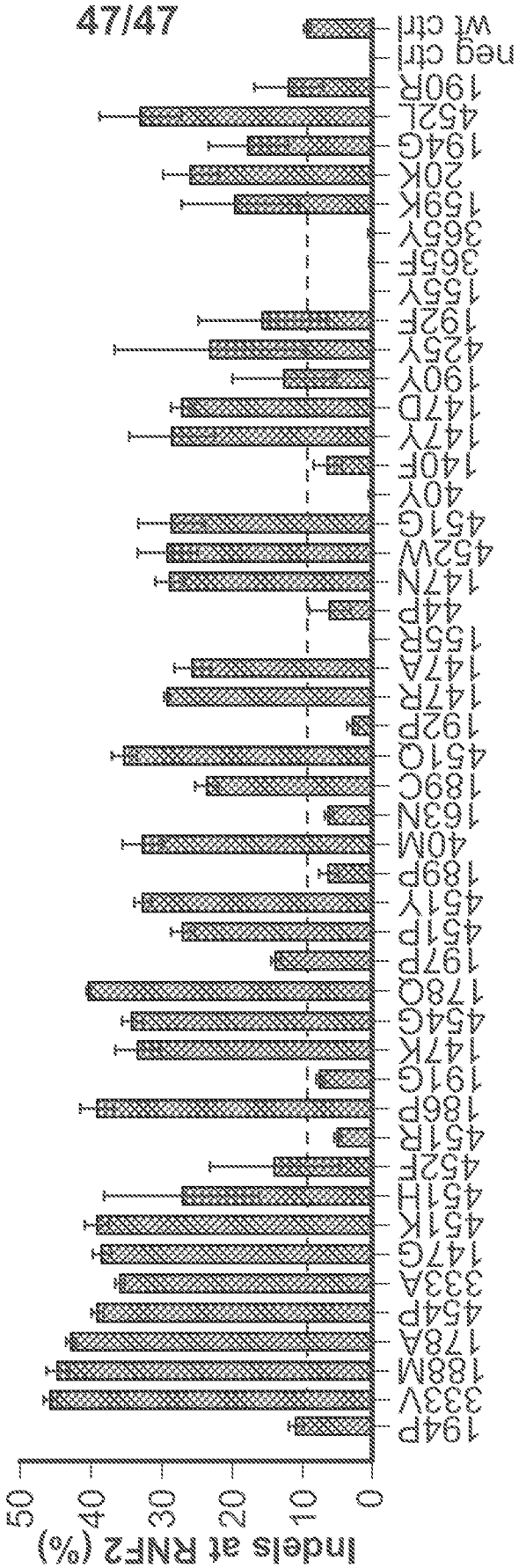
47/47
wt ctrl
neg ctrl
190R
452L
194G
20K
159K
365Y
365E
155Y
192F
425Y
190Y
147D
147Y
140F
40Y
451G
452W
147N
444P
155R
147A
147R
192P
451Q
189C
163N
40M
189P
451Y
451P
197P
178Q
454G
147K
191G
186P
451R
452F
451H
451K
147G
333A
454P
178A
188M
333V
194P
Indels at RNF2 (%)
0
10
20
30
40
50

FIG. 17

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/033749

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/22 C12N15/10
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, EMBASE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2021/139874 A1 (HOU ZHENGLIN [US] ET AL) 13 May 2021 (2021-05-13) the whole document, in particular Example 1 and SEQ ID NO 18	1-74
X	WO 2020/123887 A2 (PIONEER HI BRED INT [US]) 18 June 2020 (2020-06-18) the whole document, in particular Example 1 and SEQ ID NO: 18	1-74
X	WO 2020/214986 A1 (PIONEER HI BRED INT [US]) 22 October 2020 (2020-10-22) the whole document, in particular claim 63 and SEQ ID NO: 220	1-74
X	WO 2021/086083 A2 (GENKORE INC [KR]) 6 May 2021 (2021-05-06) the whole document	1-74

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 September 2022	08/11/2022

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bassias, Ioannis

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2022/033749

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2020/088450 A1 (UNIV CHINA AGRICULTURAL) 7 May 2020 (2020-05-07) the whole document	1-74
X,P	WO 2022/075813 A1 (GENKORE INC [KR]) 14 April 2022 (2022-04-14) the whole document	1-74
X,P	WO 2022/075808 A1 (GENKORE INC [KR]) 14 April 2022 (2022-04-14) the whole document	1-74
X,P	WO 2022/051250 A1 (UNIV LELAND STANFORD JUNIOR [US]) 10 March 2022 (2022-03-10) the whole document	1-74
X,P	XU XIAOSHU ET AL: "Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing", MOLECULAR CELL, ELSEVIER, AMSTERDAM, NL, vol. 81, no. 20, 3 September 2021 (2021-09-03), page 4333, XP086833228, ISSN: 1097-2765, DOI: 10.1016/J.MOLCEL.2021.08.008 [retrieved on 2021-09-03] the whole document	1-74

1

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2022/033749

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [ ] forming part of the international application as filed:

      [ ] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [X] furnished subsequent to the international filing date for the purposes of international search only:

      [X] in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

      [ ] on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2022/033749

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

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

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

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

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

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:
7-12, 38-43(completely); 1-6, 13-37, 44-74(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 7-12, 38-43(completely); 1-6, 13-37, 44-74(partially)

A composition comprising (a) a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid of SEQ ID NO: 1 and (b) a guide RNA (gRNA) wherein a target comprises a DNA target, a nucleic acid encoding said nuclease and said gRNA, vectors comprising said nucleic acids, a cell comprising said composition, a method of inserting or deleting one or more base pairs in a DNA with the use of said composition, a method of detecting a DNA target with the use of said composition, a method for activating or inhibiting the expression of a gene with the use of said composition, a method for nucleic acid base editing with the use of said composition and a method for activating or inhibiting the expression of a gene with epigenetic modifiers and the use of said composition.

---

2. claims: 1-6, 13-37, 44-74(all partially)

A composition comprising (a) a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid of SEQ ID NO: 2 and (b) a guide RNA (gRNA) wherein a target comprises a DNA target, a nucleic acid encoding said nuclease and said gRNA, vectors comprising said nucleic acids, a cell comprising said composition, a method of inserting or deleting one or more base pairs in a DNA with the use of said composition, a method of detecting a DNA target with the use of said composition, a method for activating or inhibiting the expression of a gene with the use of said composition, a method for nucleic acid base editing with the use of said composition and a method for activating or inhibiting the expression of a gene with epigenetic modifiers and the use of said composition.

3-19. claims: 1-6, 13-37, 44-74(all partially)

A composition comprising (a) a target specific nuclease comprising an amino acid sequence 70% identical to an amino acid selected from the group of SEQ ID NOs: 3-19 and (b) a guide RNA (gRNA) wherein a target comprises a DNA target, a nucleic acid encoding said nucleases and said gRNAs, vectors comprising said nucleic acids, a cell comprising said compositions, a method of inserting or deleting one or more base pairs in a DNA with the use of said compositions, a method of detecting a DNA target with the use of said compositions, a method for activating or inhibiting the expression of a gene with the use of said composition, a method for nucleic acid base editing with the use of said compositions and a method for activating or inhibiting the expression of a gene with epigenetic modifiers and the use

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

of said compositions.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2022/033749

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2021139874 A1	13-05-2021	AU 2019398351 A1	03-06-2021
		BR 112021011372 A2	31-08-2021
		CA 3117228 A1	18-06-2020
		CN 113166744 A	23-07-2021
		EP 3894550 A2	20-10-2021
		IL 283853 A	29-07-2021
		JP 2022514493 A	14-02-2022
		KR 20210104068 A	24-08-2021
		US 2020190494 A1	18-06-2020
		US 2021139874 A1	13-05-2021
		US 2021163908 A1	03-06-2021
		US 2022073890 A1	10-03-2022
		WO 2020123887 A2	18-06-2020
WO 2020123887 A2	18-06-2020	AU 2019398351 A1	03-06-2021
		BR 112021011372 A2	31-08-2021
		CA 3117228 A1	18-06-2020
		CN 113166744 A	23-07-2021
		EP 3894550 A2	20-10-2021
		IL 283853 A	29-07-2021
		JP 2022514493 A	14-02-2022
		KR 20210104068 A	24-08-2021
		US 2020190494 A1	18-06-2020
		US 2021139874 A1	13-05-2021
		US 2021163908 A1	03-06-2021
		US 2022073890 A1	10-03-2022
		WO 2020123887 A2	18-06-2020
WO 2020214986 A1	22-10-2020	AU 2020259484 A1	16-09-2021
		BR 112021020770 A2	04-01-2022
		CA 3131547 A1	22-10-2020
		CL 2021002714 A1	27-05-2022
		CN 113939594 A	14-01-2022
		EP 3955730 A1	23-02-2022
		US 2022240467 A1	04-08-2022
		WO 2020214986 A1	22-10-2020
WO 2021086083 A2	06-05-2021	CN 114846146 A	02-08-2022
		EP 4053285 A2	07-09-2022
		KR 20210053228 A	11-05-2021
		US 2022307018 A1	29-09-2022
		WO 2021086083 A2	06-05-2021
WO 2020088450 A1	07-05-2020	AU 2019372642 A1	17-06-2021
		CA 3118251 A1	07-05-2020
		CN 111757889 A	09-10-2020
		CN 113106081 A	13-07-2021
		CN 113136375 A	20-07-2021
		EP 3875469 A1	08-09-2021
		IL 282746 A	30-06-2021
		JP 2022512982 A	07-02-2022
		KR 20210129033 A	27-10-2021
		PH 12021550904 A1	29-11-2021
		SG 11202104347U A	28-05-2021
		US 2021395784 A1	23-12-2021
		WO 2020088450 A1	07-05-2020
WO 2022075813 A1	14-04-2022	NONE	

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2022/033749

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2022075808 A1	14-04-2022	NONE	
WO 2022051250 A1	10-03-2022	NONE	