



(51) International Patent Classification:

C12N 9/22 (2006.01)

C12N 15/90 (2006.01)

C12N 15/11 (2006.01)

C12Q 1/68 (2018.01)

(21) International Application Number:

PCT/US2020/028546

(22) International Filing Date:

16 April 2020 (16.04.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/834,582

16 April 2019 (16.04.2019)

US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 5th Floor, Oakland, California 94607 (US).

(72) Inventors: **YEO, Eugene**; c/o UC San Diego, 9500 Gilman Drive, Mail Code 0910, La Jolla, California 92093 (US).
TAN, Frederick; c/o UC San Diego, 9500 Gilman Drive, Mail Code 0910, La Jolla, California 92093 (US).

(74) Agent: **GREY, Kathryn** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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).

Declarations under Rule 4.17:

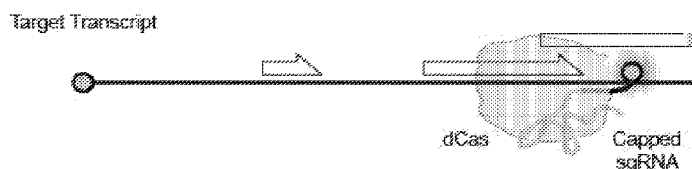
- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: PROTEIN TRANSLATIONAL CONTROL

Fig. 1E



(57) Abstract: Provided herein are compositions and methods for regulating protein translation. The compositions include a Cas polypeptide and a capped-sgRNA that includes (i) an m7G cap or an analog thereof; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide. The disclosure further provides methods of regulating translation of an mRNA in a cell, the method comprising contacting the cell with a nucleic acid comprising (a) a sequence encoding a Cas polypeptide; and (b) a sequence encoding a capped-sgRNA comprising (i) an m7G cap or analog thereof; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide.



PROTEIN TRANSLATIONAL CONTROL

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Patent Application Serial No. 62/834,582, filed April 16, 2019, which is incorporated herein by reference in its entirety.

5

STATEMENT OF GOVERNMENT SUPPORT

This invention was made with government support under EY029166 and NS103172, awarded by the National Institutes of Health. The government has certain rights in the invention.

10

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on April 16, 2020, is named Sequence_Listing.txt.

BACKGROUND

15

Existing RNA-targeting CRISPR-Cas applications provide tools to degrade RNA or modulate RNA structure but do not, on their own enhance gene expression or increase translational protein control. Likewise, existing methods of enhancing and/or increasing gene expression, such as the delivery of messenger RNAs, prove to be technically challenging. Indeed, there are few known or well characterized methods to increase mRNA translation.

20

The vast majority of gene regulatory drugs have been designed to knockdown gene expression (i.e. siRNAs, miRNAs, anti-sense, etc.). Some methods exist to enhance gene expression, such as the delivery of mRNAs; however, therapeutic delivery of such large and charged RNA molecules is technically challenging, inefficient, and not particularly practical. Classical gene therapy approaches involve delivery of a gene product as viral-encoded products (e.g. AAV or lentivirus-packaged products); however, these methods suffer from not being able to accurately reproduce the correct alternatively spliced isoforms in the correct ratios. In addition, gene delivery can exclude important non-coding regulatory sequences. Other methods of regulating protein translation involve engineered RNA binding proteins which are not easily delivered to cells and can be highly immunogenic. Often, engineered RNA binding proteins require extensive engineering for each target RNA sequence and certain of these possess limits to their applicability. More problematically, the act of

30

expression of certain of these types of engineered RNA binding proteins, when combined with translation initiation factor functions in a fusion protein context, could disrupt the stoichiometry of translation machinery maintained by the cell.

As such, there is a need to provide compositions and methods for recruiting
5 translational pre-initiation complexes in a manner which overcomes gene therapy barriers and protein engineering challenges, thereby controlling translation in cells and in gene therapy techniques.

SUMMARY

In one aspect, provided herein are complexes comprising: a Cas polypeptide; and a
10 capped-sgRNA comprising (i) an m7G cap or an analog thereof; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide. In some embodiments of any of the complexes described herein, the RNA molecule is a messenger RNA (mRNA). In some embodiments of any of the complexes described herein, the mRNA has an endogenous m7G cap. In some
15 embodiments of any of the complexes described herein, the target sequence is downstream of the endogenous m7G cap of the mRNA. In some embodiments of any of the complexes described herein, the mRNA comprises a start codon, and wherein the 5' end of the target sequence is upstream of the start codon. In some embodiments of any of the complexes described herein, the 5' end of the target sequence is between 1 to 50 nucleotides upstream of
20 the first nucleotide of the start codon. In some embodiments of any of the complexes described herein, the 5' end of the target sequence is between 1 to 15 nucleotides upstream of the first nucleotide of the start codon. In some embodiments of any of the complexes described herein, the target sequence comprises the start codon. In some embodiments of any of the complexes described herein, the mRNA comprises a start codon, and wherein the 5'
25 end of the target sequence is downstream of the start codon. In some embodiments of any of the complexes described herein, the 5' end of the target sequence is between 1 to 50 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of any of the complexes described herein, 5' end of the target sequence is between 1 to 15 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of
30 any of the complexes described herein, the 5' end of the target sequence is between 1 to 5 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of any of the complexes described herein, the spacer is at least 80% complementary to the target sequence. In some embodiments of any of the complexes described herein, the spacer is at

least 90% complementary to the target sequence. In some embodiments of any of the complexes described herein, the spacer comprises about 10 to about 40 nucleotides. In some embodiments of any of the complexes described herein, the spacer comprises about 15 to about 25 nucleotides. In some embodiments of any of the complexes described herein, the spacer comprises about 20 nucleotides. In some embodiments of any of the complexes described herein, the spacer is connected to the m7G cap or analog thereof via a linker. In some embodiments of any of the complexes described herein, the linker comprises about 5 to about 25 nucleotides. In some embodiments of any of the complexes described herein, the linker comprises about 8 to about 15 nucleotides. In some embodiments of any of the complexes described herein, the linker is not complementary to any sequence in the mRNA. In some embodiments of any of the complexes described herein, the linker is conjugated to the m7G cap or analog thereof via polyethylene glycol. In some embodiments of any of the complexes described herein, the Cas polypeptide is a nuclease-deficient Cas (dCas) polypeptide, wherein the dCas comprises an inactivated target cleavage domain and a retained guide cleavage domain. In some embodiments of any of the complexes described herein, the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d. In some embodiments of any of the complexes described herein, the direct repeat is capable of binding to a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d. In some embodiments of any of the complexes described herein, the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas9 (dCas9) polypeptide. In some embodiments of any of the complexes described herein, the direct repeat is capable of binding to a nuclease-deficient Cas9 (dCas9) polypeptide. In some aspects, provided herein is a nucleic acid comprising a sequence encoding the capped-sgRNA in any of the complexes described herein. In some embodiments, the nucleic acid further comprises a sequence encoding the Cas polypeptide in any of the complexes described herein.

In another aspect, provided herein are nucleic acids comprising a sequence encoding a capped-sgRNA, wherein the capped-sgRNA comprises: (i) an m7G cap or analog thereof; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to a Cas polypeptide. In some embodiments of any of the nucleic acids described herein, the RNA molecule is an mRNA. In some embodiments of any of the nucleic acids described herein, the mRNA has an endogenous m7G cap. In some embodiments of any of the nucleic acids described herein, the target sequence is downstream of the endogenous m7G cap of the mRNA. In some embodiments of any of the nucleic acids

described herein, the mRNA comprises a start codon, and wherein the 5' end of the target sequence is upstream of the start codon. In some embodiments of any of the nucleic acids described herein, the 5' end of the target sequence is between 1 and 50 nucleotides upstream of the first nucleotide of the start codon. In some embodiments of any of the nucleic acids described herein, the 5' end of the target sequence is between 1 to 15 nucleotides upstream of the first nucleotide of the start codon. In some embodiments of any of the nucleic acids described herein, the target sequence comprises the start codon. In some embodiments of any of the nucleic acids described herein, the mRNA comprises a start codon, and wherein the 5' end of the target sequence is downstream of the start codon. In some embodiments of any of the nucleic acids described herein, the 5' end of the target sequence is between 1 to 50 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of any of the nucleic acids described herein, the 5' end of the target sequence is between 1 to 15 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of any of the nucleic acids described herein, the 5' end of the target sequence is between 1 to 5 nucleotides downstream of the last nucleotide of the start codon. In some embodiments of any of the nucleic acids described herein, the spacer is at least 80% complementary to the target sequence. In some embodiments of any of the nucleic acids described herein, the spacer is at least 90% complementary to the target sequence. In some embodiments of any of the nucleic acids described herein, the spacer comprises about 10 to about 40 nucleotides. In some embodiments of any of the nucleic acids described herein, the spacer comprises about 15 to about 25 nucleotides. In some embodiments of any of the nucleic acids described herein, the spacer comprises about 20 nucleotides. In some embodiments of any of the nucleic acids described herein, the spacer is connected to the m7G cap or analog thereof via a linker. In some embodiments of any of the nucleic acids described herein, the linker comprises about 5 to about 25 nucleotides. In some embodiments of any of the nucleic acids described herein, the linker comprises about 8 to about 20 nucleotides. In some embodiments of any of the nucleic acids described herein, the linker is not complementary to any sequence in the mRNA. In some embodiments of any of the nucleic acids described herein, the linker is conjugated to the m7G cap or analog thereof via polyethylene glycol. In some embodiments of any of the nucleic acids described herein, the nucleic acids further comprise a sequence encoding a RNase P processing site. In some embodiments of any of the nucleic acids described herein, the nucleic acids further comprise a poly-T sequence. In some embodiments of any of the nucleic acids described herein, the nucleic acids further comprise a poly-T sequence downstream of the sequence encoding a RNase P processing site. In some

embodiments of any of the nucleic acids described herein, the nucleic acids further comprise a sequence encoding the Cas polypeptide. In some embodiments of any of the nucleic acids described herein, the Cas polypeptide is a nuclease-deficient Cas polypeptide. In some
5 polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d. In some embodiments of any of the nucleic acids described herein, the direct repeat is capable of binding to a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d. In some embodiments of any of the nucleic acids described herein, the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas9 (dCas9)
10 polypeptide. In some embodiments of any of the nucleic acids described herein, the nucleic acids further comprise one or more RNA polymerase II promoters. In some embodiments of any of the nucleic acids described herein, the sequence encoding the capped-sgRNA and the sequence encoding the Cas polypeptide are expressed from the same promoter. In some
embodiments of any of the nucleic acids described herein, the sequence encoding the capped-
15 sgRNA and the sequence encoding the Cas polypeptide are expressed from different promoters. In some aspects, provided herein are vectors comprising the nucleic acids of any of the above embodiments. In some embodiments, the vector is an AAV vector. In some
embodiments, provided herein are cells comprising the nucleic acid of any of the above
embodiments.

20 In another aspect, provided herein are methods of regulating translation of an mRNA in a cell, the method comprising contacting the cell with a nucleic acid comprising (a) a sequence encoding a Cas polypeptide; and (b) a sequence encoding a capped-sgRNA comprising (i) an m7G cap or analog thereof; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the
25 Cas polypeptide. In some embodiments of any of the methods of regulating translation of an mRNA in a cell described herein, the Cas polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide. In some embodiments of any of the methods of regulating translation of an mRNA in a cell described herein, the dCas13 polypeptide is dCas13b or dCas13d. In some
embodiments of any of the methods of regulating translation of an mRNA in a cell described
30 herein, the dCas13b comprises an inactivated target cleavage domain and a retained guide cleavage domain. In some embodiments of any of the methods of regulating translation of an mRNA in a cell described herein, the dCas13d comprises an inactivated target cleavage domain and a retained guide cleavage domain.

INCORPORATION BY REFERENCE

All publications, patents, patent applications, and information available on the internet and mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, patent application, or item of information was specifically and individually indicated to be incorporated by reference. To the extent publications, patents, patent applications, and items of information incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

DESCRIPTION OF THE DRAWINGS

A better understanding of the features and advantages can be obtained by reference to the following detailed description that sets forth illustrative embodiments and accompanying drawings (“Figure” and “Fig.” herein), of which:

FIGS. 1A-1G show modulation of translation using dCas and capped-sgRNA. FIG. 1A shows exemplary constructs for generating dCas and Capped-sgRNA. FIG. 1B shows an exemplary structure of unprocessed capped-sgRNA containing an RNase P processing site. FIG. 1C shows an exemplary chemical composition of a capped-sgRNA. FIG. 1D shows capped-sgRNA processing by RNase P. FIG. 1E shows a schematic of localized capped-sgRNA recruitment and binding of dCas to the capped-sgRNA. FIG. 1F shows an exemplary two construct system for expressing dCas and capped-sgRNA. FIG. 1G shows the sgRNA targeting window in the target ATF4 ORF. FIG. 1H shows results of capped-sgRNA modulation on the translation of the target ATF4 ORF. FIG. 1I shows results of using the control uncapped-sgRNAs on the translation of the target ATF4 ORF.

DETAILED DESCRIPTION

Embodiments according to the invention disclosed herein will be described more fully hereinafter. However, it should be understood that such embodiments are provided merely by way of example, and numerous variations, changes, and substitutions can occur to those skilled in the art without departing from the scope of this disclosure. It should also be understood that various alternatives to the specific embodiments described herein are also within the scope of this disclosure.

Where values are described in terms of ranges, it should be understood that the description includes the disclosure of all possible sub-ranges within such ranges, as well as

specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated. All numerical designations, e.g., pH, temperature, time, concentration, and molecular weight, including ranges, are approximations which are varied (+) or (-) by increments of 1.0 or 0.1, as appropriate, or alternatively by a variation of +/- 15 %, or alternatively 10%, or alternatively 5%, or alternatively 2%. It is to be understood, although not always explicitly stated, that all numerical designations are preceded by the term “about”. It also is to be understood, although not always explicitly stated, that the reagents described herein are merely exemplary and that equivalents of such are known in the art.

Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the present application and relevant art and should not be interpreted in an idealized or overly formal sense unless expressly so defined herein. While not explicitly defined below, such terms should be interpreted according to their common meaning.

Unless the context indicates otherwise, it is specifically intended that the various features of the invention described herein can be used in any combination. Moreover, the disclosure also contemplates that in some embodiments, any feature or combination of features set forth herein can be excluded or omitted. To illustrate, if the specification states that a complex comprises components A, B and C, it is specifically intended that any of A, B or C, or a combination thereof, can be omitted and disclaimed singularly or in any combination. Unless explicitly indicated otherwise, all specified embodiments, features, and terms intend to include both the recited embodiment, feature, or term and biological equivalents thereof.

As used in the description of the invention and the appended claims, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise.

The term “about,” as used herein when referring to a measurable value such as an amount or concentration and the like, is meant to encompass variations of 20%, 10%, 5%, 1 %, 0.5%, or even 0.1 % of the specified amount.

The terms “acceptable,” “effective,” “efficient” or “sufficient” when used to describe the selection of any components, ranges, dose forms, etc. disclosed herein intend that said component, range, dose form, etc. is suitable for the disclosed purpose.

Also as used herein, “and/or” refers to and encompasses any and all possible
5 combinations of one or more of the associated listed items, as well as the lack of combinations when interpreted in the alternative (“or”).

As used herein, the term “comprising” is intended to mean that the compositions and methods include the recited elements, but do not exclude others. As used herein, the transitional phrase “consisting essentially of” (and grammatical variants) is to be interpreted
10 as encompassing the recited materials or steps and those that do not materially affect the basic and novel characteristic(s) of the recited embodiment. Thus, the term “consisting essentially of” as used herein should not be interpreted as equivalent to “comprising.” “Consisting of” shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions disclosed herein. Aspects
15 defined by each of these transition terms are within the scope of the present disclosure.

As used herein, the term “functional” may be used to modify any molecule, biological, or cellular material to intend that it accomplishes a particular, specified effect.

TRANSLATIONAL ENHANCEMENT VIA M7G CAP RECRUITMENT

20 Translation initiation in mammalian cells starts with the binding of the 5' methyl-7 guanosine (m7G) cap structure by Eukaryotic Initiation Factor 4E (EIF4E), which results in the nucleation of translational pre-initiation complexes on the adjacent 5' untranslated region (5'UTR) of mRNA. The bound pre-initiation complexes then scan the 5'UTR unidirectionally (5' to 3') for suitable start codons (e.g., “AUG”) to prime and initiate
25 translation. The 5' m7G cap is an evolutionarily conserved modification of eukaryotic mRNA, and serves as a unique molecular module that recruits cellular proteins and mediates cap-related biological functions such as pre-mRNA processing, nuclear export, and cap-dependent protein synthesis.

Provided herein are compositions and methods for enhancing protein production by
30 recruiting an m7G cap to an mRNA using a capped-sgRNA and a Cas polypeptide. In the compositions and methods provided herein, the bound pre-initiation complexes do not necessarily scan the 5'UTR unidirectionally 5' to 3'.

In some aspects, provided herein is a complex comprising at least one Cas polypeptide or nucleic acid encoding the at least one Cas polypeptide, and at least one m7G

capped single guide RNA (capped-sgRNA) or nucleic acid encoding the at least one capped-sgRNA, where the at least one capped-sgRNA is capable of targeting the at least one Cas polypeptide to a target sequence in an RNA molecule (e.g., an mRNA). In some embodiments, upon hybridization between the sgRNA and the target sequence, the m7G cap of the sgRNA is brought closer to a desired start codon in the target mRNA as compared to the endogenous m7G cap of the target mRNA. Also provided are methods of regulating translation of an mRNA in a cell, the method comprising contacting the cell with a nucleic acid comprising (a) a sequence encoding a at least one Cas polypeptide; and (b) a sequence encoding at least one capped-sgRNA comprising (i) an m7G cap; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide.

In some aspects, provided herein is a complex comprising a Cas polypeptide or nucleic acid encoding the Cas polypeptide, and an m7G capped single guide RNA (capped-sgRNA) or nucleic acid encoding the capped-sgRNA, where the capped-sgRNA is capable of targeting the Cas polypeptide to a target sequence in an RNA molecule (e.g., an mRNA). In some embodiments, upon hybridization between the sgRNA and the target sequence, the m7G cap of the sgRNA is brought closer to a desired start codon in the target mRNA as compared to the endogenous m7G cap of the target mRNA. Also provided are methods of regulating translation of an mRNA in a cell, the method comprising contacting the cell with a nucleic acid comprising (a) a sequence encoding a Cas polypeptide; and (b) a sequence encoding a capped-sgRNA comprising (i) an m7G cap; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide.

Each strand of DNA or RNA has a 5' end and a 3' end, corresponding to the carbon position on the deoxyribose (or ribose) ring. "Upstream" as described herein can mean toward the 5' end of an RNA molecule and "downstream" as described herein can mean towards the 3' end of an RNA molecule. A "start codon" as described herein can refer to the first codon of a messenger RNA transcript translated by a ribosome. The most common start codon is AUG. Alternative start codons from both prokaryotes and eukaryotes include, but not limited to, GUG, UUG, AUU, and CUG.

The term "cell" as used herein may refer to either a prokaryotic or eukaryotic cell, optionally obtained from a subject or a commercially available source.

The term "encode" as it is applied to nucleic acid sequences refers to a polynucleotide which is said to "encode" a polypeptide, an mRNA, or an effector RNA if, in its native state

or when manipulated by methods well known to those skilled in the art, can be transcribed and/or translated to produce the effector RNA, the mRNA, or an mRNA that can for the polypeptide and/or a fragment thereof. The antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

5 As used herein, the term “expression” or “gene expression” refers to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The expression level of a gene may be determined by measuring the
10 amount of mRNA or protein in a cell or tissue sample; further, the expression level of multiple genes can be determined to establish an expression profile for a particular sample.

The term “target sequence” can refer to a nucleic acid sequence present in an RNA molecule to which a spacer of a guide RNA (e.g., a capped-sgRNA as disclosed herein) can hybridize, provided sufficient conditions for hybridization exist. Hybridization between the
15 spacer and the target sequence can, for example, be based on Watson-Crick base pairing rules, which enables programmability in the spacer sequence. The spacer sequence can be designed, for instance, to hybridize with any target sequence.

The “spacer” or “spacer sequence” is comprised within a single guide RNA can include a nucleotide sequence that is complementary to a specific sequence within a target
20 RNA.

“Binding” as used herein can refer to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be “associated” or “interacting” or “binding” (e.g., when a molecule X is said to interact with a molecule Y, it means that the molecule X
25 binds to molecule Y in a non-covalent manner). Binding interactions are generally characterized by a dissociation constant (K_d) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. K_d is dependent on environmental conditions, e.g., pH and temperature, as is known by those in the art. “Affinity” can refer to the strength of
30 binding, and increased binding affinity is correlated with a lower K_d .

The terms “hybridizing” or “hybridize” can refer to the pairing of substantially complementary or complementary nucleic acid sequences within two different molecules. Pairing can be achieved by any process in which a nucleic acid sequence joins with a partially, substantially or fully complementary sequence through base pairing to form a

hybridization complex. For purposes of hybridization, two nucleic acid sequences or segments of sequences are “substantially complementary” if at least 80% of their individual bases are complementary to one another. Two nucleic acid sequences or segments of sequences are “partially complementary” if at least 50% of their individual bases are complementary to one another.

As used herein, “complementary” can mean that two nucleic acid sequences have at least 50% sequence identity. Preferably, the two nucleic acid sequences have at least 60%, 70%, 80%, 90%, 95%, 96%, 97%, 98%, 99%, or 100% of sequence identity.

“Complementary” also means that two nucleic acid sequences can hybridize under low, middle, and/or high stringency condition(s).

As used herein, “substantially complementary” means that two nucleic acid sequences have at least 90% sequence identity. Preferably, the two nucleic acid sequences have at least 95%, 96%, 97%, 98%, 99%, or 100% of sequence identity. “Substantially complementary” can also mean that two nucleic acid sequences can hybridize under high stringency condition(s).

Low stringency hybridization refers to conditions equivalent to hybridization in 10% formamide, 5x Denhardt’s solution, 6x SSPE, 0.2% SDS at 22°C, followed by washing in 1x SSPE, 0.2% SDS, at 37°C. Denhardt’s solution contains 1% Ficoll, 1% polyvinylpyrrolidone, and 1% bovine serum albumin (BSA). 20x SSPE (sodium chloride, sodium phosphate, ethylene diamide tetraacetic acid (EDTA)) contains 3M sodium chloride, 0.2M sodium phosphate, and 0.025 M (EDTA). Other suitable moderate stringency and high stringency hybridization buffers and conditions are well known to those of skill in the art.

As used herein, “operably linked” refers to the situation in which part of a linear DNA sequence can influence the other parts of the same DNA molecule. For example, when a promoter controls the transcription of the coding sequence, it is operatively linked to the coding sequence.

As used herein, a “polypeptide” refers to, without limitation, proteins, fragments of proteins, and peptides, whether isolated from natural sources, produced by recombinant techniques, or chemically synthesized. A polypeptide may have one or more modifications, such as a post-translational modification (such as glycosylation, etc.) or any other modification (such as PEGylation, etc.). The polypeptide may contain one or more non-naturally-occurring amino acids (such as an amino acid with a side chain modification). Polypeptides described herein typically comprise at least about 10 amino acids.

As used herein, “contacting” a cell with a nucleic acid molecule can include allowing the nucleic acid molecule to be in sufficient proximity with the cell such that the nucleic acid molecule can be introduced into the cell.

5 A “promoter” can be a region of DNA that leads to initiation of transcription of a gene.

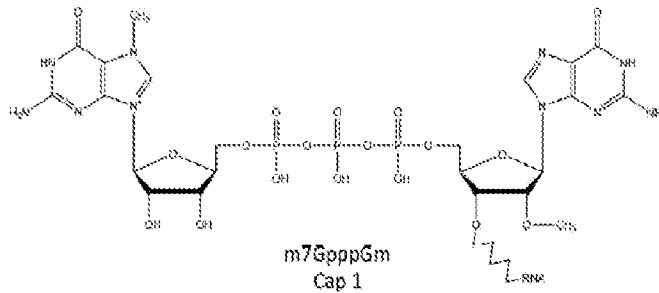
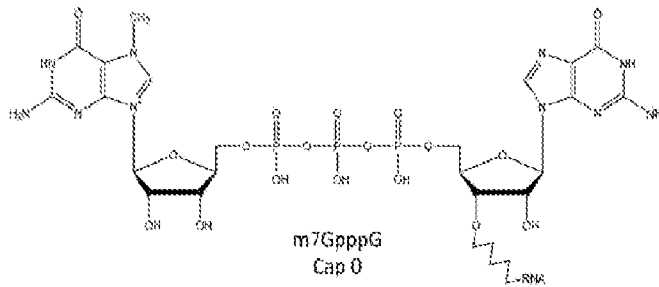
As used herein, “nuclease-deficient” may refer to a polypeptide with reduced nuclease activity, reduced endo- or exo- DNase activity or RNase activity, reduced nickase activity, or reduced ability to cleave DNA and/or RNA. “Reduced nuclease activity” means a decline in nuclease, nickase, DNase, or RNase activity of at least about 1%, 2%, 5%, 10%, 15%,
10 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100% or more or any value between any of the listed values. Alternatively, “reduced nuclease activity” may refer to a decline of at least about 1-fold, 1.5-fold, 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 15-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, 500-fold, 1000-fold, 2000-fold or more or any value between
15 any of the listed values.

“Nucleic acids” may be naturally occurring nucleic acids such as DNA and RNA, or artificial nucleic acids including peptide nucleic acid (PNA), morpholino, locked nucleic acid (LNA), glycol nucleic acid (GNA), and threose nucleic acid (TNA). Nucleic acids disclosed herein can be single-stranded or double-stranded nucleic acids.

20

I. m7G cap

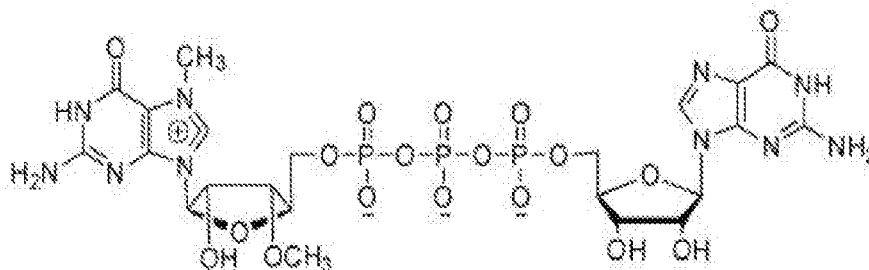
The m7G cap, or 7-methylguanosine cap, is a guanine nucleotide methylated on the 7 position and can be linked to an RNA molecule (e.g., a sgRNA or an mRNA) via a 5' to 5' triphosphate linkage. In one embodiment, capped RNA molecules (e.g., capped-sgRNAs)
25 disclosed herein include a single methyl group on the terminal G residue at the N-7 position. *In vivo*, the m7G cap structure is added enzymatically to mRNA produced by RNA polymerase II. In one embodiment, the adjacent nucleotides can be 2'-O-methylated to different extents. For example, the m7G cap can be a m7GpppN, or Cap 0, an m7GpppNm, or Cap 1, and an m7GpppNmpNm, or Cap 2. Exemplary structures of Cap 0 and Cap 1, are
30 shown below:



The 5' m7G cap is an evolutionarily conserved modification of eukaryotic mRNA, and serves as a unique molecular module that recruits cellular proteins and mediates cap-related biological functions such as pre-mRNA processing, nuclear export, and cap-dependent protein synthesis. The capped-sgRNA disclosed herein exploits these biological functions to recruit pre-initiation complexes for enhancement of protein translation.

The capped-sgRNA disclosed herein includes a nucleic acid sequence which is transcribed by, e.g. an RNA polymerase II, and the sgRNA thereby becomes 5' capped upon transcription with an m7G cap. The transcription of capped-sgRNA nucleic acid sequence is catalyzed by bacteriophage RNA polymerases, such as, without limitation, RNA polymerase T7, T3, and SP6. When bacteriophage RNA polymerases are used to generate the sgRNAs of the present disclosure, cap analogs initiate transcription. In one embodiment, the m7G(5')pppG cap analog is incorporated into the sgRNA and simulates the m7G cap structure. In another embodiment, standard cap analogs are incorporated into the sgRNA in the forward (e.g., [m7G(5')pppG(pN)]) or the reverse orientation (e.g., [G(5')pppm7G(pN)]) resulting in two forms of isomeric RNAs. In another embodiment, chemical modifications at either the 2' or 3' OH group results in the cap being incorporated solely in the forward orientation. In some embodiments, the m7G cap is an anti-reverse cap analog (ARCA), wherein one of the 3' OH groups (closer m7G) is eliminated from the cap analog and is substituted with -OCH₃. This modification forces RNA polymerases to initiate transcription with the remaining -OH group in G and thus synthesize RNA transcripts capped exclusively

in the correct orientation. An exemplary structure of ARCA (m7(3'-O-methyl)-G(5')ppp(5')G) is shown below:



Additional cap analogs contemplated herein also include unmethylated cap analogs (e.g., GpppG), trimethylated cap analogs (e.g., m₃^{2,2,7}GP₃G), and m₂^{7,3'-O}GP₃(2'OMe)ApG.

In one embodiment, the m7G cap disclosed herein includes chemical modifications relative to the naturally occurring m7G cap. For example, chemical modifications that can reduce the sensitivity of the m7G cap to cellular decapping enzymes are useful for the capped-RNAs disclosed herein. Suitable chemical modifications include, without limitation, those with 1,2-dithiodiphosphate. See those described in e.g., Strenkowska et al., *Nucleic Acids Res.* 44(20):9578-9590 (2016), phosphate-modified cap analogues described in e.g., Walczak et al., *Chem Sci.* 8(1):260-267 (2017)), as well as those described in Basolo et al., *Eur J Endocrinol.*, 145(5):599-604 (2001), and Borghardt et al., *Can Respir J.* 2018 Jun 19; 2018:2732017, all of which are incorporated herein by reference in their entirety.

II. Capped-sgRNA

Provided herein are capped-sgRNAs, nucleic acids comprising and/or encoding the capped-sgRNAs, and methods of using the same for regulating protein translation. The capped-sgRNA can include an m7G cap or an analog thereof, a spacer capable of specifically hybridizing with a target sequence in an RNA molecule, and a direct repeat capable of binding to a Cas polypeptide. In some embodiments, the capped-sgRNA includes from 5' to 3', an m7G cap or an analog thereof, a spacer sequence, and a direct repeat sequence. In one embodiment, the 5' cap is linked to the spacer sequence via a linker. In one embodiment, the capped-sgRNA is derived from an unprocessed capped-sgRNA that further includes a Ribonuclease P (RNase P) processing site, and a polyadenylated (poly-A) tail at the 3' end.

In some embodiments of the capped-sgRNAs disclosed herein, the spacer sequence and the scaffolding sequence or direct repeat sequence are not contiguous. In some embodiments, a scaffold sequence comprises a direct repeat sequence.

In some embodiments, the capped-sgRNA sequence is synthetic or comprises non-naturally occurring nucleotides. In some embodiments, a capped-guideRNA of the disclosure or a sequence encoding the guide RNA comprises or consists of modified or synthetic RNA nucleotides. Exemplary modified RNA nucleotides include, but are not limited to, pseudouridine (Y), dihydrouridine (D), inosine (I), and 7-methylguanosine (m7G), hypoxanthine, xanthine, xanthosine, 7-methylguanine, 5, 6-Dihydrouracil, 5-methylcytosine, 5-methylcytidine, 5-hydroxymethylcytosine, isoguanine, and isocytosine. Capped-sgRNAs of the disclosure may bind modified RNA within a target sequence. Within a target sequence, capped-guide RNAs of the disclosure may bind modified RNA. Exemplary epigenetically or post-transcriptionally modified RNA include, but are not limited to, 2'-O-Methylation (2'-OMe) (2'-O-methylation occurs on the oxygen of the free T - OH of the ribose moiety), N6-methyladenosine (m6A), and 5-methylcytosine (m5C). In some embodiments of the compositions of the disclosure, a capped-guide RNA of the disclosure comprises at least one sequence encoding a non-coding C/D box small nucleolar RNA (snoRNA) sequence. In some embodiments, the snoRNA sequence comprises at least one sequence that is complementary to the target RNA, wherein the target sequence of the RNA molecule comprises at least one 2'-OMe. In some embodiments, the snoRNA sequence comprises at least one sequence that is complementary to the target RNA, wherein the at least one sequence that is complementary to the target RNA comprises a box C motif (RETGAETGA) and a box D motif (CUGA).

In some embodiments, a sequence encoding a capped-guide RNA of the disclosure comprises or consists essentially of a spacer sequence and a scaffold or direct repeat sequence, wherein the spacer and the scaffold or direct repeat are operably linked. In one embodiment, the spacer and scaffold and/or direct repeat are separated by a linker sequence. In some embodiments, the linker sequence may include or consist of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, nucleotides or any number of nucleotides in between. In some embodiments, the linker sequence may include at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, nucleotides or any number of nucleotides in between.

In some embodiments, therapeutic or pharmaceutical compositions including the capped-sgRNAs or methods of gene therapy using the capped-sgRNAs of the disclosure do not include a PAMmer oligonucleotide. In other embodiments, optionally, non-therapeutic or non-pharmaceutical compositions may include a PAMmer oligonucleotide. In some embodiments of the compositions of the disclosure, a guide RNA or a portion thereof includes a sequence complementary to a protospacer flanking sequence (PFS). In some embodiments, including those wherein a guide RNA or a portion thereof includes a sequence

complementary to a PFS, the RNA binding protein may include a sequence isolated or derived from a Cas protein, such as, without limitation, a Cas9, Cas13b, or Cas13d protein. In some embodiments, including those wherein a guide RNA or a portion thereof includes a sequence complementary to a PFS, the RNA binding protein may include a sequence
 5 encoding a Cas protein, such as, without limitation, a Cas9, Cas 13b, or Cas13d protein, or an RNA-binding portion thereof. In some embodiments, the guide RNA or a portion thereof does not include a sequence complementary to a PFS.

Spacer

10 The capped-sgRNA disclosed herein comprises a “spacer” or “spacer sequence” that is complementary to a specific sequence within a target RNA. The spacer sequence can be designed to hybridize with any target sequence of interest.

In one embodiment, the spacer sequence comprised within the capped-sgRNA is about 10 to about 30 nucleotides (e.g., about 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23,
 15 24, 25, 26, 27, 28, or 29 nucleotides). In another embodiment, the spacer sequence is about 15 to about 25 nucleotides (e.g., about 18 to about 22 nucleotides, or about 20 nucleotides). In another embodiment, the spacer sequence is at least 50% complementary, at least 60% complementary, or at least 70% complementary to a target sequence in an RNA molecule (e.g., an mRNA). In another embodiment, the spacer sequence is at least 80% (e.g., at least
 20 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) complementary to a target sequence in an RNA molecule (e.g., an mRNA). In another embodiment, the spacer sequence is 100% complementary to the target sequence. Exemplary spacer sequences for the capped-sgRNAs disclosed herein are:

TTTGCTGGAATCGAGGAATGTGCTT (SEQ ID NO: 278)
 25 GTTGCGGTGCTTTGCTGGAATCGAG (SEQ ID NO: 279)
 TTTCGGTCATGTTGCGGTGCTTTGC (SEQ ID NO: 280)
 AGGAAGCTCATTTTCGGTCATGTTGC (SEQ ID NO: 281)
 CTCGCTGCTCAGGAAGCTCATTTTCG (SEQ ID NO: 282)
 CCACCAACACCTCGCTGCTCAGGAA (SEQ ID NO: 283)

30 In another embodiment of the capped-sgRNAs disclosed herein, spacer sequences may comprise a CRISPR RNA (crRNA). In another embodiment, spacer sequences of the disclosure comprise or consist of a sequence having sufficient complementarity to a target sequence of an RNA molecule to bind selectively to the target sequence. Upon binding to a target sequence of an RNA molecule, the spacer sequence may guide one or more of a

scaffold or direct repeat sequence and a Cas polypeptide or fusion protein to the RNA molecule. In some embodiments, a spacer sequence having sufficient complementarity to a target sequence of an RNA molecule to bind selectively (partially or substantially) to the target sequence has at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96,
5 97%, 98%, 99%, or any percentage identity in between to the target sequence. In some embodiments, a spacer sequence having sufficient complementarity to a target sequence of an RNA molecule to bind selectively to the target sequence has 100% identity the target sequence.

In some embodiments of the compositions of the disclosure, a capped-guide RNA or a
10 portion thereof comprises or consists of between 10 and 100 nucleotides, inclusive of the endpoints. In some embodiments, a spacer sequence of the disclosure comprises or consists of between 10 and 30 nucleotides, inclusive of the endpoints. In some embodiments, a spacer sequence of the disclosure comprises or consists of 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,
15 26, 27, 28, 29, or 30 nucleotides. In some embodiments, the spacer sequence of the disclosure comprises or consists of 20 nucleotides. In some embodiments, the spacer sequence of the disclosure comprises or consists of 21 nucleotides. In some embodiments, a scaffold or direct repeat sequence of the disclosure comprises or consists of between 10 and 100 nucleotides, inclusive of the endpoints. In some embodiments, a scaffold or direct repeat sequence of the disclosure comprises or consists of 30, 35, 40, 45, 50, 55, 60, 65, 70, 76, 80, 87, 90, 95, 100,
20 or any number of nucleotides in between. In some embodiments, the scaffold or direct repeat sequence of the disclosure comprises or consists of between 85 and 95 nucleotides, inclusive of the endpoints. In some embodiments, the scaffold or direct repeat sequence of the disclosure comprises or consists of 85 nucleotides. In some embodiments, the scaffold or direct repeat sequence of the disclosure comprises or consists of 90 nucleotides. In some
25 embodiments, the scaffold or direct repeat sequence of the disclosure comprises or consists of 93 nucleotides.

In some embodiments of the compositions of the disclosure, a capped-guide RNA or a portion thereof does not comprise a nuclear localization sequence (NLS).

In some embodiments of the compositions of the disclosure, a capped-guide RNA, or
30 a portion thereof does not comprise a sequence complementary to a protospacer adjacent motif (PAM).

In some embodiments, therapeutic or pharmaceutical compositions of the disclosure do not include a PAMmer oligonucleotide. In other embodiments, optionally, non-therapeutic or non-pharmaceutical compositions may include a PAMmer oligonucleotide. In some

embodiments of the compositions of the disclosure, a guide RNA or a portion thereof includes a sequence complementary to a protospacer flanking sequence (PFS). In some embodiments, including those wherein a guide RNA or a portion thereof includes a sequence complementary to a PFS, the RNA binding protein may include a sequence isolated or
5 derived from a Cas protein, such as, without limitation, a Cas9, Cas13b, or Cas13d protein. In some embodiments, including those wherein a guide RNA or a portion thereof includes a sequence complementary to a PFS, the RNA binding protein may include a sequence encoding a Cas protein, such as, without limitation, a Cas9, Cas 13b, or Cas13d protein, or an RNA-binding portion thereof. In some embodiments, the guide RNA or a portion thereof
10 does not comprise a sequence complementary to a PFS.

Target RNA

The “target sequence” can be a stretch of nucleic acid sequences or a sequence motif present in an RNA molecule (e.g., mRNA) of interest to which a spacer sequence of the
15 capped-sgRNA hybridizes, provided sufficient conditions for hybridization exist. Hybridization between the spacer and the target sequence is, for example, based on Watson-Crick base pairing rules, which enables programmability of the spacer sequence.

In one embodiment, the mRNA including the target sequence additionally includes one or more start codons and/or an endogenous m7G cap. In another embodiment, the target
20 sequence is located downstream of an endogenous m7G cap, with its 5' end located either upstream or downstream of a desired start codon. Any start codon in the target mRNA can be selected as the desired start codon. Upon hybridization between the spacer sequence of the capped-sgRNA and the target sequence, the m7G cap of the capped-sgRNA can be recruited to the vicinity of the desired start codon, and closer in proximity to the desired start codon
25 than the endogenous m7G cap of the mRNA. In one embodiment, the m7G cap of the bound capped-sgRNA recruits translation initiation factors (e.g., EIF4E) and initiates protein translation from the desired start codon. In another embodiment, recruitment of the translation initiation factors occurs subsequent to the binding of the m7G cap of the capped-sgRNA. Without wishing to be bound by theory, the recruitment of an m7G cap via a capped-
30 sgRNA to the vicinity of a desired start codon in a transcript allows for enhanced protein translation, as compared to protein translation initiated by the endogenous m7G cap of the transcript.

In one embodiment, the target sequence includes the desired start codon of the target mRNA. For example, the 5' end of the target sequence can be upstream of the desired start

codon and the 3' end of the target sequence can be downstream of the desired start codon. In one embodiment, the 5' end of the target sequence is located between 1 and 50 nucleotides (e.g., about 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 nucleotides) upstream of the first nucleotide of the desired start codon (e.g., an "A"). In another embodiment, the 5' end of the target sequence is located between 1 and 50 nucleotides (e.g., about 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 nucleotides) downstream of the last nucleotide of the desired start codon. In some embodiments, the target sequence does not overlap with the desired start codon. The location ranges of the target sequence between about 1 and 50 nucleotides upstream or downstream of the desired start codon accounts for the differing structural properties of the various Cas proteins capable of being used with the capped-sgRNAs disclosed herein. In another embodiment, the target sequence is not required to be located between 1 and 50 nucleotides upstream or downstream from a desired start codon, rather the target sequence is located anywhere on the transcript. This is particularly relevant if the 5'UTR is large.

Direct Repeat/ Scaffold

The capped-sgRNA disclosed herein includes both a "spacer sequence" and a "direct repeat" (or "DR" or "direct repeat sequence" or "DR sequence"). In one embodiment, DR is comprised within a scaffold sequence which is capable of binding to a corresponding (or cognate) Cas polypeptide. In another embodiment, a DR is capable of binding to a corresponding (or cognate) Cas polypeptide. A direct repeat sequence disclosed herein is a repetitive sequence found within a CRISPR locus (naturally-occurring in a bacterial genome or plasmid). It is well known that one would be able to infer the DR sequence of a corresponding Cas protein if the sequence of the associated CRISPR locus is known.

Generally, a DR is a nucleic acid sequence that consists of two or more repeats of a specific sequence, i.e., nucleotide sequences present in multiple copies in the genome. A DR sequence may or may not have intervening nucleotides. In one embodiment, a DR sequence disclosed herein includes about 10 to about 100 nucleotides (e.g. about 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, or 98 nucleotides). In another embodiment, the DR sequence is orientated either 5' or 3' to a spacer within the sgRNA. In one embodiment,

the direct repeat sequence is located 5' to the spacer. In another embodiment, the DR sequence is located 3' to the spacer.

Exemplary DR sequences for the capped-sgRNAs disclosed herein are shown below.

Exemplary direct repeat sequences for Cas13a are SEQ ID Nos 284-298.

Cas13a number	Cas13a abbreviation	Organism name	Direct Repeat sequence
Cas13a1	LshCas13a	Leptotrichia shahii	CCACCCCAATATCGAAGGGGACTAAAAC (SEQ ID NO:284)
Cas13a2	LwaCas13a	Leptotrichia wadei	GATTTAGACTACCCCAAAAACGAAGGGGACTAAAAC (SEQ ID NO:285)
Cas13a3	LseCas13a	Listeria seeligeri	GTAAGAGACTACCTCTATATGAAAGAGGACTAAAAC (SEQ ID NO:286)
Cas13a4	LbmCas13a	Lachnospiraceae bacterium MA2020	GTATTGAGAAAAGCCAGATATAGTTGGCAATAGAC (SEQ ID NO:287)
Cas13a5	LbnCas13a	Lachnospiraceae bacterium NK4A179	GTTGATGAGAAGAGCCCAAGATAGAGGGCAATAAC (SEQ ID NO:288)
Cas13a6	CamCas13a	[Clostridium] aminophilum DSM 10710	GTCTATTGCCCTCTATATCGGGCTGTTCTCCAAAC (SEQ ID NO:289)
Cas13a7	CgaCas13a	Carnobacterium gallinarum DSM 4847	ATTAAAGACTACCTCTAAATGTAAGAGGACTATAAC (SEQ ID NO:290)
Cas13a8	Cga2Cas13a	Carnobacterium gallinarum DSM 4847	AATATAAACTACCTCTAAATGTAAGAGGACTATAAC (SEQ ID NO:291)
Cas13a9	Pprcas13a	Paludibacter propionigenes WB4	CTTGTGGATTATCCCAAAATTGAAGGGAACACAAC (SEQ ID NO:292)
Cas13a10	LweCas13a	Listeria weihenstephanensis FSL R9-0317	GATTTAGAGTACCTCAAAATAGAAGAGGTCTAAAAC (SEQ ID NO:293)

Cas13a1 1	LbfCas13 a	Listeriaceae bacterium FSL M6-0635 (Listeria newyorkensis)	GATTTAGAGTACCTCAAAACAAAAGAGGACTAAAAC (SEQ ID NO:294)
Cas13a1 2	Lwa2cas1 3a	Leptotrichia wadei F0279	GATATAGATAACCCCAAAAACGAAGGGATCTAAAAC (SEQ ID NO:295)
Cas13a1 3	RcsCas13 a	Rhodobacter capsulatus SB 1003	GCCTCACATCACCGCCAAGACGACGGCGGACTGAAC (SEQ ID NO:296)
Cas13a1 4	RcrCas13 a	Rhodobacter capsulatus R121	GCCTCACATCACCGCCAAGACGACGGCGGACTGAAC (SEQ ID NO:297)
Cas13a1 5	RcdCas13 a	Rhodobacter capsulatus DE442	GCCTCACATCACCGCCAAGACGACGGCGGACTGAAC (SEQ ID NO:298)

An exemplary Cas13b direct repeat sequence is:

GTTGTGGAAGGTCCAGTTTTGAGGGGCTATTACAAC (SEQ ID NO: 299)

An exemplary Cas13d (contig e-k87_11092736) Direct Repeat Sequence is:

5 GTGAGAAGTCTCCTTATGGGGAGATGCTAC (SEQ ID NO: 300)

An exemplary Cas13d (160582958_gene49834) Direct Repeat Sequence is:

GAACTACACCCCTCTGTTCTTGTAGGGGTCTAACAC (SEQ ID NO: 301)

Additional exemplary Cas13d Direct Repeat sequences are:

CACCCGTGCAAAATTGCAGGGGTCTAAAAC (SEQ ID NO: 302)

10 GACCAACACCTCTGCAAACTGCAGGGGTCTAAAAC (SEQ ID NO: 303)

AACCCCTACCAACTGGTCGGGGTTTGAAAC (SEQ ID NO: 304)

An exemplary scaffold sequence for Cas9 is:

TTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCC
GTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC (SEQ ID NO: 305)

15 Scaffold/DR sequences of the disclosure bind the CRISPR/Cas RNA-binding protein of the disclosure. Scaffold/DR sequences of the disclosure may include a trans acting RNA (tracrRNA). Upon binding to a target sequence of an RNA molecule, the scaffold/DR sequence guides a fusion protein to the RNA molecule. In some embodiments, a scaffold/ DR sequence having sufficient complementarity to a target sequence of an RNA molecule to bind
20 selectively (partially or substantially) to the target sequence has at least 50%, 55%, 60%,

65%, 70%, 75%, 80%, 85%, 90%, 95%, 96, 97%, 98%, 99%, or any percentage identity in between to the target sequence. In some embodiments, a scaffold/ DR sequence having sufficient complementarity to a target sequence of an RNA molecule to bind selectively to the target sequence has 100% identity the target sequence. In some embodiments, scaffold/DR sequences of the disclosure comprise a secondary structure or a tertiary structure. Exemplary secondary structures include, but are not limited to, a helix, a stem loop, a bulge, a tetraloop and a pseudo not. Exemplary tertiary structures include, but are not limited to, an A-form of a helix, a B-form of a helix, and a Z-form of a helix. Exemplary tertiary structures include, but are not limited to, a twisted or helicized stem loop. Exemplary tertiary structures include, but are not limited to, a twisted or helicized pseudoknot. In some embodiments, scaffold/DR sequences of the disclosure comprise at least one secondary structure or at least one tertiary structure. In some embodiments, scaffold/ DR sequences of the disclosure include one or more secondary structure(s) or one or more tertiary structure(s).

15 *Linker*

The capped-sgRNA disclosed herein can include a “linker” or “linker sequence” between the m7G cap or analog thereof and the spacer and/or DR sequences. In one embodiment, the linker sequence includes about 5 to about 25 nucleotides (e.g., about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 nucleotides). In another embodiment, the linker sequence is non-complementary to any sequence within the RNA molecule comprising the target sequence. Exemplary sequences for such a linker include, without limitation: GTCAGATCG (SEQ ID NO: 306), GTCAGATCGCCT(SEQ ID NO: 307), GTCAGATCGCCTGGA(SEQ ID NO: 308), and GTCAGATCGCCTGGAATT(SEQ ID NO: 309). Any suitable linker sequences known in the art are also contemplated herein. In some embodiments, the linker sequence is modified to adjust the editing window.

RNase P processing site

Some embodiments provide a nucleic acid encoding the capped-sgRNA described herein, where an unprocessed capped-sgRNA is first generated upon transcription of the nucleic acid. In one embodiment, an unprocessed capped-sgRNA includes an RNase P processing site, which is downstream of the spacer and/or direct repeat and upstream of a poly-A tail. In this manner, an RNase P binds to the processing site and removes the downstream sequence (e.g., the poly-A tail) to generate a capped-sgRNA disclosed herein. Exemplary RNase P processing sites can be found at Esakova and Krasilnikov, RNA

16:1725-1747, 2010 (e.g., See Figure 1 of Esakova and Krasilnikov), incorporated herein by reference in its entirety. The RNase P processing site is known to include elements recognizable by RNase P, such as those described in Kirseborn et al. Biochimie 89: 1183–1194, 2007 and Lai et al. FEBS Lett 584: 287–296, 2010, both of which are incorporated
 5 herein by reference in their entirety. For example, the RNase P processing site can include all or a portion of a bacterial (e.g., E. coli) pre-tRNA, 4.5S rRNA precursor, or yeast pre-rRNA that includes an RNase cleavage site. Structures that resemble tmRNA, operon mRNAs, phage RNAs, OLE RNA from extremophilic bacteria are also contemplated herein as RNase P processing sites. All or a portion of a viral non-tRNA such as TYMV RNA are also useful
 10 as RNase P processing sites. In some embodiments, an RNase P processing site includes a tRNA-like small RNA (e.g., GenBank Accession No. FJ209302).

An exemplary structure of an unprocessed capped-sgRNA is shown in FIG. 1B. In this embodiment, the unprocessed capped-sgRNA includes from 5' to 3': an m7G cap, a linker, a spacer, a direct repeat, an RNase P processing site, and a poly-A tail. In another
 15 embodiment, the RNase P processing site and poly-A tail is removed upon RNase P processing, thereby generating the capped-sgRNA with the structure of FIG. 1C, wherein a' is a guanosine or adenine, b' is a spacer sequence and c' is a direct repeat sequence.

III. CRISPR/Cas Polypeptides

20 The capped-sgRNAs disclosed herein are capable of binding with their cognate or corresponding RNA-binding CRISPR/Cas polypeptides (e.g., via a direct repeat sequence in the capped-sgRNA). In one embodiment, the capped-sgRNA includes a spacer sequence that confers target specificity to the Cas/sgrRNA complex. CRISPR/Cas polypeptides are well known in the art and any particular Cas polypeptide can be adapted for use in the capped-
 25 sgRNA systems disclosed herein. In some embodiments, the Cas polypeptides for use as disclosed herein have altered activity compared to its corresponding wild type Cas polypeptide. In some embodiments, the Cas polypeptides are nuclease-deficient Cas (dCas) polypeptides. Nuclease-deficient Cas polypeptides have altered (e.g., diminished or abolished) nuclease activity without substantially diminished binding affinity to the sgRNA.
 30 These Cas polypeptides are useful, for example, in mediating the direct association between the capped-sgRNA and the target mRNA, and in protecting the 3' end of the capped-sgRNA from degradation. In some embodiments, the dCas for use with the capped-sgRNA disclosed herein is devoid of cleavage activity that is applicable to the target RNA. In some

embodiments, the dCas polypeptide for use with the capped-sgRNA disclosed herein retains cleavage activity that is applicable to the capped-sgRNA. In some embodiments, the dCas13 comprises an inactivated target cleavage domain and a retained or partially retained (i.e., activated or partially activated) guide cleavage domain.

5 In one embodiment, the Cas polypeptide is Cas13b. In another embodiment, the Cas polypeptide is dead or nuclease deficient Cas13b (dCas13b). In another embodiment, the dCas13b comprises an inactivated target cleavage domain and a retained (i.e., activated) guide cleavage domain as exemplified in Figures 1B and 1D. In one embodiment, the Cas polypeptide disclosed herein is a Type II CRISPR Cas protein. In some embodiments, the

10 Type II CRISPR Cas protein includes a Cas9 protein. Exemplary Cas9 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, bacteria or archaea. Exemplary Cas9 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, *Streptococcus pyogenes*, *Haloferax mediteranii*, *Mycobacterium tuberculosis*, *Francisella tularensis subsp. novicida*, *Pasteurella multocida*,

15 *Neisseria meningitidis*, *Campylobacter jejune*, *Streptococcus thermophilus*, *Campylobacter lari* CF89-12, *Mycoplasma gallisepticum str. F*, *Nitratifractor salsuginis str. DSM 16511*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria cinerea*, a *Gluconacetobacter diazotrophicus*, an *Azospirillum B510*, a *Sphaerochaeta globus str. Buddy*, *Flavobacterium columnare*, *Fluviicola taffensis*, *Bacteroides coprophilus*,

20 *Mycoplasma mobile*, *Lactobacillus farciminis*, *Streptococcus pasteurianus*, *Lactobacillus johnsonii*, *Staphylococcus pseudintermedius*, *Filifactor alocis*, *Treponema denticola*, *Legionella pneumophila str. Paris*, *Sutterella wadsworthensis*, *Corynebacter diphtherias*, *Streptococcus aureus*, and *Francisella novicida*.

Some embodiments of the capped-sgRNA compositions or methods disclosed herein

25 provide a Cas9 polypeptide. In certain embodiments, the Cas9 polypeptide lacks part or all of the nuclease domains (e.g., the RuvC and/or HNH domains) of a wild type Cas9 polypeptide, and therefore are nuclease-deficient. These truncated Cas9 polypeptides have a smaller size as compared to a wild type Cas9. The RuvC and HNH nuclease domains can also be inactivated, for example, as a result of point mutations within these domains. For instance,

30 D10A and H840A mutations in *Streptococcus pyogenes* Cas9 (SpCas9) results in nuclease-deficient SpCas9. An exemplary sequence of a nuclease-deficient SpCas9 is SEQ ID NO: 310. The RuvC domain is distributed among 3 non-contiguous portions of the nuclease-deficient Cas9 primary structure (residues 1-60, 719-775, and 910-1099). The HNH domain is composed of residues 776-909.

SEQ ID NO: 310

	Arg	Thr	Met	Asp	Lys	Lys	Tyr	Ser	Ile	Gly	Leu	Ala	Ile	Gly	Thr	Asn
	1				5					10				15		
5	Ser	Val	Gly	Trp	Ala	Val	Ile	Thr	Asp	Glu	Tyr	Lys	Val	Pro	Ser	Lys
			20						25				30			
	Lys	Phe	Lys	Val	Leu	Gly	Asn	Thr	Asp	Arg	His	Ser	Ile	Lys	Lys	Asn
			35				40						45			
	Leu	Ile	Gly	Ala	Leu	Leu	Phe	Asp	Ser	Gly	Glu	Thr	Ala	Glu	Ala	Thr
		50					55					60				
10	Arg	Leu	Lys	Arg	Thr	Ala	Arg	Arg	Arg	Tyr	Thr	Arg	Arg	Lys	Asn	Arg
	65					70					75				80	
	Ile	Cys	Tyr	Leu	Gln	Glu	Ile	Phe	Ser	Asn	Glu	Met	Ala	Lys	Val	Asp
					85					90					95	
15	Asp	Ser	Phe	Phe	His	Arg	Leu	Glu	Glu	Ser	Phe	Leu	Val	Glu	Glu	Asp
				100					105					110		
	Lys	Lys	His	Glu	Arg	His	Pro	Ile	Phe	Gly	Asn	Ile	Val	Asp	Glu	Val
			115					120					125			
	Ala	Tyr	His	Glu	Lys	Tyr	Pro	Thr	Ile	Tyr	His	Leu	Arg	Lys	Lys	Leu
		130					135					140				
20	Val	Asp	Ser	Thr	Asp	Lys	Ala	Asp	Leu	Arg	Leu	Ile	Tyr	Leu	Ala	Leu
	145					150					155				160	
	Ala	His	Met	Ile	Lys	Phe	Arg	Gly	His	Phe	Leu	Ile	Glu	Gly	Asp	Leu
					165					170				175		
25	Asn	Pro	Asp	Asn	Ser	Asp	Val	Asp	Lys	Leu	Phe	Ile	Gln	Leu	Val	Gln
				180					185					190		
	Thr	Tyr	Asn	Gln	Leu	Phe	Glu	Glu	Asn	Pro	Ile	Asn	Ala	Ser	Gly	Val
			195					200					205			
	Asp	Ala	Lys	Ala	Ile	Leu	Ser	Ala	Arg	Leu	Ser	Lys	Ser	Arg	Arg	Leu
		210					215					220				
30	Glu	Asn	Leu	Ile	Ala	Gln	Leu	Pro	Gly	Glu	Lys	Lys	Asn	Gly	Leu	Phe
	225					230					235				240	
	Gly	Asn	Leu	Ile	Ala	Leu	Ser	Leu	Gly	Leu	Thr	Pro	Asn	Phe	Lys	Ser
					245					250				255		
35	Asn	Phe	Asp	Leu	Ala	Glu	Asp	Ala	Lys	Leu	Gln	Leu	Ser	Lys	Asp	Thr
				260					265					270		
	Tyr	Asp	Asp	Asp	Leu	Asp	Asn	Leu	Leu	Ala	Gln	Ile	Gly	Asp	Gln	Tyr
			275				280						285			
	Ala	Asp	Leu	Phe	Leu	Ala	Lys	Asn	Leu	Ser	Asp	Ala	Ile	Leu	Leu	
		290					295				300					
40	Ser	Asp	Ile	Leu	Arg	Val	Asn	Thr	Glu	Ile	Thr	Lys	Ala	Pro	Leu	Ser
	305					310					315				320	
	Ala	Ser	Met	Ile	Lys	Arg	Tyr	Asp	Glu	His	His	Gln	Asp	Leu	Thr	Leu
					325					330				335		
45	Leu	Lys	Ala	Leu	Val	Arg	Gln	Gln	Leu	Pro	Glu	Lys	Tyr	Lys	Glu	Ile
				340					345					350		
	Phe	Phe	Asp	Gln	Ser	Lys	Asn	Gly	Tyr	Ala	Gly	Tyr	Ile	Asp	Gly	Gly
			355					360					365			
	Ala	Ser	Gln	Glu	Glu	Phe	Tyr	Lys	Phe	Ile	Lys	Pro	Ile	Leu	Glu	Lys
		370					375					380				
50	Met	Asp	Gly	Thr	Glu	Glu	Leu	Leu	Val	Lys	Leu	Asn	Arg	Glu	Asp	Leu
	385					390					395				400	
	Leu	Arg	Lys	Gln	Arg	Thr	Phe	Asp	Asn	Gly	Ser	Ile	Pro	His	Gln	Ile
				405					410					415		
55	His	Leu	Gly	Glu	Leu	His	Ala	Ile	Leu	Arg	Arg	Gln	Glu	Asp	Phe	Tyr
				420					425					430		
	Pro	Phe	Leu	Lys	Asp	Asn	Arg	Glu	Lys	Ile	Glu	Lys	Ile	Leu	Thr	Phe
			435					440					445			
	Arg	Ile	Pro	Tyr	Tyr	Val	Gly	Pro	Leu	Ala	Arg	Gly	Asn	Ser	Arg	Phe
		450					455					460				
60	Ala	Trp	Met	Thr	Arg	Lys	Ser	Glu	Glu	Thr	Ile	Thr	Pro	Trp	Asn	Phe

	465				470				475				480
	Glu	Glu	Val	Val	Asp	Lys	Gly	Ala	Ser	Ala	Gln	Ser	Phe
					485					490			495
5	Met	Thr	Asn	Phe	Asp	Lys	Asn	Leu	Pro	Asn	Glu	Lys	Val
					500				505				510
	His	Ser	Leu	Leu	Tyr	Glu	Tyr	Phe	Thr	Val	Tyr	Asn	Glu
					515			520					525
	Val	Lys	Tyr	Val	Thr	Glu	Gly	Met	Arg	Lys	Pro	Ala	Phe
		530					535					540	
10	Glu	Gln	Lys	Lys	Ala	Ile	Val	Asp	Leu	Leu	Phe	Lys	Thr
							550				555		
	Val	Thr	Val	Lys	Gln	Leu	Lys	Glu	Asp	Tyr	Phe	Lys	Lys
					565				570				575
15	Phe	Asp	Ser	Val	Glu	Ile	Ser	Gly	Val	Glu	Asp	Arg	Phe
					580				585				590
	Leu	Gly	Thr	Tyr	His	Asp	Leu	Leu	Lys	Ile	Ile	Lys	Asp
					595			600					605
	Leu	Asp	Asn	Glu	Glu	Asn	Glu	Asp	Ile	Leu	Glu	Asp	Ile
		610					615					620	
20	Leu	Thr	Leu	Phe	Glu	Asp	Arg	Glu	Met	Ile	Glu	Glu	Arg
							630						640
	Tyr	Ala	His	Leu	Phe	Asp	Asp	Lys	Val	Met	Lys	Gln	Leu
					645					650			655
25	Arg	Tyr	Thr	Gly	Trp	Gly	Arg	Leu	Ser	Arg	Lys	Leu	Ile
					660					665			670
	Arg	Asp	Lys	Gln	Ser	Gly	Lys	Thr	Ile	Leu	Asp	Phe	Leu
			675					680					685
	Gly	Phe	Ala	Asn	Arg	Asn	Phe	Met	Gln	Leu	Ile	His	Asp
		690					695					700	
30	Thr	Phe	Lys	Glu	Asp	Ile	Gln	Lys	Ala	Gln	Val	Ser	Gly
							710						720
	Ser	Leu	His	Glu	His	Ile	Ala	Asn	Leu	Ala	Gly	Ser	Pro
					725								735
35	Lys	Gly	Ile	Leu	Gln	Thr	Val	Lys	Val	Val	Asp	Glu	Leu
					740				745				750
	Met	Gly	Arg	His	Lys	Pro	Glu	Asn	Ile	Val	Ile	Glu	Met
			755					760					765
	Asn	Gln	Thr	Thr	Gln	Lys	Gly	Gln	Lys	Asn	Ser	Arg	Glu
			770				775						780
40	Arg	Ile	Glu	Glu	Gly	Ile	Lys	Glu	Leu	Gly	Ser	Gln	Ile
							790						800
	His	Pro	Val	Glu	Asn	Thr	Gln	Leu	Gln	Asn	Glu	Lys	Leu
					805					810			815
45	Tyr	Leu	Gln	Asn	Gly	Arg	Asp	Met	Tyr	Val	Asp	Gln	Glu
					820				825				830
	Asn	Arg	Leu	Ser	Asp	Tyr	Asp	Val	Asp	Ala	Ile	Val	Pro
			835					840					845
	Leu	Lys	Asp	Asp	Ser	Ile	Asp	Asn	Lys	Val	Leu	Thr	Arg
			850				855					860	
50	Asn	Arg	Gly	Lys	Ser	Asp	Asn	Val	Pro	Ser	Glu	Glu	Val
							870						880
	Met	Lys	Asn	Tyr	Trp	Arg	Gln	Leu	Leu	Asn	Ala	Lys	Leu
					885					890			895
55	Arg	Lys	Phe	Asp	Asn	Leu	Thr	Lys	Ala	Glu	Arg	Gly	Gly
					900				905				910
	Leu	Asp	Lys	Ala	Gly	Phe	Ile	Lys	Arg	Gln	Leu	Val	Glu
			915					920					925
	Ile	Thr	Lys	His	Val	Ala	Gln	Ile	Leu	Asp	Ser	Arg	Met
			930				935					940	
60	Tyr	Asp	Glu	Asn	Asp	Lys	Leu	Ile	Arg	Glu	Val	Lys	Val
							950						960

Lys Ser Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys
 965 970 975
 Val Arg Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn
 980 985 990
 5 Ala Val Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser
 995 1000 1005
 Glu Phe Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile
 1010 1015 1020
 10 Ala Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035 1040
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
 1045 1050 1055
 Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly
 1060 1065 1070
 15 Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val
 1075 1080 1085
 Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr
 1090 1095 1100
 20 Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys
 1105 1110 1115 1120
 Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe
 1125 1130 1135
 Asp Ser Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu
 1140 1145 1150
 25 Lys Gly Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile
 1155 1160 1165
 Thr Ile Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu
 1170 1175 1180
 Glu Ala Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu
 1185 1190 1195 1200
 30 Pro Lys Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu
 1205 1210 1215
 Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser
 1220 1225 1230
 35 Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys
 1235 1240 1245
 Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His
 1250 1255 1260
 40 Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys
 1265 1270 1275 1280
 Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
 1285 1290 1295
 Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile
 1300 1305 1310
 45 His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr
 1315 1320 1325
 Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val
 1330 1335 1340
 Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr
 1345 1350 1355 1360
 50 Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Ala Tyr Pro Tyr Asp Val
 1365 1370 1375
 Pro Asp Tyr Ala Ser Leu (SEQ ID NO: 310)

55 Some embodiments provide a Cas polypeptide that is a Cas9 polypeptide that lacks all or a part of (1) an HNH domain, (2) at least one RuvC nuclease domain, (3) a Cas9 polypeptide DNase active site, (4) a $\beta\beta\alpha$ -metal fold comprising a Cas9 polypeptide active site, or (5) a Cas9 polypeptide that lacks all or part of one or more of the HNH domain, at

least one RuvC nuclease domain, a Cas9 polypeptide DNase active site, and/or a $\beta\beta\alpha$ -metal fold comprising a Cas9 polypeptide active site as compared to a corresponding wild type Cas9 polypeptide.

The Cas9 polypeptides described herein can be archaeal or bacterial Cas9 polypeptides. Exemplary Cas9 polypeptide include those derived from *Haloferax mediteranii*, *Mycobacterium tuberculosis*, *Francisella tularensis subsp. novicida*, *Pasteurella multocida*, *Neisseria meningitidis*, *Campylobacter jejune*, *Streptococcus thermophilus* LMD-9 CRISPR 3, *Campylobacter lari* CF89-12, *Mycoplasma gallisepticum str. F*, *Nitratifactor salsuginis str. DSM 1651 1*, *Parvibaculum lavamentivorans*, *Roseburia intestinalis*, *Neisseria cinerea*, *Gluconacetobacter diazotrophicus*, *Azospirillum B510*, *Sphaerochaeta globus str. Buddy*, *Flavobacterium columnare*, *Fluviicola taffensis*, *Bacteroides coprophilus*, *Mycoplasma mobile*, *Lactobacillus farciminis*, *Streptococcus pasteurianus*, *Lactobacillus johnsonii*, *Staphylococcus pseudintermedius*, *Filifactor alocis*, *Treponema denticola*, *Legionella pneumophila str. Paris*, *Sutterella wadsworthensis*, *Corynebacter diphtheriae*, *Streptococcus aureus*, and *Francisella novicida*.

Any Cas polypeptides with altered nuclease activity as compared to a naturally occurring Cas polypeptide is contemplated herein. Additional types of nuclease-deficient Cas polypeptides are described in e.g., Brezgin et al. Int J Mol Sci 20(23):6041, 2019 and Xu and Lei, J Mol Biol 431:34-47, 2019, incorporated by reference in its entirety.

Exemplary Cas polypeptide sequences disclosed in the methods for translational enhancement WO2019/204828 are incorporated herein by reference in their entirety and can be used in conjunction with the corresponding capped-sgRNAs disclosed herein.

In some embodiments of the compositions of the disclosure, the CRISPR Cas protein comprises a Type V CRISPR Cas protein. In some embodiments, the Type V CRISPR Cas protein comprises a Cpf1 protein. Exemplary Cpf1 proteins of the disclosure may be isolated or derived from any species, including but not limited to, bacteria or archaea. Exemplary Cpf1 proteins of the disclosure may be isolated or derived from any species, including but not limited to, *Francisella tularensis subsp. novicida*, *Acidaminococcus sp. BV3L6* and *Lachnospiraceae bacterium sp. ND2006*. Exemplary Cpf1 proteins of the disclosure may be nuclease inactivated.

In some embodiments, the CRISPR Cas protein comprises a Type VI CRISPR Cas protein or portion thereof. In some embodiments, the Type VI CRISPR Cas protein comprises a Cas13 protein or portion thereof. Exemplary Cas13 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, bacteria or

archaea. Exemplary Cas13 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, *Leptotrichia wadei*, *Listeria seeligeri* serovar 1/2b (strain ATCC 35967 / DSM 20751 / CIP 100100 / SLCC 3954), *Lachnospiraceae* bacterium, *Clostridium aminophilum* DSM 10710, *Carnobacterium gallinarum* DSM 4847, *Paludibacter propionigenes* WB4, *Listeria weihenstephanensis* FSL R9-0317, *Listeria weihenstephanensis* FSL R9-0317, bacterium FSL M6-0635 (*Listeria newyorkensis*), *Leptotrichia wadei* F0279, *Rhodobacter capsulatus* SB 1003, *Rhodobacter capsulatus* R121, *Rhodobacter capsulatus* DE442 and *Corynebacterium ulcerans*. Exemplary Cas13 proteins of the disclosure may be DNA nuclease inactivated. Exemplary Cas13 proteins of the disclosure include, but are not limited to, Cas13a, Cas13b, Cas13c, Cas13d, and orthologs thereof. Exemplary Cas13b proteins of the disclosure include, but are not limited to, subtypes 1 and 2 referred to herein as Csx27 and Csx28, respectively.

In some embodiments of the compositions of the disclosure, the sequence encoding the RNA binding protein comprises a sequence isolated or derived from a Cas13d protein or also called CasRX/Cas13d proteins. CasRX/Cas13d is an effector of the type VI-D CRISPR-Cas systems. In some embodiments, the CasRX/Cas13d protein is an RNA-guided RNA endonuclease enzyme that can cut or bind RNA. In some embodiments, the CasRX/Cas13d protein can include one or more higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domains. In some embodiments, the CasRX/Cas13d protein can include either a wild-type or mutated HEPN domain. In some embodiments, the CasRX/Cas13d protein includes a mutated HEPN domain that cannot cut RNA but can process guide RNA. In some embodiments, the CasRX/Cas13d protein does not require a protospacer flanking sequence. Also see WO Publication No. WO2019/040664 & US2019/0062724, which is incorporated herein by reference in its entirety, for further examples and sequences of CasRX/Cas13d protein.

In some instances, the Cas polypeptide has a sequence that is at least 80% identical (e.g. at least 82%, 84%, 86%, 88%, 90%, 92%, 94%, 96%, 98%, or 99% identical) to any of the Cas polypeptides (e.g., any of the wild type or nuclease-deficient Cas polypeptides) of the present disclosure.

Exemplary Cas9 sequences include, but are not limited to:

Name	Protein Sequence
------	------------------

<i>S. pyogenes</i> Cas9	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETA EATRLKRTARRRYTRRKNRICYLQEFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPI FGNIVDEVAYHEKYPTIYHLRKKLV DSTDKADLRILIYALAHMIKFRGHFLIEGDLNPD NSDVKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNG LFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLNLLAQIGDQYADLFLAAK NLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPKEYKEIFFDQ SKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQ IHLGELHAILRRQEDFYFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETIT PWNFEVVDKGGASQSFIERMTNFDKNLPNEKVLPHKSHLLYEYFTVYNELTKVKYVTE GMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASL GTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ LKRRTYTGWGRLSRKLINGIRDKQSGKTILDFLKSDFANRNFMLIHDDSLTFKEDIQ KAQVSGQGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMAREN QTTQKGQKNSRERMKRIEEGKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYV DQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNY WRQLLNAKLITQRKFDNLTKAERGGSELDAAGFIKRLQVETROITKHVAQILD SRMN TKYDENDKLIREVKVITLKSLLVSDFRKDFQFYKREINNYHHAHDAYFLNAVVGTAI KKYPKLESEFVYGDYKVVYDVRKMIKSEQIGKATAKYFFYSNIMNFFKTEITLANGEL RKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNIKKTEVQTGGFSKESILPKRNS DKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKEVGKSKKLLSVKELLGITIMERSSF EKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYV NFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL AYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDTLIHQSI TGLYETRIDLSQLGGD (SEQ ID NO: 311)
<i>Staphylococcus aureus</i> Cas9	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRR RRHRIQRVKKLLFDYNLLTDHSELGINPYEARVKGLSQKLEEEFSAALLHLAKRRGV HNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSY VKEAKQLLKVQKAYHQLDQSFDITYIDLLETRRTYIEGPGEGSPFGWKDIKEWYEM MGHCTYFPEELRSVKYAYNADLYNALNDLNNL VITRDENEKLEYEYEFQIIE NVFKQK KKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIEAELL DQIA KILTIYQSSEDIQEELTNLNSLTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDN QIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIKKYGLPNDIII ELAREKNSKDAQKMINEMQKRNRQTNERIEEII RTTGKENAKYLIEKIKLHDMQEGKC LYSLEAIPLEDLLNPNFYEV DHIIPRSVSFDNSFNKVLVKQEENSKKGNRTPFQYLLS SDSKISYETFKKHILNLA KGGRISKTKKEYLLEERDINRFSVQKDFINRLVDTRYATR GLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWFKKERNKGKHHAE DALIIANA DFIFKEWKKLDKAKKVMENQMFEKQAESMPEIETE QEYKEIFITPHQIKHIKDFKDYK YSHRVDKKNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLL MYHHPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGN KLNAHLDITDDYPNSRNKVKLSLKP YRFDVYLDNGVYKFVTVKNLDVIKENYYEV NSKCYEEAKKLKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVMIDITYR EYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG (SEQ ID NO: 312)

<i>S. thermophilus</i> CRISPR 1 Cas9	MSDLVLGLDIGISVGVGILNKVTGEIIHKNSRIFPAAQAENNLVVRTNRQGRRLARRK KHRRVRLNRLFEEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGIS YLDDASDDGNSSVGDYAQIVKENSQKLETCTPGQIQRLERYQTYGQLRGDFTVEKD GK KHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYHGPNGNEKSRT DYGRYRTSGETLDNIFGILIGKCTFYPPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKK LSKEQKNQIINYVKNEKAMGPAKLKFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYR KMKLTLETLDIEQMDRETLDKLA YVLTNLNTEREGIQEALHEFADGFSFSQKQVDEL VQF RKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTSSSNKTKYIDEK LLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNVIEMARETNEDDEKKA IQKIQKAN KDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGT TISIHD LINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKQGRTPYQALDSMDDAWSFRE LKAFVRESKTL SNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRA HKIDTKVS VVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTL VSYSEDQLLDIETGELISDDEYKESVFKAPYQHFDVTLKSKFEFDSILFSYQVDSKFN RK ISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHD PQTFEKVIEPILENYPNKQINDKGKEVPCNPFLKYKEEHGYRKYSKKKNGNGPEIKSLKYY DSKLGHNHIDITPKDSNNKVVLSVSPWRADVYFNKTTGKYEILGLKYADLQFDKGTGT YKISQEKYNDIKKKKEGVSDSEFKFTLYKNDLLL VKDTETKEQQLFRFLSRTMPKQKH YVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKN EGDKPKLDF (SEQ ID NO: 313)
<i>N. meningitidis</i> Cas9	MAAFKPNPINYILGLDIGIASVGVAMVEIDEDENPICLIDLGV RVFERAEVPKTGDSL A MARRLARSVRRLTRRRRAHRLLRARLLKREGVLQAADFENGLIKSLPNTPWQLRAA ALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHALQTGD FRTPAELALNKFEKESGHIRNQRGDYSHTFSRKDLQAE LILLFEKQKEFGNPHVSGGLK EGIETLLMTQRPALSGDAVQKMLGHCTFEP AEPKAAKN TYTAERFIWLTKLNNLRILE QGSRPLTDTERATLMDEPYRKS KLTYAQARKLLGLEDTAFFKGLRYGKDN AEASTL MEMKAYHAISR ALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEEDITGR LKDRIQPEI LEALLKHISFDFKVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPI PADEIRNPVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEEN RKDREKAAAKFREYFPNFVGEPSKDILKLRLYEQQHGKCLYSGKEINLGR LNEKGYV EIDHALPFSRTWDDSFNNKVLVLGSENQKGNQTPYEYFNGKDNSREWQEFKARVET SRFPRSKKQRILLQKFDEDEGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASN GQITNLLRGFWGLRKVRAENDRH HALDAVVVACSTVAMQQKITRFVRYKEMNAFDG KTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLR TLLAEKL SSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKS AKRLDEGVSVLRVPLTQLKLKD LEKMVNREREPKLYEALKARLEAHKDDPAKAF AEPFYKYDKAGNRTQQVKAVRVEQ VQKTGVWVRNHN GIADNATMVRVDVFEGDKYYLVPIYSWQVAKGILPDRAVVQG KDEEDWQLIDDSFNKFS LHPNDLVEVITKKARMFYFASCHRG TGNINIRIHDLDHKI GKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR (SEQ ID NO: 314)

<i>Parvibaculum</i> <i>lavamentivorans</i> Cas9	MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIPEARDPDGTPLNQRRQKRMRR RQLRRRRIRRKALNETLHEAGFLPAYGSADWPVVMADPEYELRRRGLEEGLSAYEFGR AIYHLAQHRHFKGRELEESDTPDPDVDDKEEAANERAATLKALKNEQTTLGAWLARR PPSDRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEMRARISDTIFAQRPVFWRKN TLGECRFMPGEPLCPKGSWLSQQRRMLEKLNLAAGGNARPLDAEERDAILSKLQQQ ASMSWPGVRSALKALYKQRGEPGAESLKFNLGEGESKLLGNALAEKLAADMFGPD WPAHPRKQEIRHAVHERLWAADYGETPDKKRVILSEKDRKAHREAAANSFVADFGIT GEQAAQLQALKLPTGWEPYSIPALNLFALAELEKGERFGALVNGPDWEGWRRTNFPHR NQPTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIEIVG RDVGKSKREREEIQSGIRNEKQRKKATEDLIKNGIANPSRDDVEKWILWKEGQERCP YTGDIQGFNALFREGRYEVEHIWPRSRSFNSPRNKTLCRKDVNIEKGNRMPPFEAFGH DEDRWSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQIL AQLKRLWPDMPGEAPVKVEAVTGQVTAQLRKLWTLNNLADDGEKTRADHRHHAID ALTVACTHPGMTNKLSTRYWQLRDDPRAEKPALTPPWDITIRADAEKAVSEIVVSHRVR KKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGELDEIRDPRIKEIVAAHV AGRGGDPKKAFFPPYPCVSPGPEIRKVRILTQSKQQLNLMAQTGNNGYADLGSNHHIAIYR LPDGKADFEIVSLFDASRRLAQRNPVQRTRADGASFVMSLAAGEAIMIPEGSKKGIWI VQGVWASGQVVLERD TDADHSTTTRPMPNPILKDDAKKVSIDPIGRVRPSND (SEQ ID NO: 315)
<i>Corynebacter</i> <i>diphtheria</i> Cas9	MKYHVGIDVGTFVGLAAIEVDDAGMPIKTLVSHIHDSGLDPDEIKSAVTRLASSGI ARRTRRLYRRKRRRLQQLDKFIQRQGWVPIELEDYSDPLYPWKVRAELAASYIADEKE RGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRASGQPVPETATV GQMVTLCGLGTLKLRGEGGVL SARLQSDYAREIQEICRMQEIGQELYRKIIDVVFAAE SPKGSASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKNL VFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNS RIAPLVDWWKTASALEQHAMVKALSNAEVDFFDSPEGAKVQAFFADLDDDVHAKLD SLHLPVGRAAYSED TLVRLTRRMLSDGVLDYTARLQEFGEPSWTPPTPRIGEPVGNPA VDRVLKTVSRWLESATKTWGAPERVIIHVREGFVTEKRAREMDGDMRRRAARNAK LFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQAYCGSPITFSNSEMDHIVPRAGQG STNTRENLVAVCHRCNQSKGNTPF AIWAKNTSIEGVSVEKAEVERTRHWVTD TGMRST DFKKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFAHGTTVRVYR GSLTAEARRASGISGKLKFFDGVGKSRLDRRHHIDA AVIAFTSDYVAETLAVRSNLK QSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVRVMSNV RLRLNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANP ERHVRVNGTHVYAGDNIGLFPVSAGSIALRGGY AELGSSFHARVYKITSKKPAFAML RVYTIDLLPYRNQDLFSVELKPQTMSMRQAEEKLRDALATGNAEYLGWLVDDELV VDTSKIATDQVKA VEAE LGTIRRWVRVDGFFSPSKLRLRPLQMSKEGIKKESAPELSKIID RPGWLPVANKLFS DGNVTVVRRDSLGRVRLESTAHLPVTWKVQ (SEQ ID NO: 316)

<i>Streptococcus pasteurianus</i> Cas9	MTNGKILGLDIGIASVGVGIIIAKTGKVVHANSRLFSAANAENNAERRGFRGSRRLNRR KKHRVKRVRDLFEKYGIVTDFRNLNLPYELRVKGLTEQLKNEELFAALRTISKRRGIS YLDDEADDSTGSTDYAKSIDENRRLKKNKTPGQIQLEKYGQLRGNFTVYDENG HRLINVFSTSDYEKEARKILETQADYNKKITAEFIDYVEILTQKRKYHGPNEKSRT DYGRFRDGTTLNIFGILIGKCNFYDPDEYRASKASYTAQEYNFLNDLNNLKVSTETGK LSTEQKESLVEFAKNTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGPDLHTFEPYR KLKFNLESINIDDLRSREVIDKLADILTLNTEREGIEDAIKRNLPNQFTEEQISEIIVRKSQS TAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKSSKNTKTIDEKEVTD EIYNPVVAKSVRQTIKIINAADVKKYGDFFDKIVIEMPRDKNADDEKKFIDKRNKENKKEK DDALKRAAYLYNSSDKLPDEVFHGNKQLETIRLWYQQGERCLYSGKPISIQELVHNS NNFEIDHILPLSLSFDDSLANKVLVYAWTNQEKGGKTPYQVIDSMDAAWSFREMMDY VLKQKGLGKKRDYLLTTENIDKIEVKKKFIERNLVTRYASRVVLNSLQSLRELK DTKVSVVRGQFTSQLRRKWKIDKSRETYHHHAVALIAASSQLKLWEKQDNPMFVD YGKNQVVDKQTGEILSVSDDEYKELVFQPPYQGFVNTISSKGFEDILFSYQVDSKYNR KVSDATIIYSTRKAKIGKDKKEETYVLGKIKDIYSQNGFDTFIKKYNKDKTQFLMYQKD SLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEYRRENGLICKYSKKGKGTPIKSLK YYDKKLGNCIDITPEESRNKVLQSNPWRADVFNPETLKYELMGLKYSDDLSEKGTG NYHISQEKYDAIKEKEGIGKKSEFKFTLYRNDLILIKDIASGEQEIYRFLSRTPMPNVNHY VELKPYDKEKFDNVQELVEALGEADKVGRCIKGLNKPNIYKVRTDVLGNKYFVKK KGDKPKLDFKNNKK (SEQ ID NO: 317)
<i>Neisseria cinerea</i> Cas9	MAAFKPNPMNYILGLDIGIASVGWAIVEIDEEENPIRLIDLGVRFERAEPKTGDSLAA ARRLARSVRRLTRRRRAHRLLRARLLKREGVLQAADFDENGLIKSLPNTPWQLRAAAL DRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNTHALQTGDFR TPAELALNKFESGHIRNQRGDYSHTFNRKDLQAEINLLFEKQKEFGNPHVSDGLKE GIETLLMTQRPALSGDAVQKMLGHCTFEPTPKAAKNTYTAERFVWLTKLNNLRILEQ GSRPLTDTERATLMDEPYRKSCLTYAQARKLLDLDDTAFFKGLRYGKDNAAEASTLM EMKAYHAISRALKEKEGLKDKKSPNLSPQLQDEIGTAFSLFKTDEDITGRLKDRVQPEIL EALLKHISFDKFVQISLKAARRIVPLMEQGNRYDEACTEYGDHYGKKNTEEKIYLPPI ADEIRNPVVLRLALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENR KDREKSAKFREYFPNFVGEPKSKDILKLRLYEQQHKGCLYSGKEINLGRLEKGYVEI DHALPFSRTWDDSFNNKVLALGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSR FPRSKKQRILLQKFDEDFGKERNLNDTRYINRFLCQFVADHMLLTGKGKRRVFASNGQ ITNLLRGFWGLRKVRAENDRHHAALDAVVVACSTIAMQQKITRFVRYKEMNAFDGKTI DKETGEVLHQAHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKLSS RPEAVHKYVTPLFISRAPNRKMSGQGHMETVKSARKLDEGISVLRVPLTQLKLDLEK MVNREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKA VRVEQVQK TG VVHNHNGIADNATIVRVDVFEKGGKYYLVPIYSWQVAKGILPDRAVVGKDEE DWTVMDDSFEFKFLYANDLIKLTAKKNEFLGYFVSLNRATGAIDIRTHDSTDSTKGKN GIFQSVGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR (SEQ ID NO: 318)

<i>Campylobacter lari</i> Cas9	MRILGFDIGINSIGWAFVENDELKDCGVRIFTKAENPKNKESLALPRRNARSSRRRLKR RKARLIAIKRILAKELKLNKYDYVAADGELPKAYEGSLASVYELRYKALTQNLTKDL ARVILHIAKHRGYMKNKNEKSSNDACKGKILSALKNNALKLENYQSVGEYFYKEFFQK YKKNTKNFIKIRNTKDNYNKNCVLSDDLEKELKLILEKQKEFGYNYSEDFINEILKVAFFQ RPLKDFSHLVGACTFFEEEEKRACKNSYSAWEFVALTKIINEIKSLEKISGEIVPTQTINEV LNLILDKGSITYKKFRSCINLHESISFKSLKYDKENAENAKLIDFRKLVEFKKALGVHSL SRQELDQISTHITLIKDNVKKLTVLEKYNLSNEQINNLEIEFNDYINLSFKALGMILPLM REGKRYDEACEIANLKPKTVDEKKDFLPAFCDSIFAHELS NPVVNRAISEYRKVLNALL KKYGKVHKKIHLELARDVGLSKKAREKIEKEQKENQAVNAWALKECENIGLKASAKNI LKLKLWKEQKEICISGNKISIEHLKDEKALEVDHIYPYSRFSDDSFINKVLVFTKENQE KLNKTPFEAFGKNIEKWSKIQTALQNLPHYKKKNKILDENFKDKQQEDFISRNLDTRYI ATLIAKYTKEYLNFLLSENENANLKSGEKGSKIHVQTSIGMLTSVLRHTWGFDDKDR NNHLHHALDAIIVAYSTNSIIKAFSDFRKNQELLKARFYAKELTSDNYKHQVKFFEPFK SFREKILSKIDEIFVSKPPRKRARRALHKDTFHSENKIIDKCSYNSKEGLQIALSCGRVRK IGTKYVENDTIVRVDIFKKQNKFYAIPIYAMDFALGILPNKIVITGKDKNNNPQWQTID ESYEFCFSLYKNDLILLQKKNMQEPEFAYYNDFSISTSSICVEKHDNKNFENLTSNQKLLF SNAKEGSVKVESLGIQNLKVFEKYIITPLGDKIKADFPRENISLTKSKKYGLR (SEQ ID NO: 319)
<i>T. denticola</i> Cas9	MKKEIKDYFLGLDVGTSVGWAVTDDYKLLKANRKDLWGMRCFETAETAEVRRHLH RGARRRIERRKKRIKLQELFSQEIAKTDEGFFQRMKESPFYAEDKTILQENTLFNDKDF ADKTYHKAYPTINHLIKAWIENKVKPDPRLLYLACHNIIKKRGHFLFEGDFDSENQFDT SIQALFEYLREDMEVDIDADSQKVKEILKDSLSKNSKQSRNLKILGLKPSDKQKKAIT NLISGNKINFADLYDNPDLDKAEKNSISFSKDDFDALSDDLASILGDSFELLKAKAVY NCSVLSKVIGDEQYLSFAKVKIYEKHKTDLTCLKNVIKKHFPKDYKKVFGYNKNEKN NNNYSGYVGVCCKTSKKLIINNSVNQEDFYKFLKTILSAKSEIKEVNDILTEIETGTFLP KQISKNAEIPYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIIMLLTFKIPYIIGPI NDNHKKFFPDRCWVVKKEKSPSGKTPWNFFDHIDKEKTAFAFITSRTNFCTYLVGES VLPKSSLLYSEYTVLNEINNLIQIIIDGKNICDKLKKQKIYEDLFKKYKKITQKQISTFIKHE GICNKTDEVIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIRWATIIYDEGEKGT ILKTKIKAEYGKYCSDEQIKILNLKFSGWGRLSRKFLETVTSEMPGFSEPVIITAMRE TQNNLMELLSSEFTFTENIKKINSGFEDAEEKQFSYDGLVKPLFLSPSVKKMLWQTLKL KEISHITQAPPKKIFIEMAKGALEPARTKTRLKILQDLNNCKNDADAFSSEIKDLSGKI ENEDNLRRLSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNYDIDHIYPQSKIKDDISNR VLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNNFISLEKLNRLTRATPISDDDETAK FIARQLVETRQATKVAKVLEKMFPEKIVYSKAETVSMFRNKFIVKCREINDFHHA HDAYLNIVVGNVYNTKFTNNPWNFIKEKRDNPKIADTYNYKVFYDVKRNNITAW KGKTIITVKDMLKRNTPIYTRQAACKKGELFNQTIMKKGLGQHPLKKEGPFNSISKYGG YNKVSAAYYTLIEYEEKGNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVP KIKINSLLKINGFPCHITGKTNDLSFLLRPAVQFCCSNNEVL YFKKIIIRFSEIRSQREKIGKTI SPYEDLSFRSYIKENLWKKTKNDEIGEKEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPN SATIDILVKGKEKFKSLIENQFEVILEILKLSATRNVSDLQHIIGGSKYSVAKIGNKISS LDNCILYQSITGIFEKRIDLLKV (SEQ ID NO: 320)
<i>S. mutans</i> Cas9	MKKPYSIGLDIGTNSVGWAVVTDDYKVPKMKMKVLGNTDKSHIEKNLLGALLFDSGN TAEDRRLKRTARRRYTRRRNRILYLQEIFSEEMGKVDDSFHRLSDSFLVTEDKRGERH PIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDT RNNDVQRLFQEFLLAVYDNTFENSSLQEQNQVEEILTDKISKSARKDRVLKLPNEKSN GRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEEELVLLAQIGDNYAELFLSAK KLYDSILLSGILTVTDVGTAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFS VSKDGYAGYIDGKTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPH QIHLQEMRAIIRQAIFYPLADNQDRIEKLITFRIPYVVGFLARGKSDFAWLSRKSAD KITPWNFEIIVDKESSAEAFINRMTNYDLYLPNQKVLPHSLLYEKFYVYNELTKVKY KTEQGKTAFFDANMKQEIFDGVFVKVYRKVTKDKLMDLFLEKEFDEFRIVDLTGLDKEN KVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLNYSDDL KEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIIDGNSNRNFMQLINDDAL FKEEIAKAQVIGETDNLNQVVSIDAGSPAIIKKGILQSLKIVDELVKIMGHQPENIVVEMA RENQFTNQGRNSQQRKGLTDSIKEFGSQILKEHPVENSQQLQNDRLFLYYLQNGRDM YTGEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSY WSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGFIKQRLVETRQITKHVARILDERFNT

	ETDENNKIRQVKIVTLKSNLVSFRKEFELYKREINDYHHAHDAYLNAVIGKALLG VYPQLEPEFVYGDYPHFHGHKKNKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKD EHISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPRKTCKFYWDTKKYGG FDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLEKRGYRNVQE ENIIKLPKYSFLKLENGRKRLLASARELQKGNEIVLPNHLGTLTYHAKNIHKVDEPKHL DYVDKHKDEFKELLDVVSNSKKYTLAEGNLEKIKELYAQNGEDLKELASSFINLLT FTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQSGITGLYETRIDLNLKGGD (SEQ ID NO: 321)
<i>S. thermophilus</i> CRISPR 3 Cas9	MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGIT AEGRRKRTARRRYTRRRNRILYLQEIFSTEMATLDDAFFQRLDDSFVPDDKRDSKYP IFGNLVEEKAYHDEFPTIYHLRKYLADESTKKADRLVYLAALAHMIKYRGHFLIEGFNS KNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEVKDKISKLEKKDRILKLPGEKNSGI FSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDETLGTYIGDDYSDFLAKKL YDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKYNEVFKDDT KNGYAGYIDGKTNQEDFYVYLKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQI HLQEMRAILDKQAKFYFPLAKNKERIEKILTFRIPYYVGPLARGNSDFAWSIRKRNEKIT PWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFAES MRDYQFLDSKQKKDIVRLYFKDKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLST YHDLNINDDKEFLDDSSNEAIIIEIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLSRRH YTGWGLSAKLINGIRDEKSGNTILDYLDGINSRNFMLIHDDALSFKKKIQKAQIIG DEDKGNIEVVKSLPGSPAIIKQSIKIVDELVKVMGGRKPESIVVEMARENQYTNQ GKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYLQNGKDMYT GDDLIDRLSNYDIDHIIPQAFKDNISDNKVLVSSASNRGKSDDVPSLEVVKRRTFW YQLLKSLLISQRKFDNLTKAERGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNK KDNNRAVRTVKIITLSTLVSQFRKDFELYKREINDYHHAHDAYLNAVVASALLKK YPKLEPEFVYGDYPKYNSFRERKSATEKVYFYSNIMNIFKKSSISADGRVIERPLIEVNE ETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGPKGLFNANLSSKPKP NSNENLVGAKEYLDPKKYGGYAGISNSFTVLVKGTEIEGAKKKITNVLEFQGISILDRIN YRKDKLNLLEKGYKDIEIHELKYSLEFSDGSRMLASILSTNNKRGEIHKGNQIFLS QKFVKLLYHAKRISNTINENHRKYVENHKKFEFEELFYIIEFNENYVGAKKNGKLLNS AFQSWQNHSIDELCSSFIGPTGSRKGLFELTSRGSAADEFGLGVKIPRYRDTTPSSLLK DATLIHQSVTGLYETRIDLAKLGEG (SEQ ID NO: 322)
<i>C. jejuni</i> Cas9	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLAR RKARLNHLKHLIANEFKLNIEDYQSFEDESLAKAYKGLSPYELRFRALNELLSKQDFA RVILHIAKRRGYDDIKNSDDKEKGAILKAIQNEEKLANYQSVGEYLYKEYFQKFEN SKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEVLSVAFYKRALK DFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNLKNTEGILYTKDDLNALLN EVLKNGTLTYKQTKLLGLSDDYEFKGEKGTYFIEFKYKEFIKALGEHNSQDDLNEI AKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPMLEGKKY DEACNELNLKVAINEDKKDFLPAFNETYYKDEVNTPVVLRAIKEYRKVLNALLKKY KVHINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKILKRLF KEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYRSFDDSYMKNVLVFTKQNEKLNQT PFEAFGNDSAKWQKIEVLAKNLPKQKQKRLDKNYKDKEQKNFKDRNLNDTRYIARL VLNYTKDYLDLPLSDDENTKLNLTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRN NHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYKNKRKFFEPFSGF RQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVKALELGKIRKVN GKIVKNGDMFRVDIFKHKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILM DENYEFCSLYKDSLILIQTKDMQEPEFVYNAFTSSTVSLIVSKHDNKFETLSKNQKIL FKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK (SEQ ID NO: 323)

<i>P. multocida</i> Cas9	MQTTNLSYLGLDLGASVGWAVVEINENEDPIGLIDVGVRIFERAEPKGTGESLALSRR LARSTRRLIRRRRAHRLLLAKRFLKREGILSTIDLEKGLPNQAWELRVAGLERRLSAIEW GAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQSDDYRTPAELALKK FAKEEGHIRNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTELLMW QKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAERFVWLTKLNNLRILEDGAERALNE EERQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKHLRYSKENAESATFMELKAWHAIR KALENQGLKDTWQDLAKKPDLLDEIGTAFSLYKTDEDIQQYLTKNVPNSVINALLVSL NFDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHHYGEANQKTSQLLPAIPAQEIRNP VVLRTLSQARKVINAIIRQYGSPARVHIETGRELKGSFKERREIQKQEDNRTKRESAV QKFKELFSDFSSEPESKDILKFRLYEQQHKGCLYSGKEINIHLNEKGYVEIDHALPFSR TWDDSFNNKVLVLASENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAAKKQ RLLTQVIDDNKFIDRNLDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSR WGLIKARENNNRHHALDAIVVACATPSMQQKITRIFIRFKEVHPYKIENRYEMVDQESG EII SPHFPEPWAYFRQEVNIRVFDNHPDVLKEMLPDRPQANHGFVQPLFVSRAPTRKM SGQGHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVKEREPALYAGLKARLAEFN QDPAKAFATPFYKQGGQVKAIRVEQVQKSGVLVRENNGVADNASIVRTDVFIKNNK FFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMDEGAKFKFSLFPNDLVELKTKKEYF FGYYIGLDRATGNISLKEHDGEISKGKDGVYRVGVKLALSFEKYQVDELGKNRQICRP QQRQPVR (SEQ ID NO: 324)
<i>F. novicida</i> Cas9	MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNKGKVELSKDSYTLLMNNRTAR RHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLNRRGFSFITDGYSPEYLN IVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLMQKILEFKLMKCLTDI KDDKVSTKTLKEITSYEFELLADYLANYESLKTQKFSYTDKQGNLKELSYHHDKYN IQEFLKRHATINDRILDTLTDDLDIWNFNFEKFDKNEEKLQNQEDKDHIQAHLLHF VFAVNKIKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSNLSVK NLVNLIGNLSNLELKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGEWRVGVDQD KKDGAKYSYKDLCNELKQKVTKAGLVDFLELDPCRTIPPYLDNNNRKPPKCQSLILN PKFLDNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYKSSNQIA SGQRDYKDLDARILQFIFDRVKADELLENEIYFQAKKLLQKASSELEKLESSKLLDEVI ANSQLSQILKSQHTNGIFEQGTFLHLVCKYYKQQRQARDSRLYIMPEYRYDKKLHKYN NTGRFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFIKWL VEHIRGFKKACEDSLKIQKDNRGLLNHKNINARNTKGKCEKEIFNLICKIEGSEDKKNY KHGLAYELGVLLFGEPNEASKPEFDRKIKKFNSIYSAFIQQIAFAERKGNANTCAVCS ADNAHRMQQIKITEPVEDNKKIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDN WQNIKQVLSAKHQLHIPIITESNAFEFEPALADVKGKSLKDRRKKALERISPENIFKDNK NRIEFAKGISAYSGANLTDGDFDGAKEELDHIIPRSHKKYGTLNDEANLICVTRGDNDK NKGNRIFCLRDLDADNYKLLQFETDDLEIEKKIADTIWDANKKDFKGNYSFINLTPQ EQKAFRHALFLADENPIKQAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLN TDKISFDYFGIPTIGNRGIAEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIA ADEHRNDGSIGLEIDKNYSLYPLDKNTGEVFTKDIFSQKITDNEFSDDKLLVRKKAIEGF NTHRQMTRDGIYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVY CLKFVDKPIDIQISTLEELRNILTTNNAATAEYYYINLKTQKLHEYYIENYNTALGYK KYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYREWQN TTIKDDYEFLLKSFNVKSITKLHKKVRKDFSLPISTNEGKFLVKKRTWDNNFIYQILNDS DSRADGTPKPFIPAFDISKNEIVEAIIIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEV ETPSDLRDIGIATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEI LKQSTIIEFESSGFNKTIKEMLGMLAGIYNETSNN (SEQ ID NO: 325)
<i>Lactobacillus buchneri</i> Cas9	MKVNNYHIGLDIGTSSIGWVAIGKDGKPLRVKGKTAIGARLFQEGNPAADRRMFRTTR RRLSRRKWRLKLLIEIFDPYITPVDSTFFARLKQSNLSPKDSRKEFKGSMLFPDLTDMQ YHKNYPTIYHLRHALMTQDKKFDIRMVYLAHHIVKYRGNFLNSTPVDSFKASKVDFV DQFKKLNELYAAINPEESFKINLANSEDIGHQFLDPSIRKFDKKKQIPKIVPVMNDKVT DRLNGKIASIIEHAILGYKAKLDVVLQCTPVDSKPAWALKFDDDEDIDAKLEKILPEMDEN QQSIVAILQNLYSQVTLNQIVPNGMSLSSESMIEKYNDHHDHLKLYKKLIDQLADPKKK AVLKKAYSQYVGDDGKVIEQAEFWSSVKKNLDDSELSKQIMDLIDAEKFMKQRTSQ NGVIPHQLHQRELDIEHQSKEYYPWLVEINPNKHDLHLAKYKIEQLVAFRVPYYVGP MITPKDQAESAETVFSWMERKGTETGQITPWNFDEKVDKASANRFIKRMTTKDTYLI GEDVLPDESLLYEKFKVLNELNMVRVNGKLLKVADKQAIFQDLFENYKHVSVKKLQN

	YIKAKTGLPSDPEISGLSDPEHFNNLSLGTYNDFKKLFGSKVDEPDLQDDFEKIVEWSTVF EDKKILREKLNEITWLSQQKDVLSSRYQGWGRLSKLLTGIVNDQGERIIDKLWNT NKNFMQIQSDDDFAKRIHEANADQMQA VDVEDVLADAYTSPQNKKAIRQVVKVDD IQKAMGGVAPKYISIEFTRSEDRNPRRTISRQRQLENTLKD TAKSLAKSINPELLSELDN AAKSKKGLTDRLYL YFTQLGKDIYTGE PINIDELNKYDIDHILPQAFIKDNSLDNRVLVL TAVNNGKSDNVPLRMFGAKMGHFWKQLAEAGLISKRKLKLNLTDPDTISKYAMHGFI RRQLVETSQVIKLVANILGDKYRNDDTKIIEITARMNHQMRDEFGFIKNREINDYHHAF DAYLTAFLGRYL YHRYIKLRPYFVYGDFFKFREDKVTMRNFNFLHDLTDDTQEKIAD AETGEVIWDRENSIQQLKDVYHYKFMLISHEVYTLRGAMFNQTVYPASDAGKRKLIPV KADRPVNVYGGYSGSADAYMAIVRIHNKKGDKYRVVGVPMRALDRLDAAKNVSDA DFDRALKDVLAPQLTKTKKSRKTGEITQVIEDFEIVLGKVMYRQLMIDGDKKFMGLSS TYQYNAKQLVLSQSVKTLASKGRLDPLQESMDYNNVYTEILDKNQYFSLYDMNKF RHKLNLGFSKFISFPNHNVDGNTKVSSGKREILQEILNGLHANPTFGNLKDVGITTPFG QLQQPNGILLSDETKIRYQSPTGLFERTVSLKDL (SEQ ID NO: 326)
<i>Listeria innocua</i> Cas9	MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQ TAADRRMARTARRRIERRNRISYLGIFAEEMSKTDANFFCRLSDSFYVDNEKRNRSR HPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALD TQNTSVDIGYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKL YPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKDSYEEDLESLLALIGDEYAE LFVAAKNAYS AVVLSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYE EIFSNTEKHGYAGYIDGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFD NGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLANGQSEFAWL TRKADGEIRPWNIEEKVDFGKSAVD FIEKMTNKD TYLPKENVLPKHSLCYQKYL VYNE LTKVRYINDQGKTSYFSGQEKEQIFNDL FKQKRKVKKDLFLRNMSHVESPTIEGLE DSFNSSYSTYHDLLKVGKQEILDNPVNTEMLENIVKILTVFEDKRMKEQLQQFSDVL DGVVLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLINDS NLSFKSIIKEQVTTADKDIQSIVADLAGSPAIIKGILQSLKIVDELVSVMGYPPQTIVVE MARENQTTGKGKNNSRPRYKSLEKAIKEFGSQLKEHPTDNQELRNNRLYLYYLQNGK DMYTGGDLDIHNL SNYDIDHIVPQS FIDNSIDNLVLTSSAGNREKGDDVPPEIVRKRK VFWEKLYQGNLMSKRKFDYLTKAERGGGLEADKARFIHRQLVETRQITKNVANILHQ RFNYEKDDHGNTMKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVV ANTLLKVYPQLEPEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEI LWDKKYLDTVKKVMSYRQMNIVKKTEIQKGEFSKATIKPGNSSKLIPRKTNWDPMK YGGLDSPNMAYAVVIEYAKGKNKL VFEKKIIRVTIMERKA FEKDEKAFLEEQGYRQPK VLAKLPKYTLYECEEGRRRMLASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSL DYIESNREMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIQAQSFVDLMAF NAMGAPASFKEFFETTERKRYNNLKELLNSTIYQSITGLYESRKRLDD (SEQ ID NO: 327)

<i>L. pneumophila</i> Cas9	MESSQILSPIGIDLGGKFTGVCLSHLEAFAELPNHANTKYSVILIDHNNFQLSQAQRRAT RHRVRNKKRNQFVKRVALQLFQHILSRDLNAKEETALCHYLNNRGYTYVDTDLDEYI KDETTINLLKELLPSESEHNFIDWFLQKMQSSEFRKILVSKVEEKDDKELKNAVKNIK NFITGFEEKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKIPSVCLSNLLGHLSNLQWK NLHRYLAKNPKQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQOLEQSODYISILE KTPPEITIPPYEARTNTGMEKDQSLLLNPEKLNLYPNWRNLIPGIIDAHFLEKDLEHT KLRDRKRISPSKQDEKRDSYILQRYLDLNNKKIDKFKIKKQLSFLGQGGKQLPANLIETQK EMETHFNSSLVSVLIQIASAYNKEREDAAQGIWFDNAFSLCELSNINPPRKQKILPLLVG AILEDFFINNKKDKWAKFKIFWNTHKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEALNH PEHSNNKALIKIQTIPDIIQAIQSHLGHNDQSALYHNPFSLSQLYTILETKRDGFHNKCV AVTCENYWRSQKTEIDPEISYASRLPADSVRPFDFGLARMQRLAYEIAMAKWEQIK HIPDNSSLLIPIYLEQNRFEFEESFKKIKGSSSDKTLEQAIEKQNIQWEEKFQRIINASMNI CPYKGASIGGQGEIDHIYPRSLSKKHFGVIFNSEVNLIYCSSQGNREKKEEHYLLLEHLS LYLKHQFGTDNVSDIKNFISQNVANIKKYISFHLTPEQQAARHALFLDYDDEAFKTI TKFLMSQQKARVNGTQKFLGKQIMEFLSTLADSKQLQLEFSIKQITAEVVDHRELLSK QEPKL VKSRQQSFPSHAIDATLTMSIGLKEFPQFSQELDN SWFINHLMPEVHLNPVRS KEYKNPNISSTPLFKDSL YAEFIPVWVKGETFAIGFEKDLFEIKPSNKEKLFLLKTY STKNPGESLQELQAKSAKAWLYFPINKTLAEFLHHYFHKIEIVTPDDTTVCHFINSLRY YTKKESITVKILKEPMPVLSVKFESSKKNVLGSFKHTIALPATKDWERLFNHPNFLALK ANPAPNPKEFNEFIRKYFLSDNNPNSDIPNNGHNIKPKQKHKA VRKVFSLPVPGNAGTM MRIRRKDNKGQPL YQLQTIDDTPSMGIQINEDRLVKQEVLM DAYKTRNLSTIDGINNSE GQAYATFDNWL TLPVSTFKPEIKLEMKPHSKTRRYIRITQSLADFIKTIDEALMIKPSDS IDDP LNMPNEIVCKNKLFGNELKPRDGKMKIVSTGKIVTYEFESDSTPQWIQTL YVTQL KKQP (SEQ ID NO: 328)
<i>N. lactamica</i> Cas9	MAAFKPNPMNYILGLDIGIASVGWAMVEVDEENPIRLIDLGVRFERAEPKTDGDSL AMARRLARSVRRRLTRRRARHLLRARRLLKREGVLQADDFDENGVLKSLPNTPWQLRA AALDRKLTCLWSAVLLHLVKHRGYLSQRKNEGETADKELGALLKGADNAHALQT GDFRTPAELALNKFESGHIRNQRGDYSHTFSRKDLQAEINLLFEKQKEFGNPHVSD GLKEDIETLLMAQRPALSGDAVQKMLGHCTFEPAEPKAAKNITYAERFIWLTCLNNLR ILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEAS TLMEMKAYHAISRALEKEGLKDKKSPLNLSLTELQDEIGTAFSLFKTDKIDITGRKDRVQ PEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYCKKNAAEKIYL PIPADEIRNPVRLRALSQARKVINCVRVRRYGSAPARIHIETAREVGKSFKDRKEIEKRQE ENRKDREKAAAKFREYFPNFVGEPSKDILKLRLYEQQHGKCLYSKGEINLVRNLEKG YVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARV ETSRFPRSKKQRILLQKFDEEGFKERNLNDTRYVNRFLCQFVADHILLTGKGKRRVFAS NGQITNLLRGFWGLRKVRTENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFD GKTIDKETGEVLHQKAHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAE KLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGISVLRVPLTQLKLD GLEKMNVREREPKLYDALKAQLETHKDDPAKAFAPFYKYDKAGSRTQQVKAVRIEQ VQKTGVWVRNHNHGIADNATMVRVDVFEKGGKYLVPIYSWQVAKGILPDRAVVAFAK DEEDWTVMDDSEFRFVLYANDLIKLTAKKNEFLGYFVSLNRATGAIDIRTHDSTDSTK GKNGIFQSVGVKLTALSQKQNDIDELGKEIRPCRLKKRPPVR (SEQ ID NO: 329)
<i>N. meningitides</i> Cas9	MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRFERAEPKTDGSLA MARRLARSVRRRLTRRRARHLLRARRLLKREGVLQADDFDENGVLKSLPNTPWQLRAA ALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGADNAHALQTGD FRTPAELALNKFESGHIRNQRGDYSHTFSRKDLQAEILLLFEKQKEFGNPHVSGGLK EGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNITYAERFIWLTCLNNLRILE QGGERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTL MEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDDEITGRKLDRIQPEI LEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPI PADEIRNPVRLRALSQARKVINGVRVRRYGSAPARIHIETAREVGKSFKDRKEIEKRQEEN RKDREKAAAKFREYFPNFVGEPSKDILKLRLYEQQHGKCLYSKGEINLGRNLEKGYV EIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVET SRFPRSKKQRILLQKFDEEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKRRVFASN GQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDG KTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKL SSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSARKLDEGVSVLRVPLTQLKLD LEKMNVREREPKLYEALKARLEAHKDDPAKAFAPFYKYDKAGNRTQQVKAVRVEQ VQKTGVWVRNHNHGIADNATMVRVDVFEKGDKYLVPIYSWQVAKGILPDRAVVGQ

	KDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRTGNINIRIHDLDHKI GKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR (SEQ ID NO: 330)
<i>B. longum</i> Cas9	MLSRQLLGASHLARPVSYSYNVQDNDVHCSYGERCFMRGKRYRIGIDVGLNSVGLAA VEVSDENSPVRLNAQSVIHGGVDPQKNKEAITRKNMSGVARRTRMRRRKRERLH KLDMLLGKFGYPVIEPESLDKPFEEWHVRAELATRYIEDDELRESISIALRHMARHRG WRNPYRQVDSLISDNPYSKQYGELKEKAKAYNDATAAEEESTPAQLVVAMLDAGY AEAPRLRWRTGSKKPDAEGYLPVRLMQEDNANELKQIFRVQRPADDEWKPLFRSVFY AVSPKGSAEQRVGQDPLAPEQARALKASLAFQEYRIANVITNLRIKDASAELRKLTVDE KQSIYDQLVSPSSEDITWSDLCDFLGFKRSQKGVGSLTEDGEERISSRPPRLTSVQRIYE SDNKIRKPLVAWWKSASDNEHEAMIRLLSNTVDIDKVREDVAYASAIEFIDGLDDDAL TKLDSVDLPSGRAAYSVELTQKLTRQMLTTDDDLHEARKTLFNVTDTSWRPPADPIGEP LGNPVSDRVLKNVNRVLMNCQQRWGNPVSVNIEHVRSSFSVAFARKDKREYEKNNE KRSIFRSSLSQLRADEQMEKVRESLRLLEAIQRQNGQCLYCGRTTIFRTCEMDHIVP RKGVGSTNTRTNFAAVCAECNRMKSNTPFAIWARSEDAQTRGVSLAEAKKRVTMFTF NPKSYAPREVKAFAKQAVIARLQQTEDDAIDNRSIESVAWMADDELHRRIDWYFNAKQ YVNSASIDDAEAETMKTTVSVFQGRVTASARRAAGIEGKIHFQGSQSKTRLDRRHAV DASVIAMMNTAAAQTLMERESLRESQRLIGLMPGERSWKEYPYEGTSRYESFHLWLD NMDVLELLELNDALDNDRIA VMQSQRVVLGNSIAHDATIHPLEKVPLGSAMSADLIRRA STPALWCALTRLPDYDEKEGLPEDSHREIRVHDTRY SADDEMGFFASQAAQIAVQEGS ADIGSAIHHARVYRCWKTNAKGVRKYFYGMIRVFQTDLLRACHDDLFTVPLPPQSISM RYGEPRVVQALQSGNAQYLGLSLVVGDEIEMDFSSLDVDGQIGEYLFQFSQFSGGNLAW KHWVVDGFFNQTLRIRPRYLA AEGLAKAFSDDVVPDGVQKIVTKQGWLPVNTASK TAVRIVRRNAFGEPRLSAHMPCSWQWRHE (SEQ ID NO: 331)
<i>A. muciniphila</i> Cas9	MSRSLTFSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRLRR NIRRRVRIRIGRLLVQAQIITPEMKETSGHPAPFYLA SEALKGHR TLAPIELWHVLRW YAHNRGYDNNASWSNSLSEDGGNGEDTERVKHAQDLM DKHGTATMAETICRELKLE EGKADAPMEVSTPAYKNLNTAFPR LIVEKEVRRILELSAPLIPGLTAEIIEIAQHPLTT EQRGVLLQHGIKLARRYRGSLLFQGLIPRFDNR IISRC PVTWAQVYEAELKKGNSEQSA RERA EKLSKVPTANCPEFYEYRMARILCNIRADGEPLSAEIRRELMNQARQEGKLT KAS LEKAISSRLGKETETNVSNYFTLHPDSEEALYLNPAVEVLQRSGIGQILSPSVYRI AANR LRRGKSVTPNYLLNLLKSRGESGEALEKKIEKESKKKEADYADTLPKPKYATGRAPYA RTVLKKVVEEILDGEDPTRPARGEAHPDGELKAHDGCLYCLLDTSSVNQH QKERRL DTMTNNHLVRHRMLILDRLLKDLIQDFADGQKDRISRVCEVGKELTTFSAMD SKKIQ RELTLRQKSHTDAVNRLKRKLPGKALSANLIRK CRIAMDMNWTCPFTGATYGDHELE NLELEHIVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVEQE QENPVPDKPNLHIC SLNNYRELVEKLDDKKGHEDDRRRKKKRRKALLMVRGLSHKHQSQNHEAMKEIGMTE GMMTQSSHLMKLACKSIKTS LPAHIDMIPGAVTA EVRKA WDVFVFKELCPEAADP DSGKILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLLRRV LAMRIPEKLIPQVRPV ANQRHYVLNDDGRMMLRDL SASLKENIREQLMEQRVIQHVPADMGGALLKETMQRV LSVDGSGEDAMVSLSKKKGKKEKNQVKASKLVGVFPEGPSK LKALKAAIEIDGNYG VALDPKPVVIRHIKVFKRIMALKEQNGGKPVRLKKGMLIHLTSSKDPKHAGVWRIESI QDSKGGVKLDLQRAHCAVPKNKTHECNWREVDLISLLKKYQMKRYPTSYTGTPR (SEQ ID NO: 332)

<i>O. laneus</i> Cas9	METTLGIDLGTSNIGLALVDQEEHQILYSGVRIFPEGINKDTIGLGEKEESRNATRRAKR QMRQYFRKKLRKAKLLELLIAYDMCPLKPEDVRRWKNWDKQQKSTVRQFPDTPAF REWLKQNPYELRKQAVTEDVTRPELGRILYQMIQRRGFLSSRKGKEEGKIFTGKDRMV GIDETRKNLQKQTLGAYLYDIAPKNGEKYRFRTERVRARYTLRDMYIREFEIHWQRQA GHLGLAHEQATRKKNFLEGSATNVRNSKLITHLQAKYGRGHVLIEDTRITVTFQLPLK EVLGGKIEIEEEQLKFKSNESVLFWQRPLRSQKSLLSKCVFEGRNFYDPVHQKWIIAGPT PAPLSHPEFEEFRAYQFINNIYGKNEHLTAIQREAVFELMCTESKDFNFEEKIPKHLKLFE KFNFD DTTKVPACTTISQLRKLFPHPVWEEKREEIWHCFYFYDDNTLLFEKLQKDYAL QTNDLEKIKKIRLSESYGNVSLKAIRRNIPYLKKGYAYSTAVLLGGIRNSFGKRFEYFKE YEPEIEKAVCRILKEKNAEGEVIRKIKDYL VHNRFGFAKNDRAFKLYHHSQAITTAQ KERLPETGNLRNPVQQGLNELRRTVNKLLATCREKYGPSFKFDHIHVEMGRELRSSKT EREKQSRQIRENEKKNEAAKVKLAEYGLKAYRDNQKYLKYKEIEEKGTVCCPYTGK TLNISHTLGSDNSVQIEHIIPYSISLDDSLANKTLCDATFNREK GELTPYDFYQKDP WGASSWEEIEDRAFRLLPYAKAQRFRIRK PQESNEFISRLNDTRYISKKA VEYLSAICS DVKAFFGQLTAE LRHLWGLNNILQSAPDITFPLPVSATENHREYYVITNEQNEVIRLFPK QGETPRTEK GELLLTGEVERK VFRCKGMQEFQTDVSDGKYWRIKLSSSVTWSPLFAP KPISADGQIVLK GRIEKG VFCNQLKQKLKTGLPDGSYWISLPVISQTFKEGESVNNSKL TSQQVQLFGRVREGIFRCHNYQCPASGADGNFWCTLDTDTAQPAFTPIKNAPPGVGGG QIILTGDVDDKGIFHADDLHYELPASLPKGKYYGIFTVESCDPTLPIEL SAPKTSKGEN LIEGNIWVDEHTGEVRFDPKKNREDQRHHAIDAIVIALSSQSFLQRLSTYNARRENKKR GLDSTEHPSPWPGFAQDVRSVVP LLVSYKQNP KTLCKISKTLKYDGKKIHS CGNAV RGQLHKETVYGQRTAPGATEKSYHIRKDIREL KTSKHIGKVVDITIRQMLLKH LQENY HIDITQEFNIPSNAFFKEGVYRIFLPNKHGEPVPIKKIRMKEELGNAERLKDNIQYVNP RNNHHVMYQDADGNLKEEIVSFWSVIERQNGQPIYQLPREGRNIVSILQINDTFLIGL KEEEPEVYRNDLSTLSKHL YRVQKLSGMYTFRHHLASTLNNERE EFRIQSLEAWKRA NPVKVQIDEIGRITFLNGPLC (SEQ ID NO: 333)
-----------------------	--

- Nuclease deficient *S. pyogenes* Cas9 proteins may comprise a substitution of an Alanine (A) for an Aspartic Acid (D) at position 10 and an alanine (A) for a Histidine (H) at position 840. Exemplary nuclease deficient *S. pyogenes* Cas9 proteins of the disclosure may
- 5 comprise or consist of the amino acid sequence (D10A and H840A bolded and underlined):
- 1 MDKKYSIGL**A**IGTNSVGWAV ITDEYKVPSK KFKVLGNTDR HSIKKNLIGA LLFDSGETAE
61 ATRLKRTARR RYTRRKNRIC YLQEISNEM AKVDDSFHHR LEESFLVEED KKHERHPIFG
121 NIVDEVAYHE KYPTIYHLRK KLVDSTDKAD LRLIYLALAH MIKFRGHFLI EGDLPNDNSD
181 VDKLFIQLVQ TYNQLFEENP INASGVDAKA ILSARLSKSR RLENLIAQLP GEKKNGLFGN
10 241 LIALSLGLTP NFKSNFDLAE DAKLQLSKDT YDDDLNLLA QIGDQYADLF LAAKNLSDAI
301 LLSDILRVNT EITKAPLSAS MIKRYDEHHQ DLTLLKALVR QQLPEKYKEI FFDQSKNGYA
361 GYIDGGASQE EFKYFIKPIE EKMDGTEELL VKLNREDLLR KQRTFDNGSI PHQIHLGELH
421 AILRRQEDFY PFLKDNREKI EKILTFRIPI YVGPLARGNS RFAWMTRKSE ETITPWNFEE
481 VVDKGASASQ FIERMTNFDK NLPNEKVLPK HSLLYEYFTV YNELTKVKYV TEGMRKPAFL
15 541 SGEQKKAIVD LLFKTNRKVT VKQLKEDYFK KIECFDSVEI SGVEDRFNAS LGTYHDLLKI
601 IKDKDFLDNE ENEDILEDIV LTLTLFEDRE MIEERLKTYA HLFDDKVMKQ LKRRRYTGWG
661 RLSRKLINGI RDKQSGKTIL DFLKSDGFAN RNFMQLIHDD SLTFKEDIQK AQVSGQGDSL
721 HEHIANLAGS PAIKKGILQT VKVVDLVKV MGRHKPENIV IEMARENQTT QKGQKNSRER
781 MKRIEEGIKE LGSQILKEHP VENTQLQNEK LYLYYLQNGR DMYVDQELDI NRLSDYDVDA**A**
20 841 IVPQSFLKDD SIDNKVLTRS DKNRGKSDNV PSEEVVKKMK NYWRQLLNAK LITQRKFDNL
901 TKAERGGLSE LDKAGFIKRQ LVETRQITKH VAQILDSRMN TKYDENDKLI REVKVITLKS
961 KLVSDFRKDF QFYKVREINN YHHAHDAYLN AVVG TALIKK YPKLESEFVY GDYKVYDVVRK

1021 MIAKSEQEIG KATAKYFFYS NIMNFFKTEI TLANGEIRKR PLIETNGETG EIVWDKGRDF
 1081 ATVRKVL SMP QVNIVKKTEV QTGGFSKESI LPKRNSDKLI ARKKDWDPKK YGGFDSPTVA
 1141 YSVLVVAKVE KGKSKKLKSV KELLGITIME RSSFEKNPID FLEAKGYKEV KKDILIKLPK
 1201 YSLFELENGR KRMLASAGEL QKGNELALPS KYVNFLYLAS HYEKLKGSPE DNEQKQLFVE
 5 1261 QHKHYLDEII EQISEFSKRV ILADANLDKV LSAYNKHRDK PIREQAENII HLFTLTNLGA
 PAAFKYFDTT IDRKYRTSTK EVLDATLIHQ SITGLYETRI DLSQLGGD (SEQ ID NO: 334).

Exemplary wild type *Francisella tularensis subsp. Novicida* Cpf1 (FnCpf1) proteins of the disclosure may comprise or consist of the amino acid sequence:

10 1 MSIQEFVNK YSLSKTLRFE LIPQGKTLEN IKARGLILDD EKRAKDYKKA KQIIDKYHQF
 61 FIEEILSSVC ISEDLLQNY S DVYFKLKKSD DDNLQKDFKS AKDTIKKQIS EYIKDSEKFK
 121 NLENQNLIDA KKGQESDLIL WLKQSKDNGI ELFKANS DIT DIDEALEI IK SFKGWTTYFK
 181 GFHENRKNVY SSNDIPT SII YRIVDDNL PK FLENKAKYES LKDKAPEAIN YEQIKKDLAE
 241 ELTFDIDYKT SEVNQRV FSL DEVFEIANFN NYLNQSGITK FNTIIGGKFV NGENTKRKGI
 15 301 NEYINLYSQQ INDKTLKKYK MSVLFKQILS DTESKSFVID KLEDDSDVVT TMQSFYEQIA
 361 AFKTVEEKSI KETLSLLFDD LKAQKLDLSK IYFKNDKSLT DLSQQVFDDY SVIGTAVLEY
 421 ITQQIAPKNL DNPSKKEQEL IAKKTEKAKY LSLETIKLAL EEFNKHRDID KQCRFEEILA
 481 NFAAIPMIFD EIAQNKD NLA QISIKYQNGG KKDLLQASAE DDVKA IKDL DQTNLLHLKL
 541 KIFHISQSED KANILDKDEH FYLVFEECYF ELANIVPLYN KIRNYITQKP YSDEKFKLNF
 20 601 ENSTLANGWD KNKEPDNTAI LFIKDDKYLL GVMNKKNNKI FDDKAIKENK GEGYKKIVYK
 661 LLPGANKMLP KVFFSAKSIK FYNPSEDILR IRNHSTHTKN GSPQKGYEKF EFNIEDCRKF
 721 IDFYKQSISK HPEWKDFGFR FSDTQRYNSI DEFYREVENQ GYKLTFENIS ESYIDSVVNQ
 781 GKLYLFQIYN KDFSAYS KGR PNLHTLYWKA LFDERNLQDV VYKLNGEAE L FYRKQSI PKK
 841 ITHPAKEAIA NKNKDNPKKE SVFEYDLIKD KRFTEDKFFF HCPITINFKS SGANKFNDEI
 25 901 NLLLKEKAND VHILSIDRGE RHLAYYTLVD GKGNI IKQDT FNIIGNDRMK TNYHDKLAAI
 961 EKDRDSARKD WKKINNIKEM KEGYLSQVVH EIAKL VIEYN AIVVFEDLNF GFKRGRFKVE
 1021 KQVYQKLEKM LIEKLNLYLV KDNEFDKTGG VLRAYQLTAP FETFKKMGKQ TGIIYYVPAG
 1081 FTSKICPVTG FVNQLYPKYE SVSKSQEFFS KFDKICYNLD KGYFEFSFDY KNFGDKAAKG
 1141 KWTIASFGSR LINFRNSDKN HNWDTREVPY TKELEKLLKD YSIEYGHGEC IKA AICGESD
 30 1201 KKFFAKLTSV LNTILQMRNS KTGT ELDYLI SPVADVNGNF FDSRQAPKNM PQDADANGAY
 1261 HIGLKGLMLL GRIKNNQEGK KLNLVIKNEE YFEFVQNRNN (SEQ ID NO: 335)

Exemplary wild type *Lachnospiraceae bacterium sp. ND2006* Cpf1 (LbCpf1) proteins of the disclosure may comprise or consist of the amino acid sequence:

35 1 AASKLEKFTN CYLSKTLRF KAIPVGKTQE NIDNKRLIVE DEKRAEDYKG VKKLLDRYYL
 61 SFINDVLHSI KLKLN NYIS LFRKKT RTEK ENKELENLEI NLRKEIAKAF KGAAGYKSLF
 121 KKDIIETILP EAADDKDEIA LVNSFNGFTT AFTGFFDNRE NMFSEEAKST SIAFRCINEN
 181 LTRYISNMDI FEKVDAIFDK HEVQEIKEKI LNSDYDVEDF FEGEFFNFVL TQEGIDVYNA
 241 IIGGFVTESG EKI KGLNEYI NLYNAKTKQA LPKFKPLYKQ VLSDRESLSF YGEGYTSDEE
 40 301 VLEVFRNTLN KNSEIFSSIK KLEKLFKNFD EYSSAGIFVK NGPAISTISK DIFGEWN LIR

361 DKWNAEYDDI HLKKKAVVTE KYEDDRRKSF KKIGSFSLEQ LQEYADADLS VVEKLKEIII
 421 QKVDEIYKVY GSSEKLFDDAD FVLEKSLKKN DAVVAIMKDL LDSVKSFENY IKAFFGEGKE
 481 TNRDESFYGD FVLAYDILLK VDHIYDAIRN YVTQKPYSKD KFKLYFQNPQ FMGGWDKDK
 541 TDYRATILRY GSKYYLAIMD KKYAKCLQKI DKDDVNGNYE KINYKLLPGP NKMLPKVFFS
 5 601 KKWMAYYNPS EDIQKIYKNG TFKKGDMFNL NDCHKLIDFF KDSISRYPKW SNAYDFNFSE
 661 TEKYKDIAGF YREVEEQGYK VSFESASKKE VDKLVEEGKL YMFQIYNKDF SDKSHGTPNL
 721 HTMYFKLLFD ENNHGQIRLS GGAELEFMRA SLKKEELVVH PANSPIANKN PDNPKKTTTL
 781 SYDVYKDKRF SEDQYELHIP IAINKCPKNI FKINTEVRVL LKHDDNPYVI GIDRGERNLL
 841 YIVVVDGKGN IVEQYSLNEI INNENGIKRIK TDYHSLDDK EKERFEARQN WTSIENIKEL
 10 901 KAGYISQVNH KICELVEKYD AVIALEDLNS GFKNRVRKVE KQVYQKFEKM LIDKLNMYMD
 961 KKSNPCATGG ALKGYQITNK FESFKSMSTQ NGFIFYIPAW LTSKIDPSTG FVNLLKTKYT
 1021 SIADSKKFIS SFDRIYVPE EDLFEFALDY KNFSRTDADY IKKWKLYSYG NRIRIFAAAK
 1081 KNNVFAWEEV CLTSAYKELF NKYGINYQQG DIRALLCEQS DKAFYSSFMA LMSLMLQMRN
 1141 SITGRDQVDF LISPVKNSDG IFYDSRNYEA QENAILPKNA DANGAYNIAR KVLWAIGQFK
 15 1201 KAEDEKLDKV KIAISNKEWL EYAQTSVK (SEQ ID NO: 336)

Exemplary wild type *Acidaminococcus* sp. *BV3L6* Cpf1 (AsCpf1) proteins of the disclosure may comprise or consist of the amino acid sequence:

1 MTQFEGFTNL YQVSKTLRFE LIPQGKTLKH IQEQGFIEED KARNHDHYKEL KPIIDRIYKT
 20 61 YADQCLQLVQ LDWENLSAAI DSYRKEKTEE TRNALIEEQA TYRNAIHDF IGRTDNLTD
 121 INKRHAIEYK GLFKAELFNG KVLKQLGTVT TTEHENALLR SFDKFTTYFS GFYENRKNVF
 181 SAEDISTAIP HRIVQDNFPK FKENCHIFTR LITAVPSLRE HFENVKKAIG IFVSTSIEEV
 241 FSFPFYNQLL TQTQIDLYNQ LLGGISREAG TEKIKGLNEV LNLAIQKND TAHIIASLPH
 301 RFIPLFKQIL SDRNTLSFIL EEFKSDEEVI QSFCKYKTL RNENVLETAE ALFNELNSID
 25 361 LTHIFISHKK LETISSALCD HWDTLRNALY ERRISELTGK ITKSAKEKVQ RSLKHEDINL
 421 QEIIISAAGKE LSEAFKQKTS EILSHAAAL DQPLPTTLKK QEEKEILKSQ LDSLLGLYHL
 481 LDWFAVDESN EVDPEFSARL TGIKLEMEPS LSFYNKARNY ATKKPYSVEK FKLNFQMPTL
 541 ASGWDVNKEK NNGAILFVKN GLYYLGIMPK QKGRYKALSF EPTEKTSEGF DKMYDYDFPD
 601 AAKMIPKCST QLKAVTAHFQ THTTPILLSN NFIEPLEITK EIYDLNNEK EPKKFQTAYA
 30 661 KKTGDQKGYR EALCKWIDFT RDFLSKYTKT TSIDLSSLRP SSQYKDLGEY YAEINPLLYH
 721 ISFORIAEKE IMDAVETGKL YLFQIYNKDF AKGHHGKPNL HTLYWTGLFS PENLAKTSIK
 781 LNGQAELFYR PKSRMKMAH RLGEKMLNKK LKDQKTPIPD TLYQELYDYV NHRLSHDLSD
 841 EARAALPNVI TKEVSHEIIK DRRFTSDKFF FHVPITLNYQ AANSPSKFNQ RVNAYLKEHP
 901 ETPIIGIDRG ERNLIYITVI DSTGKILEQR SLNTIQQFDY QKKLDNREKE RVAARQAWSV
 35 961 VGTIKDLKQG YLSQVIHEIV DLMIHQAVV VLENLNFQFK SKRTGIAEKA VYQQFEKMLI
 1021 DKLNLCLVLKD YPAEKVGGVL NPYQLTDQFT SFAKMGTSQS FLFYVPAPYT SKIDPLTGTV
 1081 DPFVWKTIKN HESRKHFLEG FDFLHYDVKT GDFILHFKMN RNLSFQRGLP GEMPAWDIVF
 1141 EKNETQFQAK GTPFIAGKRI VPVIENHRFT GRYRDLYPAN ELIALLEEKG IVFRDGSNII
 1201 PKLLENDSSH AIDTMVALIR SVLQMRNSNA ATGEDYINSP VRDLNGVCDF SRFQNPPEWPM

1261 DADANGAYHI ALKGQLLLNH LKESKDLKLQ NGISNQDWLA YIQELRN (SEQ ID NO: 337)

In some embodiments of the compositions of the disclosure, the sequence encoding the RNA binding protein comprises a sequence isolated or derived from a CRISPR Cas protein or RNA-binding portion thereof. In some embodiments, the CRISPR Cas protein comprises a Type VI CRISPR Cas protein. In some embodiments, the Type VI CRISPR Cas protein comprises a Cas13 protein. Exemplary Cas13 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, bacteria or archaea. Exemplary Cas13 proteins of the disclosure may be isolated or derived from any species, including, but not limited to, *Leptotrichia wadei*, *Listeria seeligeri* serovar 1/2b (strain ATCC 35967 / DSM 20751 / CIP 100100 / SLCC 3954), *Lachnospiraceae* bacterium, *Clostridium aminophilum* DSM 10710, *Carnobacterium gallinarum* DSM 4847, *Paludibacter propionisigenes* WB4, *Listeria weihenstephanensis* FSL R9-0317, *Listeria weihenstephanensis* FSL R9-0317, bacterium FSL M6-0635 (*Listeria newyorkensis*), *Leptotrichia wadei* F0279, *Rhodobacter capsulatus* SB 1003, *Rhodobacter capsulatus* R121, *Rhodobacter capsulatus* DE442 and *Corynebacterium ulcerans*. Exemplary Cas13 proteins of the disclosure may be DNA nuclease inactivated. Exemplary Cas13 proteins of the disclosure include, but are not limited to, Cas13a, Cas13b, Cas13c, Cas13d, and orthologs thereof. Exemplary Cas13b proteins of the disclosure include, but are not limited to, subtypes 1 and 2 referred to herein as Csx27 and Csx28, respectively. Exemplary Cas13a proteins include, but are not limited to:

Cas13a number	Cas13a abbreviation	Organism name	Accession number
Cas13a1	LshCas13a	<i>Leptotrichia shahii</i>	WP_018451595.1 (SEQ ID NO: 338)
Cas13a2	LwaCas13a	<i>Leptotrichia wadei</i>	WP_021746774.1 (SEQ ID NO: 339)
Cas13a3	LseCas13a	<i>Listeria seeligeri</i>	WP_012985477.1 (SEQ ID NO: 340)
Cas13a4	LbmCas13a	<i>Lachnospiraceae</i> bacterium MA2020	WP_044921188.1 (SEQ ID NO: 341)
Cas13a5	LbnCas13a	<i>Lachnospiraceae</i> bacterium NK4A179	WP_022785443.1 (SEQ ID NO: 342)
Cas13a6	CamCas13a	[<i>Clostridium</i>] <i>aminophilum</i> DSM 10710	WP_031473346.1 (SEQ ID NO: 343)

Cas13a7	CgaCas13a	Carnobacterium gallinarum DSM 4847	WP_034560163.1 (SEQ ID NO: 344)
Cas13a8	Cga2Cas13a	Carnobacterium gallinarum DSM 4847	WP_034563842.1 (SEQ ID NO: 345)
Cas13a9	Pprcas13a	Paludibacter propionigenes WB4	WP_013443710.1 (SEQ ID NO: 346)
Cas13a10	LweCas13a	Listeria weihenstephanensis FSL R9-0317	WP_036059185.1 (SEQ ID NO: 347)
Cas13a11	LbfCas13a	Listeriaceae bacterium FSL M6-0635 (Listeria newyorkensis)	WP_036091002.1 (SEQ ID NO: 348)
Cas13a12	Lwa2cas13a	Leptotrichia wadei F0279	WP_021746774.1 (SEQ ID NO: 349)
Cas13a13	RcsCas13a	Rhodobacter capsulatus SB 1003	WP_013067728.1 (SEQ ID NO: 350)
Cas13a14	RcrCas13a	Rhodobacter capsulatus R121	WP_023911507.1 (SEQ ID NO: 351)
Cas13a15	RcdCas13a	Rhodobacter capsulatus DE442	WP_023911507.1 (SEQ ID NO: 352)

Exemplary wild type Cas13a proteins of the disclosure may comprise or consist of the amino acid sequence:

1 MGNLFGHKRW YEVRDKKDFK IKRKVKVKN YDGNKYILNI NENNNKEKID NNFIRKYIN
 5 61 YKKNNDNLKE FTRKFHAGNI LFKLKGKEGI IRIENNDDFL ETEEVVLYIE AYGKSEKLKA
 121 LGITKKKIID EAIRQGITKD DKKIEIKRQE NEEIEIDIR DEYTNKTLND CSIILRIEN
 181 DELETKKSIY EIFKNINMSL YKIEKIIEN ETEKVFENRY YEEHLREKLL KDDKIDVILT
 241 NFMEIREKIK SNLEILGFVK FYLNVGGDKK KSKNKKMLVE KILNINVDLT VEDIADFKIK
 301 ELEFWNITKR IEKVKKVNE FLEKRRNRTY IKSIVLLDKH EKFKIERENK KDKIVKFFVE
 10 361 NIKNNSIEK IEKILAEFKI DELIKKLEKE LKKGNCDEI FGIFKKHYKV NFDSKKFSKK
 421 SDEEKELYKI IYRYLKGRIE KILVNEQKVR LKMEKIEIE KILNESILSE KILKRVKQYT
 481 LEHIMYLGKL RHNDIDMTTV NTDDFSRLHA KEELDLELIT FFASTNMELN KIFSRENINN
 541 DENIDFFGGD REKNYVLDK ILNSKIKIIR DLDFIDNKN ITNNFIRKFT KIGTNERNRI
 601 LHAISKERDL QGTQDDYNKV INIIQNLKIS DEEVSKALNL DVVFKDKNI ITKINDIKIS
 15 661 EENNNDIKYL PSFSKVLPEI LNLYRNPNK EPFDTIETEK IVLNALIYVN KELYKKLILE
 721 DDLEENESKN IFLQELKCTL GNIDEIDENI IENYKNAQI SASKGNKAI KKYQKKVIEC

781 YIGYLKKNYE ELFDSDFKM NIQEIKKQIK DINDNKTYER ITVKTSDKTI VINDDFEYII
 841 SIFALLNSNA VINKIRNRFF ATSVWLNTSE YQNIIDILDE IMQLNTRLNE CITENWNLNL
 901 EEFIQMKKEI EKDFDDFKIQ TKKEIFNNYY EDIKNNILTE FKDDINGCDV LEKKLEKIVI
 961 FDDETKFEID KKSNIQDEQ RKLSNINKKD LKKKVDQYIK DKDQEIKSKEI LCRIIFNSDF
 5 1021 LKKYKKEIDN LIEDMESENE NKFQEIYYPK ERKNELYIYK KNLFLNIGNP NFDKIYGLIS
 1081 NDIKMADAKF LFNIDGKNIR KKKISEIDAI LKNLNDKLNG YSKEYKEYI KKLKENDDFE
 1141 AKNIQNKNYK SFEKDYNRVS EYKKIRDLVE FNYLNKIESY LIDINWKLAI QMARFERDMH
 1201 YIVNGLRELG IIKLSGYNTG ISRAYPKRNG SDGFYTTTAY YKFFDEESYK KFEKICYGFG
 1261 IDLSENSEIN KPENESIRNY ISHFYIVRNP EADYSIAEQI DRVSNLLSYS TRYNNSTYAS
 10 1321 VFEVFKKDVN LDYDELKKKF KLIGNNDILE RLMKPKKVSF LELESYNSDY IKNLIIELLT
 1381 KIENTNDTL (SEQ ID NO: 353)

Exemplary Cas13b proteins include, but are not limited to:

Species	Cas13b Accession	Cas13b Size (aa)
Paludibacter propionigenes WB4	WP_013446107.1 (SEQ ID NO: 354)	1155
Prevotella sp. P5-60	WP_044074780.1 (SEQ ID NO: 355)	1091
Prevotella sp. P4-76	WP_044072147.1 (SEQ ID NO: 356)	1091
Prevotella sp. P5-125	WP_044065294.1 (SEQ ID NO: 357)	1091
Prevotella sp. P5-119	WP_042518169.1 (SEQ ID NO: 358)	1091
Capnocytophaga canimorsus Cc5	WP_013997271.1 (SEQ ID NO: 359)	1200
Phaeodactylibacter xiamenensis	WP_044218239.1 (SEQ ID NO: 360)	1132
Porphyromonas gingivalis W83	WP_005873511.1 (SEQ ID NO: 361)	1136
Porphyromonas gingivalis F0570	WP_021665475.1 (SEQ ID NO: 362)	1136
Porphyromonas gingivalis ATCC 33277	WP_012458151.1 (SEQ ID NO: 363)	1136
Porphyromonas gingivalis F0185	ERJ81987.1 (SEQ ID NO: 364)	1136

Porphyromonas gingivalis F0185	WP_021677657.1 (SEQ ID NO: 365)	1136
Porphyromonas gingivalis SJD2	WP_023846767.1 (SEQ ID NO: 366)	1136
Porphyromonas gingivalis F0568	ERJ65637.1 (SEQ ID NO: 367)	1136
Porphyromonas gingivalis W4087	ERJ87335.1 (SEQ ID NO: 368)	1136
Porphyromonas gingivalis W4087	WP_021680012.1 (SEQ ID NO: 369)	1136
Porphyromonas gingivalis F0568	WP_021663197.1 (SEQ ID NO: 370)	1136
Porphyromonas gingivalis	WP_061156637.1 (SEQ ID NO: 371)	1136
Porphyromonas gulae	WP_039445055.1 (SEQ ID NO: 372)	1136
Bacteroides pyogenes F0041	ERI81700.1 (SEQ ID NO: 373)	1116
Bacteroides pyogenes JCM 10003	WP_034542281.1 (SEQ ID NO: 374)	1116
Alistipes sp. ZOR0009	WP_047447901.1 (SEQ ID NO: 375)	954
Flavobacterium branchiophilum FL-15	WP_014084666.1 (SEQ ID NO: 376)	1151
Prevotella sp. MA2016	WP_036929175.1 (SEQ ID NO: 377)	1323
Myroides odoratimimus CCUG 10230	EHO06562.1 (SEQ ID NO: 378)	1160
Myroides odoratimimus CCUG 3837	EKB06014.1 (SEQ ID NO: 379)	1158
Myroides odoratimimus CCUG 3837	WP_006265509.1 (SEQ ID NO: 380)	1158
Myroides odoratimimus CCUG 12901	WP_006261414.1 (SEQ ID NO: 381)	1158
Myroides odoratimimus CCUG 12901	EHO08761.1 (SEQ ID NO: 382)	1158

Myroides odoratimimus (NZ_CP013690.1)	WP_058700060.1 (SEQ ID NO: 383)	1160
Bergeyella zoohelcum ATCC 43767	EKB54193.1 (SEQ ID NO: 384)	1225
Capnocytophaga cynodegmi	WP_041989581.1 (SEQ ID NO: 385)	1219
Bergeyella zoohelcum ATCC 43767	WP_002664492.1 (SEQ ID NO: 386)	1225
Flavobacterium sp. 316	WP_045968377.1 (SEQ ID NO: 387)	1156
Psychroflexus torquis ATCC 700755	WP_015024765.1 (SEQ ID NO: 388)	1146
Flavobacterium columnare ATCC 49512	WP_014165541.1 (SEQ ID NO: 389)	1180
Flavobacterium columnare	WP_060381855.1 (SEQ ID NO: 390)	1214
Flavobacterium columnare	WP_063744070.1 (SEQ ID NO: 391)	1214
Flavobacterium columnare	WP_065213424.1 (SEQ ID NO: 392)	1215
Chryseobacterium sp. YR477	WP_047431796.1 (SEQ ID NO: 393)	1146
Riemerella anatipestifer ATCC 11845 = DSM 15868	WP_004919755.1 (SEQ ID NO: 394)	1096
Riemerella anatipestifer RA-CH-2	WP_015345620.1(SEQ ID NO: 395)	949
Riemerella anatipestifer	WP_049354263.1 (SEQ ID NO: 396)	949
Riemerella anatipestifer	WP_061710138.1 (SEQ ID NO: 397)	951
Riemerella anatipestifer	WP_064970887.1 (SEQ ID NO: 398)	1096
Prevotella saccharolytica F0055	EKY00089.1 (SEQ ID NO: 399)	1151
Prevotella saccharolytica JCM 17484	WP_051522484.1 (SEQ ID NO: 400)	1152

Prevotella buccae ATCC 33574	EFU31981.1 (SEQ ID NO: 401)	1128
Prevotella buccae ATCC 33574	WP_004343973.1 (SEQ ID NO: 402)	1128
Prevotella buccae D17	WP_004343581.1 (SEQ ID NO: 403)	1128
Prevotella sp. MSX73	WP_007412163.1 (SEQ ID NO: 404)	1128
Prevotella pallens ATCC 700821	EGQ18444.1 (SEQ ID NO:405)	1126
Prevotella pallens ATCC 700821	WP_006044833.1 (SEQ ID NO:406)	1126
Prevotella intermedia ATCC 25611 = DSM 20706	WP_036860899.1 (SEQ ID NO:407)	1127
Prevotella intermedia	WP_061868553.1 (SEQ ID NO:408)	1121
Prevotella intermedia 17	AFJ07523.1 (SEQ ID NO:409)	1135
Prevotella intermedia	WP_050955369.1 (SEQ ID NO: 410)	1133
Prevotella intermedia	BAU18623.1 (SEQ ID NO:411)	1134
Prevotella intermedia ZT	KJJ86756.1 (SEQ ID NO:412)	1126
Prevotella aurantiaca JCM 15754	WP_025000926.1 (SEQ ID NO: 413)	1125
Prevotella pleuritidis F0068	WP_021584635.1(SEQ ID NO: 414)	1140
Prevotella pleuritidis JCM 14110	WP_036931485.1(SEQ ID NO: 415)	1117
Prevotella falsenii DSM 22864 = JCM 15124	WP_036884929.1 (SEQ ID NO: 416)	1134
Porphyromonas gulae	WP_039418912.1 (SEQ ID NO: 417)	1176
Porphyromonas sp. COT-052 OH4946	WP_039428968.1 (SEQ ID NO: 418)	1176

Porphyromonas gulae	WP_039442171.1 (SEQ ID NO: 419)	1175
Porphyromonas gulae	WP_039431778.1 (SEQ ID NO: 420)	1176
Porphyromonas gulae	WP_046201018.1 (SEQ ID NO: 421)	1176
Porphyromonas gulae	WP_039434803.1 (SEQ ID NO: 422)	1176
Porphyromonas gulae	WP_039419792.1 (SEQ ID NO: 423)	1120
Porphyromonas gulae	WP_039426176.1 (SEQ ID NO: 424)	1120
Porphyromonas gulae	WP_039437199.1 (SEQ ID NO: 425)	1120
Porphyromonas gingivalis TDC60	WP_013816155.1 (SEQ ID NO: 426)	1120
Porphyromonas gingivalis ATCC 33277	WP_012458414.1 (SEQ ID NO: 427)	1120
Porphyromonas gingivalis A7A1-28	WP_058019250.1 (SEQ ID NO: 428)	1176
Porphyromonas gingivalis JCVI SC001	EOA10535.1 (SEQ ID NO: 429)	1176
Porphyromonas gingivalis W50	WP_005874195.1 (SEQ ID NO: 430)	1176
Porphyromonas gingivalis	WP_052912312.1 (SEQ ID NO: 431)	1176
Porphyromonas gingivalis AJW4	WP_053444417.1 (SEQ ID NO: 432)	1120
Porphyromonas gingivalis	WP_039417390.1 (SEQ ID NO: 433)	1120
Porphyromonas gingivalis	WP_061156470.1 (SEQ ID NO: 434)	1120

Exemplary wild type *Bergeyella zoohelcum* ATCC 43767 Cas13b (BzCas13b)

proteins of the disclosure may comprise or consist of the amino acid sequence:

1 menktslgnn iyynpfkpqp ksyfagyfna amentdsvfr elgkrlkgxe ytsenffdai

61 fkenislvey eryvklldsy fpmarlldkk evpikerken fknfkgiik avrdlnrfyt
 121 hkehgeveit deifgvldem lkstvlvtvkk kkvktldktke ilkksiekql dilcqkkley
 181 lrdtarkiee krrnqrerge kelvapfkys dkrddliaai yndafdvaid kkkdskless
 241 kakyntksdp qqeegdlkip iskngvfl1 slfltkqeih afkskiagfk atvideatvs
 5 301 eatvshgkns icfmatheif shlaykklkr kvrtaeinyg eaenaeqlsv yaketlmmqm
 361 ldelskvpdv vyqnlsvdv ktfieldwney lkenngdvgt meeeqvihpv irkryedkfn
 421 yfairfldef aqfptlrfqv hlgnylhdsr pkenlisdr ikekitvfgr lselehkkal
 481 fikntetned rehyweifpn pnydfpkeni svndkdfpia gsildrekqp vagkigikvk
 541 llngqyvsev dkavkahqlk qrkaskpsiq niieivpin esnpkeavf gggptaylsm
 10 601 ndihsilyef fdkwekkkek lekkgekelr keigkelekk ivgkiqaiq qiidkdnak
 661 ilkpyqdgns taidkeklik dlkqeqnilq klkdeqtvre keyndfiayq dknreinkvr
 721 drnhkqylkd nlkrkypeap arkevlyyre kgkvavwlan dikrfmptdf knewkgeqhs
 781 llqkslayye qckeelknll pekvfqhlpf klgyfyqqky lyqfytcyld krleyisglv
 841 qqaenfksen kvfkkenec fkflkkqnyt hkeldarvqs ilgypifler gfmdekptii
 15 901 kgktfkgnea lfadwfryyk eyqnfqtfyd tenyplvele kkqadrkrxt kiyqqkndv
 961 ftllmakhif ksvfkqdsid qfsledlyqs reerlgnqer arqtgernn yiwnktvdlk
 1021 lcdgkitven vllknvgdfi kyeydqrva flkyeeniew qaflikeske eenypyvver
 1081 eieqyekvrr eellkevhli eeyilekvkd keilkkgdnq nfkyiylngl lkqlknedve
 1141 sykvfnlnte pedvninglk qeatdleqka fvltyirnkf ahnqlpkkef wdycqekyqk
 20 1201 ieeketyaey faevfkkeke alik (SEQ ID NO: 435)

An exemplary nuclease deficient Cas13b (dCas13b) nucleic acid sequence with C-terminal nuclear export sequence is:

ATGaacatccccgctctggtggaaaaccagaagaagtaactttggcacctacagcgtgatggccatgctgaacgct
 25 cagaccgtgctggaccacatccagaaggtggccgatattgagggcgagcagaacgagaacaacgagaatctgtgg
 tttcaccccgctgatgagccacctgtacaacgccaaagacggctacgacaagcagcccgagaaaaccatgttcac
 atcgagcggctgcagagctacttccattcctgaagatcatggccgagaaccagagagagtagacgaacggcaag
 tacaagcagaaccgctggaagtgaacagcaacgacatcttcgaggtgctgaagcgcgcccttcggcgtgctgaag
 atgtacaggacctgaccaacgcAtacaagacctacgaggaaaagctgaacgacggctgcgagttcctgaccagc
 30 acagagcaacctctgagcggcatgatcaacaactactacacagtgccctgcggaacatgaacgagagatacggc
 tacaagacagaggacctggccttcacccaggacaagcggttcaagtctgtgaaggacgcctacggcaagaaaaag
 tcccaagtgaataccggattcttcctgagcctgcaggactacaacggcgacacacagaagaagctgcacctgagc
 ggagtggaatcgccctgctgatctgcctgttcctggacaagcagtagatcaacatctttctgagcaggctgcc
 atcttctccagctacaatgccagagcgaggaacggcgatcatcatcagatccttcggcatcaacagcatcaag
 35 ctgccaagagaccggatccacagcgagaagtccaacaagagcgtggccatggatgatgctcaacgaagtgaagcgg
 tgccccgacgagctgttcacaacactgtctgccgagaagcagtcgggttcagaatcatcagcgacgaccacaat
 gaagtgtgatgaagcggagcagcgacagattcgtgcctctgctgctgcagtatatcgattacggcaagctgttc
 gaccacatcaggttccacgtgaacatgggcaagctgagatacctgctgaagggcgacaagacctgcacgagcgc
 cagaccagagtcagagtgatcgagcagccctgaacggccttcggcagactggaagaggccgagacaatgcggaag
 40 caagagaacggcaccttcggcaacagcggcatccggatcagagacctcgagaacatgaagcgggacgacgccaat

cctgccaaactatccctacatcgtaggacacctacacacactacatcctggaaaacaacaaggtagagatgtttatc
 aacgacaaagaggacagcgccccactgctgcccgtgatcgaggatgatagatacgtgggtcaagacaatccccagc
 tgccggatgagcaccctggaaattccagccatggccttcacatgtttctgttcggcagcaagaaaaccgagaag
 ctgatcgtaggacgtgcacaaccggtacaagagactgttccaggccatgcagaaagaagaagtgaccgccgagaat
 5 atcgccagcttcggaatcgccgagagcgacactgcctcagaagatcctggatctgatcagcggcaatgccacggc
 aaggatgtggacgccttcacagactgacgtggacgacatgctgaccgacaccgagcggagaatcaagagattc
 aaggacgaccggaagtccattcgagcgccgacaacaagatgggaaagagaggcttcaagcagatctccacaggc
 aagctggccgacttcctggccaaggacatcgtagctgtttcagcccagcgtgaacgatggcgagaacaagatcacc
 ggctgaactaccggatcatgcagagcgccattgcccgtgtacgatagcggcgacgattacgaggccaagcagcag
 10 ttcaagctgatgttcgagaaggcccgctgatcggcaagggcacaacagagcctcatccatttctgtacaagggtg
 ttgccccgcagcatccccgccaatgccgtcgagttctacgagcgctacctgatcgagcggaaagttctacctgacc
 ggctgtccaacgagatcaagaaaggcaacagagtggatgtgcccctcatccggcgggaccagaaacaagtggaaa
 acaccgcctgaagaccctgggcagaatctacagcgaggatctgcccgtggaaactgccagacagatgttcgac
 aatgagatcaagtccacctgaagtccctgccacagatggaaggcatcgacttcaacaatgccaacgtgacctat
 15 ctgatcgccgagtacatgaagagagtgtgtagcagcacttccagaccttctaccagtggaaaccgcaactaccgg
 tacatggacatgcttaagggcgagtacgacagaaagggtccctgcagcactgcttcaccagcgtggaagagaga
 gaaggcctctggaaagagcgggctccagaacagagcgggtacagaaagcaggccagcaacaagatccgagcaac
 cggcagatgagaaaagccagcagcgaagagatcgagacaatcctggataagcggctgagcaacagccggaacgag
 taccagaaaagcgagaaagtgatccggcgctacagagtgcaggatgcctgctgtttctgctggccaaaaagacc
 20 ctgaccgaactggccgatttcgacggcgagaggttcaaactgaaagaaatcatgcccgcgagcggagaagggaatc
 ctgagcgagatcatgccatgagcttcaccttcgagaaaggcggaagaagtacaccatcaccagcagggcatg
 aagctgaagaactacggcgacttctttgtgctggctagcgacaagaggatcggcaacctgctggaactcgtgggc
 agcgacatcgtgtccaaagaggatatcatggaagagttcaaaaaacgaccagtgcaggcccagatcagctcc
 atcgtgttcaacctggaaaagtgggcttcgacacataccccgagctgtctgcccagagtggaccgggaagagaag
 25 gtggacttcaagagcatcctgaaaatcctgctgaacaacaagaacatcaacaaagagcagagcgacatcctgcgg
 aagatccggaacgccttcgatgcAaacaattaccccgacaaaggcgtggtggaaatcaaggccctgcctgagatc
 gccatgagcatcaagaaggcctttggggagtacccatcatgaaggatccCTTCAACTGCCTCCACTTGAAGA
 CTGACACTGctcgagAGAGATTAG (SEQ ID NO: 436)

- 30 An exemplary nuclease deficient Cas13b (dCas13b) nucleic acid sequence with stop codon (making it an independent reading frame) is as follows:

ATGaacatccccgctctggtggaaaaccagaagaagtactttggcacctacagcgtgatggccatgctgaacgct
 cagaccgtgtctggaccacatccagaagggtggccgatatgtaggggcagcagaaacgagaacaacgagaatctgtgg
 35 tttcaccctgtgatgagccacctgtacaacgccaagaacggctacgacaagcagcccagaaaaaccatgttcatc
 atcgagcggctgcagagctacttccattcctgaagatcatggccgagaaccagagagagtacgcaacgggaag
 tacaagcagaaccgctggaagtgaacagcaacgacatcttcgaggtgctgaagcgcgcttccggcgtgctgaag
 atgtacagggacctgaccaacgcAtacaagacctacgaggaaaagctgaacgacggctgcgagttcctgaccagc
 acagagcaacctctgagcggcatgatcaacaactactacagtgccctgcggaacatgaacgagagatacggc
 tacaagacagaggacctggccttcacccaggacaagcgggttcaagttcgtgaaggacgcctacggcaagaaaaag
 40 tcccaagtgaataccggattcttctgagcctgcaggactacaacggcgacacacagaagaagtgcacctgagc

ggagtgggaatcgccctgctgatctgctgttcctggacaagcagacacatcaacatctttctgagcaggctgccc
 atcttctccagctacaatgccagagcgaggaacggcggtatcatcatcagatccttcggcatcaacagcatcaag
 ctgcccgaaggaccggatccacagcgagaagtccaacaagagcggtggccatggatatgctcaacgaagtgaagcgg
 tgcctcgacgagctgttcacaacactgtctgccgagaagcagtcctcggttcagaatcatcagcgacgaccacaat
 5 gaagtgctgatgaagcggagcagcgacagattcgtgctctgctgctgcagtatatcgattacggcaagctgttc
 gaccacatcaggttccacgtgaacatgggcaagctgagatacctgctgaaggccgacaagacctgcatcgacggc
 cagaccagagtcagagtgatcgagcagccctgaacggcttcggcagactggaagaggccgagacaatgcggaag
 caagagaacggcaccttcggcaacagcggcatccggatcagagacctcgagaacatgaagcgggacgacgccaat
 cctgccaaactatccctacatcgtggacacctacacacactacatcctggaaaacaacaaggctcgagatgtttatc
 10 aacgacaaagaggacagcgccccactgctgccctgctgatcgaggatgatagatacgtgggtcaagacaatcccagc
 tgcgggatgagcaccctggaaattccagccatggccttcacatgtttctgttcggcagcaagaaaaccgagaag
 ctgatcgtggacgtgcacaaccgggtacaagagactgttcaggccatgcagaaagaagaagtgaccgcccagagaat
 atcgccagcttcggaatcgccgagagcgacactgcctcagaagatcctggatctgatcagcggcaatgccacggc
 aaggatgtggacgccttcacagactgaccgtggacgacatgctgaccgacaccgagcggagaatcaagagattc
 15 aaggacgaccggaagtccattcgagcgcgcgacaacaagatgggaaagagaggcttcaagcagatctccacaggc
 aagctggccgacttcctggccaaggacatcgtgctgtttcagcccagcgtgaacgatggcgagaacaagatcacc
 ggctgaactaccggatcatgacagcgcattgctgtacgatagcggcgacgattacgaggccaagcagcag
 ttcaagctgatgttcgagaaggcccggtgatcggcaagggcacaacagagcctcatccatttctgtacaaggtg
 ttcgcccgcagcatccccgccaatgccgtcgagtctacgagcgcctacctgatcgagcggaaagttctacctgacc
 20 ggcctgtccaacgagatcaagaaaggcaacagagtggtatgtgcccttcacccggcgggaccagaacaagtggaaa
 acacccgccatgaagaccctgggcagaatctacagcgaggatctgccctggaaactgccagacagatgttcgac
 aatgagatcaagtcccacctgaagtccctgccacagatggaaggcatcgacttcaacaatgccaaactgacat
 ctgatcgccgagtacatgaagagagtgtgtggacgacgacttcagaccttctaccagtggaaaccgcaactaccgg
 tacatggacatgcttaagggcgagtacgacagaaagggctccctgcagcactgcttcaccagcgtggaagagaga
 25 gaaggcctctggaaagagcgggctccagaacagagcgggtacagaaagcaggccagcaacaagatccgagcaac
 cggcagatgagaaacgccagcagcgaagagatcgagacaatcctggataagcggctgagcaacagccggaacgag
 taccagaaaagcgagaaagtgatccggcgctacagagtgcaggatgcctgctgtttctgctggccaaaagacc
 ctgaccgaactggccgatttcgacggcgagaggttcaaactgaaagaaatcatgcccgacgccgagaagggaatc
 ctgagcagagatcatgccatgagcttcaccttcgagaaaggcggcaagaagtacaccatcaccagcagggcatg
 30 aagctgaagaactacggcgacttctttgtgctggctagcgacaagaggatcggaacctgctggaactcgtgggc
 agcgacatcgtgtccaaagaggatatcatggaagagttcaacaaaacgaccagtgcaggcccagatcagctcc
 atcgtgttcaacctggaaaagtgggccttcgacacataccccgagctgtctgccagagtggaccgggaagagaag
 gtggacttcaagagcatcctgaaaatcctgctgaacaacaagaacatcaacaaagagcagagcgacatcctgcgg
 aagatccggaacgccttcgatgcAaacaattacccgcacaaaggcgtggtggaaatcaaggccctgcctgagatc
 35 gccatgagcatcaagaaggcctttggggagtacgccatcatgaagTAG (SEQ ID NO: 437)

Exemplary wild type Cas13d proteins of the disclosure may comprise or consist of the amino acid sequences:

Cas13d (<i>Ruminococcus flavefaciens</i> XPD3002)	IEKKKSFAKGMGVKSTLVSGSKVYMTTFAEGSDARLEKIVEGDSI RSVNEGEAFSAEMADKNAGYKIGNAKF'SHPKGYAVVANNPLYTGP VQQDMLGLKETLEKRYFGESADGNDNICIQVIHNILDIEKILAEY
--	--

	ITNAAYAVNNISGLDKDIIGFGKFSTVYTYDEFKDP EHHRAAFNN NDKLINAIKAQYDEFDNFLDNPRLG YFGQAFFSKEGRNYIINYGN ECYDILALLSGLAHWV VANN EESRISRTWLYNLDKNLDNEYIST LNYLYDRITNELTNSFSKNSAANVNYIAETLGINPAEFAEQYFRF SIMKEQKNLGFNITKLREVM LDRKDMSEIRKNHKV FDSIRTKVYT MMDFVIYRY YIEEDAKVAAANKSLPDNEKSLSEKDI FVINLRGSF NDDQKDALYYDEANRIWRKLENIMHNIKEFRGNKTREYKKKDAPR LPRILPAGRDVS AF SKLMYALTMFLDGKEINDLLTTLINKFDNIQ SFLKVMPLIGVNAKFVEEYAFFKDSAKIADELRLIKS FARMGEPI ADARRAMYIDAIRILGTNLSYDELKALADTFSLDENGK LKKGKH GMRNFIINNVISNKR FHYLIRYGDPAHLHEIAKNEAVVKFVLGRI ADIQKKQGQNGKNQIDRY YETCIGKDKGKSVSEKVDALTKIITGM NYDQFDKKRSVIEDTGRENAEREKFKKIISLYLTVIYHILKNIVN INARYVIGFHCVERDAQLYKEKGYDINLKKLEEKGFSSVTKL CAG IDETAPDKRK DVEKEMAERAKESIDSLESANPKLYANYIKYSDEK KAEETFTRQINREKAKTALNAYLRNTKWNVI IREDLLRIDNKTCTL FANKA VALEVAR YVHAYINDIAEVNSYFQLYHYIMQRIIMNERYE KSSGKVSEYF DAVNDEKKYNDRLLKLLCVPFGYCI PRFKNLSIEA LFD RNEAAKF DKEKKS GNS (SEQ ID NO: 438)
Cas13d (contig e- k87_11092736)	MKRQKTFAKRIGIKSTVAYGQGYAITTFGKGSKAEIAVRSADPP EETLPTESDATLSIHAKFAKAGRDGREFKCGDVDETRIHTSRSEY ESLISNPAESPREDYLG LKGT LERKFFGDEYPKDNLR IQIIYSIL DIQKILGLYVEDILHFVDGLQDEPEDLVGLGLGDEKMQKLLSKAL PYMGFFGSTDVFKVTKKREERAAADEHNAKVFRALGAIRQKLAHF KWKESLAIFGANANMPIRFFQGATGGRQLWNDVIAPLWKKRIERV RKSFLSNSAKNLWVLYQVFKDDTDEKKKARARQYYHFSVLKEGKN LGFNLTKTREYFLDKFFPIFHSSAPDVKRKVDTFRSKFYAILDFI IYEASVSVANSQGOMGVAPWKGAIDNALVKLREAPDEEAKEKIYN VLAASIRNDSLFLRLKSACDKFGAEQNRPVFPNELRNNRDIRNVR SEWLEATQDVDAAAFVQLIAFLCNFLEGKEINELVTALIKKFEGI QALIDLLRNLEGVDSIRFENE FALFNDDKGNMAGRIARQLRLLAS VGKMKPDMTDAKRVLYKSALEILGAPPDEVSEWLAENILLDKSN NDYQKAKKTVNPFERNYIAKNVITSRSFYLVRYAKPTAVRKLMSN PKIVRYVLKRLPEKQVASYYSAIWTQSESNSNEMVKLIEMIDRLT TEIAGFSFAVLKDKKDSIVSASRESRAVNLEVERLKKLT TLYMSI AYIAVKSLVKVNARYFIAYSALERDLYFFNEKYGEEFRLHFIPYE LNGKTCQFEYLAILKYLLARDEETLKRKCEICEEIKVGCEKHKKN ANPPYEYDQEWIDKKKALNSERKACERRLHFSTHWAQYATKRDEN MAKHPQKWYDILASHYDELLALQATGWLATQARND AEHLNPVNEF DVYIEDLRRYPEGTPKNKDYHIGSYFEIYHYIRQRAYLEEVLAKR KEYRDSGSFTDEQLDKLQKILDDIRARGSYDKNLLKLEYLPFAYN LPRYKNLTTEALFDDDSVSGKKRVAEWREREK TREAREQRRQR (SEQ ID NO: 439)
Cas13d (contig e- k87_11092736) Direct Repeat Sequence	GTGAGAAGTCTCCTTATGGGGAGATGCTAC (SEQ ID NO: 300)
Cas13d (160582958_gene4983 4)	MKNSVTFKLIQAQENKEAARKKAKDIAEQARIAKRNGVVKKEENR INRIQIEIQ TQKKSNTQNAYHLKSLAKAAGVKS VFAIGNDLLMTG FGPGNDATIEKRVFQNRAIETLSSPEQYSAEFQNKQFKIKGNIKV LNHSTQKMEEIQTELQDNYNRP HFDLLGCKNVLEQKYFGRTFSDN

	IHVQIAYNIMDIEKLLTPYINNIIYTLNELMRDNSKDDFFGCDSH FSVAYLYDELKAGYSDRLKTKPNLSKNIDRIWNNFCNYMNSDSGN TEARLAYFGELFYKPKETGDAKSDYKTHLSNNQKEEWELKSDKEV YNIFAILCDLRHFCTHGESITPSGKPPYNLEKNLFPEAKQVLNS LFEEKAESLGAEAFGKTAGKTDVSILLKVFEKEQASQKEQQALLK EYYDFKVQKTYKNMGFSIKKLREAIMEIPDAAKFKDDLYSSLRHK LYGLFDFILVKHFLDTSSENQNNIDIFRQLRACRCEEEKDQVYR SIAVKVWEKVKKELNMFQVVI PSLSKDELKQMEMTKNTELLS SIETISTQASLFSEMI FMMTYLLDGKEINLLCTSLIEKFENIASF NEVLKSPQIGYETKYTEGYAFFKNADKTAKELRQVNNMARMTKPL GGVNTKCVMYNEAAKILGAKPMSKAELESVFNLNDHDTYSPSGK KIPNKNFRNFI INNVITSRRFLYLIRYGNPEKIRKIAINPSIISF VLKQIPDEQIKRYPPCIGKRTDDVTLMRDELGKMLQSVNFEQFS RVNNKQNAKQNPNGEKARLQACVRLYLTVPYLFIKNMVNINARYV LAFHCLERDHALCFNSRKLNDDSYNEMANKFQMRKAKKEQYEKE YKCKKQETGTAHTKKIEKLNQQAIDKDIKMHSTCRNYRNLV AHLNVVSKLQNYVSELPNDYQITSYFSFYHYCMQLGLMEKVSSKN IPLVESLKNEANDAQSYSAKKTLEYFDLIEKNRTYCKDFLKALNA PFSYNLPRFKNLSIEALFDKNIVYEQADLKKE (SEQ ID NO: 440)
Cas13d (160582958_gene4983 4) Direct Repeat Sequence	GAACTACACCCCTCTGTTCTTGTAGGGGTCTAACAC (SEQ ID NO: 301)
Cas13d (contig tpg DJXD01000002.1 ; uncultivated <i>Ruminococcus</i> assembly, UBA7013, from sheep gut metagenome)	MKKQKSCKTVSKTSGLKEALSVQGTVIMTSFGKGNMANLSYKIPS SQKPQNLNSSAGLKNVEVSGKKIKFQGRHPKIATTDNPLFKPQPG MDLLCLKDKLEMHYFGKT FDDNIHIQLIYQILDIEKILAVHVNNI VFTLDNVLHPQKEELTEDFIGAGGWRINLDYQTLRGQTNKYDRFK NYIKRKELLYFGAEFYHENERRYEEDI FAILTLLSALRQFCFHSD LSSDESDHVNSFWLYQLEDQLSDEFKETLSILWEEVTERIDSEFL KTNTVNLHILCHVFPKESKETIVRAYYEFLIKKS FKNMGFSIKKL REIMLEQSDLKSFKEDKYNSVRKLYKLFDFIITYYYDHHA FEKE ALVSSLRSSLTEENKEEIYIKTARTLASALGADFKKAAADVNAKN IRDYQKKANDYRISFEDIKIGNTGIGYFSELIYMLTLLLDGKEIN DLLTTLINKFDNII SFIDILKKLNLEFKFKPEYADFFNMTNCRYT LEELRVINSIARMQKPSADARKIMYRDALRILGMDNRPDEEIDRE LERTMPVGADGKFIKGKQGFRNF IASNVIESSRFHYLVRYNNPHK TRTLVKNPVNVKFVLEGI PETQIKRYFDVCKGQEIPPTSDKSAQI DVLARIISSVDYKIFEDVPQSAKINKDDPSRNFSDALKKQRYQAI VSLYLTVMYLITKNLVVNSRYVIAFHCLERDAFLHGVTLPKMNK KIVYSQLTTHLLTDKNYTTYGHLKNQKGHRKWWYVLVKNLQNSDI TAVSSFRNIVAHISVVRNSNEYISGIGELHSYFELYHYLVQSMIA KNNWYDTSHQPKTAEYLNKHKHTYCKDFVKAYCIPFGYVVPY KNLTINELFDRNNPNPEPKEEV (SEQ ID NO: 441)
Cas13d (contig tpg DJXD01000002.1 ; uncultivated <i>Ruminococcus</i> assembly, UBA7013,	CAACTACAACCCCGTAAAAATACGGGGTTCTGAAAC (SEQ ID NO: 442)

from sheep gut metagenome)	
Cas13d (Gut_metagenome_contig6049000251)	SEQ ID NO: 443
Cas13d (Gut_metagenome_contig546000275)	SEQ ID NO: 444
Cas13d (Gut_metagenome_contig4114000374)	SEQ ID NO: 445
Cas13d (Gut_metagenome_contig721000619)	SEQ ID NO: 446
Cas13d (Gut_metagenome_contig2002000411)	SEQ ID NO: 447
Cas13d (Gut_metagenome_contig13552000311)	SEQ ID NO: 448
Cas13d (Gut_metagenome_contig10037000527)	SEQ ID NO: 449
Cas13d (293. Cas13d from Gut_metagenome_contig238000329)	SEQ ID NO: 450
Cas13d (Gut_metagenome_contig2643000492)	SEQ ID NO: 451
Cas13d (Gut_metagenome_contig874000057)	SEQ ID NO: 452
Cas13d (Gut_metagenome_contig4781000489)	SEQ ID NO: 453
Cas13d (Gut_metagenome_contig12144000352)	SEQ ID NO: 454
Cas13d (Gut_metagenome_contig5590000448)	SEQ ID NO: 455
Cas13d (Gut_metagenome_contig525000349)	SEQ ID NO: 456

Cas13d (Gut_metagenome_contig7229000302)	SEQ ID NO: 457
Cas13d (Gut_metagenome_contig3227000343)	SEQ ID NO: 458
Cas13d (Gut_metagenome_contig7030000469)	SEQ ID NO: 459
Cas13d (gut_metagenome_P17E0k2120140920,_c87000043)	SEQ ID NO: 460
Cas13d (Metagenomic hit (no protein accession): contig emb OBVH01003037.1 , human gut metagenome sequence (also found in WGS contigs emb OBXZ01000094.1 and emb OBJF01000033.1))	SEQ ID NO: 461
Cas13d (Metagenomic hit (no protein accession): contig OGZC01000639.1 (human gut metagenome assembly))	SEQ ID NO: 462
Cas13d (Metagenomic hit (no protein accession): contig emb OHBM01000764.1 (human gut metagenome assembly))	SEQ ID NO: 463
Cas13d (Metagenomic hit (no protein accession): contig emb OHCP01000044.1 (human gut metagenome assembly))	SEQ ID NO: 464

Cas13d (Metagenomic hit (no protein accession): contig emb OGDF01008514.1 (human gut metagenome assembly))	SEQ ID NO: 465
Cas13d (Metagenomic hit (no protein accession): contig emb OGPN01002610.1 (human gut metagenome assembly))	SEQ ID NO: 466
Cas13d (Metagenomic hit (no protein accession): from contig emb OBLI01020244 and emb OBLI01038679 (from pig gut metagenome))	SEQ ID NO: 467
Cas13d (Metagenomic hit (no protein accession): contig OIZX01000427.1)	SEQ ID NO: 468
Cas13d (Metagenomic hit (no protein accession): contig OCTW011587266.1)	SEQ ID NO: 469
Cas13d (Metagenomic hit (no protein accession): contig emb OGNF01009141.1)	SEQ ID NO: 470
Cas13d (Metagenomic hit (no protein accession): contig emb OIEN01002196.1)	SEQ ID NO: 471
Cas13d (Ga0129306_1000735)	SEQ ID NO: 472
Cas13d (Ga0129317_1008067)	SEQ ID NO: 473
Cas13d (Ga0224415_10048792)	SEQ ID NO: 474

Cas13d (250twins_35838_GLO 110300)	SEQ ID NO: 475
Cas13d (250twins_36050_GLO 158985)	SEQ ID NO: 476

Additional exemplary Cas13d sequences and direct repeat sequences for Cas13d are listed as SEQ ID Nos: 1-277 in the corresponding sequence listing.

5 IV. Nucleic acids encoding the capped-sgRNA and/or the Cas polypeptide

Some embodiments disclosed herein provide compositions comprising a nucleic acid sequence encoding the capped-sgRNAs described herein, and vectors (e.g., expression vector(s)) comprising the nucleic acid sequences. In some embodiments, nucleic acid sequences encoding the capped-sgRNAs are operably linked to one or more promoters.

10 Suitable promoters include, without limitation, RNA polymerase II promoters such as, without limitation, CMV, PGK, and EF1 α promoters. In one embodiment, the RNA polymerase II promoter is an RNA Pol II transcribed non-coding RNA. The sgRNA is transcribed by the RNAase polymerase II, acquires an m7G cap and becomes polyadenylated. Additional promoters suitable for driving expression of the capped-sgRNA are also

15 contemplated, such as, without limitation, bacteriophage promoters (e.g., RNA polymerase T3, T7, and SP6), ubiquitous promoters, tissue-specific promoters, inducible promoters, and constitutive promoters. For example, liver-specific promoters as described in PCT/US06/00668 are contemplated herein. For example, promoters for genes encoding genes encoding c-jun; jun-b; c-fos; c-myc; serum amyloid A; apolipoprotein B editing catalytic

20 subunit; liver regeneration factors, such as LRF-I; signal transducers; activators of transcription, such as STAT-3; serum alkaline phosphates (SAP); insulin-like growth factor-binding proteins, such as IGFBP-I; cyclin D1; active protein- 1; CCAAT enhancer core binding protein; ornithine decarboxylase; phosphatase of regenerating liver- 1; early growth response gene-1; hepatocyte growth factors; hemopexin; insulin-like growth factors (IGF),

25 such as IGF-I and IGF-2; hepatocyte nuclear family 1; hepatocyte nuclear family 4; hepatocyte Arg-Ser-rich domain-containing proteins; glucose 6-phosphatase; acute phase proteins, such as serum amyloid A and serum amyloid P (SAA/SAP); steroid hydroxylases; leukotriene hydroxylases; fatty acid hydroxylases; desmolase; peptidyl isomerases; and sterol demethylases.

In one embodiment, vectors comprising the nucleic acids that encode the capped-sgRNA further include a sequence that encodes a Cas polypeptide (e.g., any of the Cas polypeptides described herein, such as, without limitation, a truncated nuclease-deficient Cas protein). The capped-sgRNA and the Cas transcript can be transcribed from the same promoter, or from different promoters. RNA polymerase II promoters (e.g. CMV, PGK, and EF1 α promoters), for example, are suitable for driving expression of both the sgRNA and the Cas gene. In some embodiments, the Cas transcript is expressed from one promoter, such as a PGK promoter, and the capped-sgRNA is expressed from a different promoter, such as an EF1 α promoter.

Some embodiments disclosed herein provide nucleic acids encoding the Cas polypeptides and vectors comprising the nucleic acids that encode the Cas polypeptides. Nucleic acids encoding the Cas polypeptides can be operably linked to one or more promoters. Suitable promoters include RNA polymerase II promoters (e.g., CMV, PGK, and EF1 α promoters), bacteriophage promoters (e.g., RNA polymerase T3, T7, and SP6), ubiquitous promoters, tissue-specific promoters, inducible promoters, and constitutive promoters. The Cas polypeptide can be associated with or include or be in operable linkage with a tag or detectable agent, such as a fluorescent agent, a fluorescent protein, an enzyme.

In some embodiments, a sequence encoding a capped-guide RNA of the disclosure includes a sequence encoding a promoter to drive expression of the guide RNA. In some embodiments, a vector that includes a sequence encoding a capped-guide RNA of the disclosure includes a sequence encoding a promoter to drive expression of the guide RNA. In some embodiments, a promoter driving expression of the guide RNA includes a sequence encoding a constitutive promoter, or an inducible promoter. In some embodiments, a promoter to drive expression of the guide RNA includes a sequence encoding a hybrid or a recombinant promoter. In some embodiments, a promoter to drive expression of the guide RNA is a promoter capable of expressing the guide RNA in a mammalian cell or a human cell. In some embodiments, a promoter to drive expression of the guide RNA is a promoter capable of expressing the guide RNA and restricting the guide RNA to the nucleus of the cell. In some embodiments, a promoter to drive expression of the guide is a human RNA polymerase promoter or a sequence isolated or derived from a human RNA polymerase promoter. In some embodiments, a promoter to drive expression of the guide RNA is a U6 promoter or a sequence isolated or derived from a U6 promoter. In some embodiments, a promoter to drive expression of the guide RNA is a human tRNA promoter or a sequence isolated or derived from a human tRNA promoter. In some embodiments, a promoter to drive

expression of the guide RNA is a human valine tRNA promoter or a sequence isolated or derived from a human valine tRNA promoter.

In some embodiments of the compositions of the disclosure, a promoter to drive expression of the capped-guide RNA further includes a regulatory element. In some
5 embodiments, a vector that includes promoter to drive expression of the guide RNA further includes a regulatory element. In some embodiments, a regulatory element enhances expression of the guide RNA. Exemplary regulatory elements include, but are not limited to, an enhancer element, an intron, an exon, or a combination thereof.

In some embodiments, the nucleic acid sequences encoding the Cas polypeptides are
10 linked to one or more localization signals. Localization signals are amino acid sequences on a protein that tags the protein for transportation to a particular location in a cell. An exemplary localization signal is a nuclear localization signal (NLS), which is an amino acid sequence that tags a protein for import into the cell nucleus by nuclear transport. In one embodiment, one or more localization signals is/are operably linked to the sequence encoding a Cas
15 polypeptide. In some embodiments, the localization signal is a nuclear localization signal (NLS). An exemplary NLS is SV40 large T antigen NLS (PKKKRRV (SEQ ID NO: 477)) and nucleoplasmin NLS (KRPAATKKAGQAKKKK (SEQ ID NO: 478)). Other NLSs are known in the art; see, e.g., Konermann et al., Cell 173:665-676, 2018; Cokol et al., EMBO Rep. 1(5):411-415 (2000); Freitas and Cunha, Curr Genomics 10(8): 550-557 (2009),
20 incorporated herein by reference in their entirety. Without limitation, additional NLSs are those that have K(K/R)X(K/R) as a putative consensus sequence (e.g., PAAKRVKLD (SEQ ID NO: 479)). Other additional NLSs include KRSWSMAF (SEQ ID NO: 480) and KRKYF (SEQ ID NO: 481). In some embodiments, vectors comprising the nucleic acids that encode the Cas polypeptide further encode a capped-sgRNA (e.g., any of the capped-sgRNAs
25 described herein). In some embodiments, the localization signal is a nuclear export signal (NES). Incorporating an NES is particularly suited for altering molecular machinery in the cytoplasm. In one embodiment, an NES is the HIV-REV NES or the PKI NES.

In other embodiments, the nucleic acids encoding the capped-sgRNA and/or the Cas polypeptide can be further operably linked to a sequence that encodes one or more reporter
30 genes, or effector genes such as, without limitation, endonucleases that have nuclease activity. In one embodiment, a nucleic acid sequence encodes a capped-sgRNA disclosed herein and a fusion protein that includes a dCas polypeptide and an endonuclease. Any suitable reporter genes are contemplated, including but not limited to, fluorescent reporters.

In addition, any suitable endonucleases are contemplated for fusing with a Cas polypeptide, in particular a dCas polypeptide.

V. Vectors

5 Vectors contemplated for the present disclosure can include those that are suitable for expression in a selected host, whether prokaryotic or eukaryotic, for example, phage, plasmid, and viral vectors. Viral vectors may be either replication competent or replication defective retroviral vectors. Viral propagation generally will occur only in complementing host cells comprising replication defective vectors, for example, when using replication
10 defective retroviral vectors in methods provided herein viral replication will not occur. Vectors may comprise kozak sequences (Lodish et al., Molecular Cell Biology, 4th ed., 1999) and may also contain the ATG start codon. Promoters that function in a eukaryotic host are from, without limitation, SV40, LTR, CMV, EF-1a, white cloud mountain minnow β -actin.

 In some embodiments of the compositions of the disclosure, a vector of the disclosure
15 includes one or more of a sequence encoding at least one capped-guide RNA of the disclosure, one or more promoters to drive expression of the one or more guide RNAs and a sequence encoding a regulatory element. In some embodiments of the compositions of the disclosure, the vector further includes a sequence encoding a Cas polypeptide, a dCas polypeptide or a dCas-fusion protein.

20 Copy number and positional effects are considered in designing transiently and stably expressed vectors. Copy number can be increased by, for example, dihydrofolate reductase amplification. Positional effects can be optimized by, for example, Chinese hamster elongation factor-1 vector pDEF38 (CHEF1), ubiquitous chromatin opening elements (UCOE), scaffold/matrix-attached region of human (S/MAR), and artificial chromosome
25 expression (ACE) vectors, as well as by using site-specific integration methods known in the art. The expression constructs containing the vector can further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs can include a translation initiating codon at the beginning and a termination codon (UAA, UGA, or UAG)
30 appropriately positioned at the end of the polypeptide to be translated.

 Considering the above-mentioned factors, exemplary vectors suitable for expressing Cas polypeptides and/or sgRNAs in bacteria include PiggyBac transposon vectors, pTT vectors (e.g., from Biotechnology Research Institute (Montreal, Canada)), pQE70, pQE60,

and pQE-9 (e.g. those available from Qiagen (Mississauga, Ontario, Canada)); vectors derived from pcDNA3, available from Invitrogen (Carlsbad, CA); pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH6a, pNH18A, pNH46A, available from Stratagene (La Jolla, CA); and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from
 5 Pharmacia (Peapack, NJ). Among suitable eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXTI, and pSG available from Stratagene (La Jolla, CA); and pSVK3, pBPV, pMSG and pSVL, available from Pharmacia (Peapack, NJ).

A pTT vector backbone can be used for expressing the Cas polypeptide and/or sgRNA (Durocher et al., Nucl. Acids Res. 30:E9 (2002)). Briefly, the backbone of a pTT
 10 vector may be prepared by obtaining pIRESpuro/EGFP (pEGFP) and pSEAP basic vector(s), for example from Clontech (Palo Alto, CA), and pcDNA3.1, pCDNA3.1 /Myc-(His)₆ and pCEP4 vectors can be obtained from, for example, Invitrogen (Carlsbad, CA). As used herein, the pTT5 backbone vector can generate a pTT5-Gateway vector and be used to transiently express proteins in mammalian cells. The pTT5 vector can be derivatized to
 15 pTT5-A, pTT5-B, pTT5-D, pTT5-E, pTT5-H, and pTT5-I, for example. As used herein, the pTT2 vector can generate constructs for stable expression in mammalian cell lines.

A pTT vector can be prepared by deleting the hygromycin (BsmI and Sail excision followed by fill-in and ligation) and EBNA1 (ClaI and NsiI excision followed by fill-in and ligation) expression cassettes. The ColEI origin (FspI-SalI fragment, including the 3' end of
 20 the β -lactamase open reading frame (ORF) can be replaced with a FspI-SalI fragment from pcDNA3.1 containing the pMBI origin (and the same 3' end of β -lactamase ORF). A Myc-(His)₆ C-terminal fusion tag can be added to SEAP (HindIII-HpaI fragment from pSEAP-basic) following in-frame ligation in pcDNA3.1 /Myc-His digested with HindIII and EcoPvV. Plasmids can subsequently be amplified in *E. coli* (DH5a) grown in LB medium and purified
 25 using MAXI prep columns (Qiagen, Mississauga, Ontario, Canada). To quantify, plasmids can be subsequently diluted in, for example, 50 mM Tris-HCl pH 7.4 and absorbencies can be measured at 260 nm and 280 nm. Plasmid preparations with A₂₆₀/A₂₈₀ ratios between about 1.75 and about 2.00 are suitable for producing the Fc-fusion constructs.

The expression vector pTT5 allows for extrachromosomal replication of the cDNA
 30 driven by a cytomegalovirus (CMV) promoter. The plasmid vector pCDNA- pDEST40 is a Gateway-adapted vector which can utilize a CMV promoter for high-level expression. SuperGlo GFP variant (sgGFP) can be obtained from Q-Biogene (Carlsbad, CA). Preparing a pCEP5 vector can be accomplished by removing the CMV promoter and polyadenylation signal of pCEP4 by sequential digestion and self-ligation using Sail and XbaI enzymes

resulting in plasmid pCEP4A. A GblII fragment from pAdCMV5 (Massie et al., J. Virol. 72:2289-2296 (1998)), encoding the CMV5-poly(A) expression cassette ligated in BglII-linearized pCEP4A, resulting in the pCEP5 vector.

Additional vectors include optimized for use in CHO-S or CHO-S-derived cells, such as pDEF38 (CHEF 1) and similar vectors (Running Deer et al., Biotechnol. Prog. 20:880-889 (2004)). The CHEF vectors contain DNA elements that lead to high and sustained expression in CHO cells and derivatives thereof. They may include, but are not limited to, elements that prevent the transcriptional silencing of transgenes.

Vectors may include a selectable marker for propagation in a host. A selectable marker can allow the selection of transformed cells based on their ability to thrive in the presence or absence of a chemical or other agent that inhibits an essential cell function. Selectable markers confer a phenotype on a cell expressing the marker, so that the cell can be identified under appropriate conditions. Suitable markers include genes coding for proteins which confer drug resistance or sensitivity thereto, impart color to, or change the antigenic characteristics of those cells transfected with a molecule encoding the selectable marker, when the cells are grown in an appropriate selective medium.

Suitable selectable markers include dihydro folate reductase or G418 for neomycin resistance in eukaryotic cell culture; and tetracycline, kanamycin, or ampicillin resistance genes for culturing in E. coli and other bacteria. Suitable selectable markers also include cytotoxic markers and drug resistance markers, whereby cells are selected by their ability to grow on media containing one or more of the cytotoxins or drugs; auxotrophic markers, by which cells are selected for their ability to grow on defined media with or without particular nutrients or supplements, such as thymidine and hypoxanthine; metabolic markers for which cells are selected, for example, for ability to grow on defined media containing a defined substance, for example, an appropriate sugar as the sole carbon source; and markers which confer the ability of cells to form colored colonies on chromogenic substrates or cause cells to fluoresce.

Retroviral vectors are contemplated herein. One such vector, the ROSA geo retroviral vector, which maps to mouse chromosome six, was constructed with the reporter gene in reverse orientation with respect to retroviral transcription, downstream of a splice acceptor sequence (U.S. Patent No. 6,461,864; Zambrowicz et al., Proc. Natl. Acad. Sci. 94:3789-3794 (1997)). Infecting embryonic stem (ES) cells with ROSA geo retroviral vector resulted in the ROSA geo26 (ROSA26) mouse strain by random retroviral gene trapping in the ES cells.

Adeno-associated viral vectors (AAV vectors) are contemplated herein. The term “adeno-associated virus” or “AAV” as used herein can refer to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae. Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. At least 11 or 12, sequentially numbered, are disclosed in the prior art. Non-limiting exemplary serotypes useful in the methods disclosed herein include any of the 11 or 12 serotypes, e.g., AAV2, AAV5, and AAV8, or variant serotypes, e.g. AAV-DJ. The AAV structural particle is composed of 60 protein molecules made up of VP1, VP2, and VP3. Each particle contains approximately 5 VP1 proteins, 5 VP2 proteins and 50 VP3 proteins ordered into an icosahedral structure.

AAV is a naturally occurring defective virus that requires another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle. (For a review, see Muzyczka et al., Curr. Topics in Micro and Immunol. 158:97-129 (1992)). AAV vectors efficiently transduce various cell types and can produce long-term expression of transgenes *in vivo*. AAV vectors have been extensively used for gene augmentation or replacement and have shown therapeutic efficacy in a range of animal models as well as in the clinic; see, e.g., Mingozzi and High, Nature Reviews Genetics 12, 341-355 (2011); Deyle and Russell, Curr Opin Mol Ther. 2009 Aug; 11(4): 442-447; Asokan et al., Mol Ther. 2012 April; 20(4): 699-708. AAV vectors containing as little as 300 base pairs of AAV can be packaged and can produce recombinant protein expression. For example, AAV2, AAV5, AAV2/5, AAV2/8 and AAV2/7 vectors have been used to introduce DNA into photoreceptor cells (see, e.g., Pang et al., Vision Research 2008, 48(3):377-385; Khani et al., Invest Ophthalmol Vis Sci. 2007 Sep; 48(9):3954-61; Allocca et al., J. Virol. 2007 81(20):11372-11380). In some embodiments, the AAV vector can include (or include a sequence encoding) an AAV capsid polypeptide described in PCT/US2014/060163; for example, a virus particle comprising an AAV capsid polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, and 17 of PCT/US2014/060163, and a Cas sequence and capped-guide RNA sequence as described herein. In some embodiments, the AAV capsid polypeptide is an Anc80 polypeptide, e.g., Anc80L27; Anc80L59; Anc80L60; Anc80L62; Anc80L65; Anc80L33; Anc80L36; or Anc80L44. In some embodiments, the AAV incorporates inverted terminal repeats (ITRs) derived from the AAV2 serotype. Exemplary left and right ITRs are presented in Table 6 of WO 2018/026976 and are listed below:

AAV2 Left ITR (SEQ ID NO: 482) TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGG
CGACCAAAGGTCGCCCCGACGCCCGGGCTTTGCCCGGGCGGCCTC
AGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATC
ACTAGGGGGTTCCT

AAV2 Right ITR (SEQ ID NO: 483) AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTC
GCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCCCGGGCGTCGGGGCG
ACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGA
GGGAGTGGCCAA

It should be noted, however, that numerous modified versions of the AAV2 ITRs are used in the field, and the ITR sequences shown below are exemplary and are not intended to be limiting. Modifications of these sequences are known in the art, or will be evident to skilled artisans, and are thus included in the scope of this disclosure. Expression of the Cas polypeptide and/or sgRNA in the AAV vector can be driven by a promoter described herein or known in the art. In some embodiments, AAV vectors capable of delivering ~4.5kb are used for packaging of the nucleic acids encoding capped-sgRNAs or Cas polypeptides. In some embodiments, AAVs capable of packaging larger transgenes such as about 4.6kb, 4.7kb, 4.8kb, 4.9kb, 5.0kb, 5.1kb, 5.2kb, 5.3kb, 5.4kb, 5.5kb, 5.6kb, 5.7kb, 5.8kb, 5.9kb, 6.0kb, 6.1kb, 6.2kb, 6.3kb, 6.4kb, 6.5kb, 6.6kb, 6.7kb, 6.8kb, 6.9kb, 7.0kb, 7.5kb, 8.0kb, 9.0kb, 10.0kb, 11.0kb, 12.0kb, 13.0kb, 14.0kb, 15.0kb, or larger are used.

A DNA insert comprising nucleic acids (optionally contained in a vector or vectors) encoding Cas9 polypeptides or sgRNAs can be operatively linked to an appropriate promoter, such as the phage lambda PL promoter; the E. coli lac, trp, phoA, and tac promoters; the SV40 early and late promoters; and promoters of retroviral LTRs. Suitable vectors and promoters also include the pCMV vector with an enhancer, pcDNA3.1 ; the pCMV vector with an enhancer and an intron, pCIneo; the pCMV vector with an enhancer, an intron, and a tripartate leader, pTT2, and CHEF1. The promoter sequences include at least the minimum number of bases or elements necessary to initiate transcription of a gene of interest at levels detectable above background. Within the promoter sequence may be a transcription initiation site, as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. In alternative embodiments, eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes.

The expression constructs can further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding

portion of the transcripts expressed by the constructs can include a translation initiating codon at the beginning and a termination codon (UAA, UGA, or UAG) appropriately positioned at the end of the polypeptide to be translated.

In some embodiments of the compositions and methods of the disclosure, the vector is
 5 or comprises an “RNA targeting system” comprising (a) nucleic acid sequence encoding an Cas polypeptide or dCas polypeptide or dCas polypeptide fusion protein; and (b) a capped-single guide RNA (capped-sgRNA) sequence comprising: an RNA sequence (or spacer sequence) that hybridizes to or binds to a target RNA sequence and an RNA sequence (direct repeat or scaffold sequence) capable of binding to or associating with the Cas polypeptide
 10 and wherein the RNA targeting system recognizes the target RNA and enhances translation of the target RNA. In some embodiments, the nucleic acid sequence or vector is a single vector.

VI. Cells

15 Some embodiments disclosed herein provide cells comprising the nucleic acid or nucleic acids (e.g., vector or vectors) that encode Cas9 polypeptides or capped-sgRNAs. In some embodiments, the cells transfected may be a prokaryotic cell, a eukaryotic cell, a yeast cell, an insect cell, an animal cell, a mammalian cell, a human cell, etc. The proteins expressed in mammalian cells have been glycosylated properly. Examples of useful
 20 mammalian host cell lines are HEK293, CHO, sp2/0, NSO, COS, BHK, and PerC6.

In some embodiments, the target mRNA is in a cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a mammalian cell. In some embodiments, the cell is a bovine, murine, feline, equine, porcine, canine, simian, or human cell. In some embodiments, the cell is a plant cell.
 25 In some embodiments, the cell is in a subject or patient. In some embodiments, the cell is *in vivo*, *in vitro*, *ex vivo*, or *in situ*. In some embodiments, the composition comprises a vector comprising composition comprising the capped-guide RNA of the disclosure and/or a Cas polypeptide or dCas polypeptide. In some embodiments, the vector is a viral vector for transducing a cell. In some embodiments, the viral vector is an AAV vector.

30 Transfection of animal cells typically involves opening transient pores or “holes” in the cell membrane, to allow the uptake of material. There are various methods of introducing foreign DNA into a eukaryotic cell. Transfection can be carried out using calcium phosphate, by electroporation, or by mixing a cationic lipid with the material to produce liposomes,

which fuse with the cell membrane and deposit their cargo inside. Many materials have been used as carriers for transfection, which can be divided into three kinds: (cationic) polymers, liposomes, and nanoparticles.

5 VII. Methods of modulating protein translation

Some embodiments disclosed herein provide compositions for and methods of enhancing protein translation in a cell, comprising introducing capped-sgRNAs (e.g., any of the capped-sgRNAs described herein) and Cas polypeptides (e.g., any of the Cas polypeptides described herein) into the cell. The methods of enhancing protein translation can include introducing or administering a nucleic acid or nucleic acids (e.g., vector or vectors) encoding the capped-sgRNA and the Cas polypeptides into a cell. In some embodiments, provided herein are methods of regulating translation of an mRNA in a cell that include contacting the cell with a nucleic acid comprising (a) a sequence encoding a Cas polypeptide; and (b) a sequence encoding a capped-sgRNA comprising (i) an m7G cap; (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and (iii) a direct repeat capable of binding to the Cas polypeptide.

Methods of measuring levels of protein translation are known in the art. Exemplary methods include, without limitation, western blot, mass spectrometry, antibody staining, and mean fluorescence intensity flow cytometry. In instances where a reporter construct is linked to the target mRNA, protein translation can also be measured based on the levels of the reporter molecule.

In some embodiments, enhancing translation or increasing or upregulating gene expression refers to an increase in the amount of peptide translated from the target mRNA as compared to a control. In some embodiments, the control includes a level of peptide translated from the target mRNA in the absence of the capped-sgRNA compositions and methods. In some embodiments, the control includes the level of the peptide translated from the target mRNA prior to addition of the compositions disclosed herein. In some embodiments, translation is increased about 1.1 fold, about 1.2 fold, about 1.3 fold, about 1.4 fold, about 1.5 fold, about 1.6 fold, about 1.7 fold, about 1.8 fold, about 1.9 fold, about 2 fold, about 2.5 fold, about 3 fold, about 4 fold, about 5 fold, about 6 fold, about 7 fold, about 8 fold, about 9 fold, about 10 fold, about 20 fold, about 50 fold, about 100 fold, about 1000 fold, or about 10,000 fold relative to the control. The amount of peptide translated can be determined by any method known in the art.

Gene Therapy/Therapeutic Targets

In certain embodiments, methods of modulating protein translation are useful for treating patients afflicted with a disease or disorder. In one embodiment, methods of using the capped-sgRNA compositions disclosed herein are haploinsufficiencies. Exemplary haploinsufficiency diseases or disorders include, without limitation, Autosomal dominant Retinitis Pigmentosa (RP11) caused by mutations in *PRPF31*, Autosomal dominant Retinitis Pigmentosa (RP31) caused by mutations in *TOPORS*, Frontotemporal dementia caused by mutations in *GRN*, DeVivo Syndrome (Glut1 deficiency) caused by mutations in *SLC2A1*, Dravet syndrome caused by mutations in *SCN1A*, 1q21.1 Deletion Syndrome, 5q-Syndrome in Myelodysplastic Syndrome (MDS), 22q11.2 Deletion Syndrome, CHARGE Syndrome, Cleidocranial Dysostosis, Ehlers-Danlos Syndrome, Frontotemporal Dementia caused by mutations in Progranulin, Haploinsufficiency of A20, Holoprosencephaly (caused by haploinsufficiency in the Sonic Hedgehog gene), Holt-Oram Syndrome, Marfan Syndrome, Dyskeratosis Congenita, and Phelan-McDermid Syndrome.

In another embodiment, methods of using the capped-sgRNA compositions disclosed herein for treating haploinsufficiency diseases or disorders, such as, without limitation, those listed in the preceding paragraph, involving mutations which lead to introduction of a premature termination codon (PTC) resulting in degradation from mutant allele or loss of function of the protein (or less protein to be produced) are contemplated herein.

In another embodiment, methods of translation enhancement using the capped-sgRNA compositions disclosed herein are useful for treating cancer. In one embodiment, the methods can be used for upregulating protein expression of tumor suppressor genes (TSG) in tissue predisposed to cancer due to hereditary (or acquired) mutations of TSG. In another embodiment, the methods can be used for upregulating protein expression from genes that would prevent cancer from metastasizing (ie angiogenesis genes). In another embodiment, the methods can be used for upregulating protein expression from genes that would result in the cancer being more susceptible to follow-up treatments. In another embodiment, the methods can be used for translational enhancement to prevent cancer evasion of the immune system.

As used herein, the “administration” of the compositions disclosed herein (e.g., a fusion RNA, viral particle, vector, polynucleotide, cell, population of cells, or pharmaceutical composition or formulation) to a subject includes any route of introducing or delivering to a subject the agent to perform its intended function. Administration can be carried out by any

suitable route, including orally, intranasally, intraocularly, ophthalmically, parenterally (intravenously, intramuscularly, intraperitoneally, or subcutaneously), or topically.

Administration includes self-administration and the administration by another.

In some aspects, the disclosure provides a method of treating a disease or disorder comprising administering to a subject a therapeutically effective amount of a capped-sgRNA composition(s) of the disclosure. Also provided herein are methods for treating a disease or condition in a subject in need thereof, the methods comprising, consisting of, or consisting essentially of administering a nucleic acid sequence comprising/encoding (a) a capped-sgRNA disclosed herein; and (b) a dCas polypeptide, a vector comprising the nucleic acid sequence, or a viral particle comprising the vector to the subject, thereby enhancing translation of a target mRNA in the subject or patient. In some embodiments, the target mRNA is involved in the etiology of a disease or condition in the subject.

In some embodiments of the methods described herein, the subject or patient is an animal. In some embodiments, the subject is a mammal. In some embodiments, the mammal is a bovine, equine, porcine, canine, feline, simian, murine, or human. In some embodiments, the subject is a human.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, a genetic disease or disorder. In some embodiments, the genetic disease or disorder is a single-gene disease or disorder. In some embodiments, the single-gene disease or disorder is an autosomal dominant disease or disorder, an autosomal recessive disease or disorder, an X-chromosome linked (X-linked) disease or disorder, an X-linked dominant disease or disorder, an X-linked recessive disease or disorder, a Y-linked disease or disorder or a mitochondrial disease or disorder. In some embodiments, the genetic disease or disorder is a multiple-gene disease or disorder. In some embodiments, the genetic disease or disorder is a multiple-gene disease or disorder. In some embodiments, the single-gene disease or disorder is an autosomal dominant disease or disorder including, but not limited to, Huntington's disease, neurofibromatosis type 1, neurofibromatosis type 2, Marfan syndrome, hereditary nonpolyposis colorectal cancer, hereditary multiple exostoses, Von Willebrand disease, and acute intermittent porphyria. In some embodiments, the single-gene disease or disorder is an autosomal recessive disease or disorder including, but not limited to, Albinism, Medium-chain acyl-CoA dehydrogenase deficiency, cystic fibrosis, sickle-cell disease, Tay-Sachs disease, Niemann-Pick disease, spinal muscular atrophy, and Roberts syndrome. In some embodiments, the single-gene disease or disorder is X-linked disease or disorder including, but not limited to, muscular

dystrophy, Duchenne muscular dystrophy, Hemophilia, Adrenoleukodystrophy (ALD), Rett syndrome, and Hemophilia A. In some embodiments, the single-gene disease or disorder is a mitochondrial disorder including, but not limited to, Leber's hereditary optic neuropathy.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, an immune disease or disorder. In some embodiments, the immune disease or disorder is an immunodeficiency disease or disorder including, but not limited to, B-cell deficiency, T-cell deficiency, neutropenia, asplenia, complement deficiency, acquired immunodeficiency syndrome (AIDS) and immunodeficiency due to medical intervention (immunosuppression as an intended or adverse effect of a medical therapy). In some embodiments, the immune disease or disorder is an autoimmune disease or disorder including, but not limited to, Achalasia, Addison's disease, Adult Still's disease, Agammaglobulinemia, Alopecia areata, Amyloidosis, Anti-GBM/Anti-TBM nephritis, Antiphospholipid syndrome, Autoimmune angioedema, Autoimmune dysautonomia, Autoimmune encephalomyelitis, Autoimmune hepatitis, Autoimmune inner ear disease (AIED), Autoimmune myocarditis, Autoimmune oophoritis, Autoimmune orchitis, Autoimmune pancreatitis, Autoimmune retinopathy, Autoimmune urticaria, Axonal & neuronal neuropathy (AMAN), Balo disease, Behcet's disease, Benign mucosal pemphigoid, Bullous pemphigoid, Castleman disease (CD), Celiac disease, Chagas disease, Chronic inflammatory demyelinating polyneuropathy (CIDP), Chronic recurrent multifocal osteomyelitis (CRMO), Churg-Strauss Syndrome (CSS) or Eosinophilic Granulomatosis (EGPA), Cicatricial pemphigoid, Cogan's syndrome, Cold agglutinin disease, Congenital heart block, Coxsackie myocarditis, CREST syndrome, Crohn's disease, Dermatitis herpetiformis, Dermatomyositis, Devic's disease (neuromyelitis optica), Discoid lupus, Dressler's syndrome, Endometriosis, Eosinophilic esophagitis (EoE), Eosinophilic fasciitis, Erythema nodosum, Essential mixed cryoglobulinemia, Evans syndrome, Fibromyalgia, Fibrosing alveolitis, Giant cell arteritis (temporal arteritis), Giant cell myocarditis, Glomerulonephritis, Goodpasture's syndrome, Granulomatosis with Polyangiitis, Graves' disease, Guillain-Barre syndrome, Hashimoto's thyroiditis, Hemolytic anemia, Henoch-Schonlein purpura (HSP), Herpes gestationis or pemphigoid gestationis (PG), Hidradenitis Suppurativa (HS) (Acne Inversa), Hypogammaglobulinemia, IgA Nephropathy, IgG4-related sclerosing disease, Immune thrombocytopenic purpura (ITP), Inclusion body myositis (IBM), Interstitial cystitis (IC), Juvenile arthritis, Juvenile diabetes (Type 1 diabetes), Juvenile myositis (JM), Kawasaki disease, Lambert-Eaton syndrome, Leukocytoclastic vasculitis, Lichen planus, Lichen sclerosus, Ligneous conjunctivitis, Linear

IgA disease (LAD), Lupus, Lyme disease chronic, Meniere's disease, Microscopic
 polyangiitis (MPA), Mixed connective tissue disease (MCTD), Mooren's ulcer, Mucha-
 Habermann disease, Multifocal Motor Neuropathy (MMN) or MMNCB, Multiple sclerosis,
 Myasthenia gravis, Myositis, Narcolepsy, Neonatal Lupus, Neuromyelitis optica,
 5 Neutropenia, Ocular cicatricial pemphigoid, Optic neuritis, Palindromic rheumatism (PR),
 PANDAS, Paraneoplastic cerebellar degeneration (PCD), Paroxysmal nocturnal
 hemoglobinuria (PNH), Parry Romberg syndrome, Pars planitis (peripheral uveitis),
 Parsonage-Turner syndrome, Pemphigus, Peripheral neuropathy, Perivenous
 encephalomyelitis, Pernicious anemia (PA), POEMS syndrome, Polyarteritis nodosa,
 10 Polyglandular syndromes type I, II, III, Polymyalgia rheumatica, Polymyositis,
 Postmyocardial infarction syndrome, Postpericardiotomy syndrome, Primary biliary cirrhosis,
 Primary sclerosing cholangitis, Progesterone dermatitis, Psoriasis, Psoriatic arthritis, Pure red
 cell aplasia (PRCA), Pyoderma gangrenosum, Raynaud's phenomenon, Reactive Arthritis,
 Reflex sympathetic dystrophy, Relapsing polychondritis, Restless legs syndrome (RLS),
 15 Retroperitoneal fibrosis, Rheumatic fever, Rheumatoid arthritis, Sarcoidosis, Schmidt
 syndrome, Scleritis, Scleroderma, Sjogren's syndrome, Sperm & testicular autoimmunity,
 Stiff person syndrome (SPS), Subacute bacterial endocarditis (SBE), Susac's syndrome,
 Sympathetic ophthalmia (SO), Takayasu's arteritis, Temporal arteritis/Giant cell arteritis,
 Thrombocytopenic purpura (TTP), Tolosa-Hunt syndrome (THS), Transverse myelitis, Type
 20 1 diabetes, Ulcerative colitis (UC), Undifferentiated connective tissue disease (UCTD),
 Uveitis, Vasculitis, Vitiligo, Vogt-Koyanagi-Harada Disease, or Wegener's granulomatosis.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, an inflammatory disease or disorder.

In some embodiments of the compositions and methods of the disclosure, a disease or
 25 disorder of the disclosure includes, but is not limited to, a metabolic disease or disorder.

In some embodiments of the compositions and methods of the disclosure, a disease or
 disorder of the disclosure includes, but is not limited to, a degenerative or a progressive
 disease or disorder. In some embodiments, the degenerative or a progressive disease or
 disorder includes, but is not limited to, amyotrophic lateral sclerosis (ALS), Huntington's
 30 disease, Alzheimer's disease, and aging.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, an infectious disease or disorder.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, a pediatric or a developmental disease or disorder.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, a cardiovascular disease or disorder.

In some embodiments of the compositions and methods of the disclosure, a disease or disorder of the disclosure includes, but is not limited to, a proliferative disease or disorder. In some embodiments, the proliferative disease or disorder is a cancer. In some embodiments, the cancer includes, but is not limited to, Acute Lymphoblastic Leukemia (ALL), Acute Myeloid Leukemia (AML), Adrenocortical Carcinoma, AIDS-Related Cancers, Kaposi Sarcoma (Soft Tissue Sarcoma), AIDS-Related Lymphoma (Lymphoma), Primary CNS Lymphoma (Lymphoma), Anal Cancer, Appendix Cancer, Gastrointestinal Carcinoid Tumors, Astrocytomas, Atypical Teratoid/Rhabdoid Tumor, Central Nervous System (Brain Cancer), Basal Cell Carcinoma, Bile Duct Cancer, Bladder Cancer, Bone Cancer, Ewing Sarcoma, Osteosarcoma, Malignant Fibrous Histiocytoma, Brain Tumors, Breast Cancer, Burkitt Lymphoma, Carcinoid Tumor, Carcinoma, Cardiac (Heart) Tumors, Embryonal Tumors, Germ Cell Tumor, Primary CNS Lymphoma, Cervical Cancer, Cholangiocarcinoma, Chordoma, Chronic Lymphocytic Leukemia (CLL), Chronic Myelogenous Leukemia (CML), Chronic Myeloproliferative Neoplasms, Colorectal Cancer, Craniopharyngioma, Cutaneous T-Cell Lymphoma, Ductal Carcinoma In Situ, Embryonal Tumors, Endometrial Cancer (Uterine Cancer), Ependymoma, Esophageal Cancer, Esthesioneuroblastoma (Head and Neck Cancer), Ewing Sarcoma (Bone Cancer), Extracranial Germ Cell Tumor, Extragonadal Germ Cell Tumor, Eye Cancer, Childhood Intraocular Melanoma, Intraocular Melanoma, Retinoblastoma, Fallopian Tube Cancer, Fibrous Histiocytoma of Bone, Malignant, and Osteosarcoma, Gallbladder Cancer, Gastric (Stomach) Cancer, Gastrointestinal Carcinoid Tumor, Gastrointestinal Stromal Tumors (GIST) (Soft Tissue Sarcoma), Childhood Gastrointestinal Stromal Tumors, Germ Cell Tumors, Childhood Extracranial Germ Cell Tumors, Extragonadal Germ Cell Tumors, Ovarian Germ Cell Tumors, Testicular Cancer, Gestational Trophoblastic Disease, Hairy Cell Leukemia, Head and Neck Cancer, Heart Tumors, Hepatocellular (Liver) Cancer, Histiocytosis, Hodgkin Lymphoma, Hypopharyngeal Cancer (Head and Neck Cancer), Intraocular Melanoma, Islet Cell Tumors, Pancreatic Neuroendocrine Tumors, Kaposi Sarcoma (Soft Tissue Sarcoma), Kidney (Renal Cell) Cancer, Langerhans Cell Histiocytosis, Laryngeal Cancer (Head and Neck Cancer), Leukemia, Lip and Oral Cavity Cancer (Head and Neck Cancer), Liver Cancer, Lung Cancer

(Non-Small Cell and Small Cell), Childhood Lung Cancer, Lymphoma, Male Breast Cancer, Malignant Fibrous Histiocytoma of Bone and Osteosarcoma, Melanoma, Merkel Cell Carcinoma (Skin Cancer), Mesothelioma, Metastatic Squamous Neck Cancer with Occult Primary (Head and Neck Cancer), Midline Tract Carcinoma With NUT Gene Changes,

5 Mouth Cancer (Head and Neck Cancer), Multiple Endocrine Neoplasia Syndromes, Multiple Myeloma/Plasma Cell Neoplasms, Mycosis Fungoides (Lymphoma), Myelodysplastic Syndromes, Myelodysplastic/Myeloproliferative Neoplasms, Nasal Cavity and Paranasal Sinus Cancer (Head and Neck Cancer), Nasopharyngeal Cancer (Head and Neck Cancer), Neuroblastoma, Non- Hodgkin Lymphoma, Non-Small Cell Lung Cancer, Oral Cancer, Lip

10 and Oral Cavity Cancer and Oropharyngeal Cancer, Osteosarcoma and Malignant Fibrous Histiocytoma of Bone, Ovarian Cancer, Pancreatic Cancer, Pancreatic Neuroendocrine Tumors (Islet Cell Tumors), Papillomatosis, Paraganglioma, Parathyroid Cancer, Penile Cancer, Pharyngeal Cancer (Head and Neck Cancer), Pheochromocytoma , Plasma Cell Neoplasm/Multiple Myeloma, Pleuropulmonary Blastoma, Pregnancy and Breast Cancer,

15 Primary Central Nervous System (CNS) Lymphoma, Primary Peritoneal Cancer, Prostate Cancer, Rectal Cancer, Recurrent Cancer, Renal Cell (Kidney) Cancer, Retinoblastoma, Rhabdomyosarcoma, Childhood (Soft Tissue Sarcoma), Salivary Gland Cancer (Head and Neck Cancer), Sarcoma, Childhood Rhabdomyosarcoma (Soft Tissue Sarcoma), Childhood Vascular Tumors (Soft Tissue Sarcoma), Ewing Sarcoma (Bone Cancer), Kaposi Sarcoma

20 (Soft Tissue Sarcoma), Osteosarcoma (Bone Cancer), Uterine Sarcoma, Sezary Syndrome, Lymphoma, Skin Cancer, Small Cell Lung Cancer, Small Intestine Cancer, Soft Tissue Sarcoma, Squamous Cell Carcinoma of the Skin, Squamous Neck Cancer, Stomach (Gastric) Cancer, T-Cell Lymphoma, Testicular Cancer, Throat Cancer (Head and Neck Cancer), Nasopharyngeal Cancer, Oropharyngeal Cancer, Hypopharyngeal Cancer, Thymoma and

25 Thymic Carcinoma , Thyroid Cancer, Transitional Cell Cancer of the Renal Pelvis and Ureter, Renal Cell Cancer, Urethral Cancer, Uterine Sarcoma, Vaginal Cancer, Vascular Tumors (Soft Tissue Sarcoma), Vulvar Cancer, Wilms Tumor and Other Childhood Kidney Tumors.

In some embodiments of the methods of the disclosure, a subject of the disclosure has

30 been diagnosed with the disease or disorder. In some embodiments, the subject of the disclosure presents at least one sign or symptom of the disease or disorder. In some embodiments, the subject has a biomarker predictive of a risk of developing the disease or disorder. In some embodiments, the biomarker is a genetic mutation.

In some embodiments of the methods of the disclosure, a subject of the disclosure is female. In some embodiments of the methods of the disclosure, a subject of the disclosure is male. In some embodiments, a subject of the disclosure has two XX or XY chromosomes. In some embodiments, a subject of the disclosure has two XX or XY chromosomes and a third
5 chromosome, either an X or a Y.

In some embodiments of the methods of the disclosure, a subject of the disclosure is a neonate, an infant, a child, an adult, a senior adult, or an elderly adult. In some embodiments of the methods of the disclosure, a subject of the disclosure is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or 31 days old.
10 In some embodiments of the methods of the disclosure, a subject of the disclosure is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 months old. In some embodiments of the methods of the disclosure, a subject of the disclosure is at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or any number of years or partial years in between of age.

15 In some embodiments of the methods of the disclosure, a subject of the disclosure is a mammal. In some embodiments, a subject of the disclosure is a non-human mammal.

In some embodiments of the methods of the disclosure, a subject of the disclosure is a human.

In some embodiments of the methods of the disclosure, a therapeutically effective
20 amount comprises a single dose of a composition of the disclosure. In some embodiments, a therapeutically effective amount comprises a therapeutically effective amount comprises at least one dose of a composition of the disclosure. In some embodiments, a therapeutically effective amount comprises a therapeutically effective amount comprises one or more dose(s) of a composition of the disclosure. In some embodiments of the methods of the disclosure, a
25 therapeutically effective amount eliminates a sign or symptom of the disease or disorder. In some embodiments, a therapeutically effective amount reduces a severity of a sign or symptom of the disease or disorder.

In some embodiments of the methods of the disclosure, a therapeutically effective amount eliminates the disease or disorder.

30 In some embodiments of the methods of the disclosure, a therapeutically effective amount prevents an onset of a disease or disorder. In some embodiments, a therapeutically effective amount delays the onset of a disease or disorder. In some embodiments, a therapeutically effective amount reduces the severity of a sign or symptom of the disease or

disorder. In some embodiments, a therapeutically effective amount improves a prognosis for the subject.

In some embodiments of the methods of the disclosure, a composition of the disclosure is administered to the subject systemically. In some embodiments, the composition
5 of the disclosure is administered to the subject by an intravenous route. In some embodiments, the composition of the disclosure is administered to the subject by an injection or an infusion.

In some embodiments of the methods of the disclosure, a composition of the disclosure is administered to the subject locally. In some embodiments, the composition of
10 the disclosure is administered to the subject by an intraosseous, intraocular, intracerebrospinal, or intraspinal route. In some embodiments, the composition of the disclosure is administered directly to the cerebral spinal fluid of the central nervous system. In some embodiments, the composition of the disclosure is administered directly to a tissue or fluid of the eye and does not have bioavailability outside of ocular structures. In some
15 embodiments, the composition of the disclosure is administered to the subject by an injection or an infusion.

Also provided herein is a pharmaceutical composition comprising any one or more of the capped-sgRNAs and Cas or dCas polypeptide, or the nucleic acid sequences encoding the polypeptide, and a carrier. In some embodiments, a composition can be one or more
20 polynucleotides encoding a capped-guide nucleotide sequence-Cas polypeptide or fusion polypeptide. In some embodiments, a composition can be any of the nucleic acids or proteins described herein. In some embodiments, a composition can be any polynucleotide described herein. In some embodiments, the carrier is a pharmaceutically acceptable carrier. In some embodiments, the composition is a pharmaceutical composition comprising a capped-guide
25 nucleotide sequence-Cas polypeptide or fusion polypeptide, and a pharmaceutically acceptable carrier. In some embodiments, the composition or pharmaceutical composition further comprises one or more gRNAs, capped-sgRNAs, crRNAs, and/or tracrRNAs.

Briefly, pharmaceutical compositions disclosed herein may comprise a capped-guide nucleotide sequence-Cas polypeptide or fusion polypeptide or a nucleotide sequence
30 encoding the same, optionally comprised in an AAV, which is optionally also immune orthogonal, in combination with one or more pharmaceutically or physiologically acceptable carriers, diluents or excipients. Such compositions may comprise buffers such as neutral buffered saline, phosphate buffered saline and the like; carbohydrates such as glucose, mannose, sucrose or dextrans, mannitol; proteins; polypeptides or amino acids such as

glycine; antioxidants; chelating agents such as EDTA or glutathione; adjuvants (e.g., aluminum hydroxide); and preservatives. Compositions of the present disclosure may be formulated for oral, intravenous, topical, enteral, subpial, and/or parenteral administration. In certain embodiments, the compositions of the disclosure are formulated for intravenous administration.

In some embodiments of the methods of the disclosure, a composition of the disclosure is administered to the subject locally. In some embodiments, the composition of the disclosure is administered to the subject by an intraosseous, intraocular, intracerebrospinal, intraspinal, or subpial route. In some embodiments, the composition of the disclosure is administered directly to the cerebral spinal fluid of the central nervous system. In some embodiments, the composition of the disclosure is administered directly to a tissue or fluid of the eye and does not have bioavailability outside of ocular structures. In some embodiments, the composition of the disclosure is administered to the subject by an injection or an infusion.

In some embodiments, the compositions disclosed herein are formulated as pharmaceutical compositions. Briefly, pharmaceutical compositions for use as disclosed herein may include a protein(s) or a polynucleotide encoding the protein(s), optionally comprised in an AAV, which is optionally also immune orthogonal, in combination with one or more pharmaceutically or physiologically acceptable carriers, diluents or excipients. Such compositions may include buffers such as neutral buffered saline, phosphate buffered saline and the like; carbohydrates such as glucose, mannose, sucrose or dextrans, mannitol; proteins; polypeptides or amino acids such as glycine; antioxidants; chelating agents such as EDTA or glutathione; adjuvants (e.g., aluminum hydroxide); and preservatives. Compositions of the disclosure may be formulated for oral, intravenous, topical, enteral, intraocular, and/or parenteral administration. In some embodiments, intraocular administration includes, without limitation, subretinal, intravitreal, or topical (via eye drops) administration. In certain embodiments, the compositions of the disclosure are formulated for intravenous administration.

EXAMPLES

The following examples are provided to better illustrate the claimed invention and are not to be interpreted as limiting the scope of the invention. To the extent that specific materials are mentioned, it is merely for purposes of illustration and is not intended to limit

the invention. One skilled in the art can develop equivalent means or reactants without the exercise of inventive capacity and without departing from the scope of the invention.

Example 1: Modulation of translation using dCas and capped-sgRNA

5 Exemplary constructs for generating nuclease dead Cas (dCas13b) and m7G capped-sgRNA are shown in FIG. 1A. dCas and capped-sgRNA were transcribed from either the same promoter (i), or independent promoters (ii) by RNA polymerase II. An exemplary structure for the unprocessed capped-sgRNA is shown in FIG. 1B. From 5' to 3', the capped-sgRNA contains the m7G cap, a linker of variable length, a spacer, a direct repeat, an RNase
10 P processing site such as a tRNA-like small RNA, and a poly-A tail. An exemplary chemical composition of a capped-sgRNA is shown in FIG. 1C. Variants of the m7G cap include either guanosine or adenine for the second nucleotide of the di-nucleotide m7G cap (at the a' position). FIG. 1D shows RNase P processing to trim downstream sequences, including the poly-A tail, from the transcript. The capped-RNA complexed with dCas, binds to a site on the
15 target messenger RNA thereby bringing the m7G cap of the sgRNA to the vicinity of a desired start codon (FIG. 1E).

A two construct system was used to test the ability of dCas and capped-sgRNA to regulate translation *in trans*. The first construct includes a dCas13b gene driven by a PGK promoter, followed by a NLS-Turquoise reporter which is driven by an EF1 α promoter. The
20 second construct includes a sequence encoding a capped sgRNA, a 5'UTR region of ATF4, and the ATF4 ORF linked to an NL-Citrine promoter. Both the sequence encoding a capped sgRNA and the ATF4 ORF are driven by a TRE promoter. The second construct further includes an rtTA-P2A-Puro construct linked to an NLS-Cherry reporter via an IRES sequence. The rtTA-P2A-Puro-IRES-NLS-Cherry portion is driven by the EF1 α promoter. A
25 number of different types of capped sgRNAs (sg(1), sg(2), sg(3), sg(4), sg(5), sg(6)) were generated, each included a spacer that targets the ATF4 transcript at a different site. The sequences of the sgRNAs are listed in below. The target sequences were within the "sgRNA targeting window" as shown in FIG. 1G. Control capped-sgRNAs were also generated that do not target a sequence within the ATF4 transcript. These control capped-sgRNAs are referred
30 to as non-targeting sgRNAs. Each capped-sgRNA was tested using dCas gradient in combination with a target ATF4 transcript gradient. The plots in FIG. 1H reflect log2 fold difference in Citrine/Cherry ratio between each capped sgRNA – sg (a1), sg(a2), sg(a3), sg(a4), sg(a5) and sg(a6), and non-targeting sgRNAs. A spacer targeting a region just 3'

proximal to the start codon AUG was shown to result in the biggest increase in protein expression. The results demonstrate that optimal translational control is a function of the target sequence chosen, the expression level of the target RNA, and the expression level of dCas.

- 5 Uncapped-sgRNAs that correspond to each of the capped-sgRNA tested were subjected to the same gradient test. As shown in FIG. 1I, uncapped-sgRNAs were either ineffective in enhancing translation, or only minimally enhanced translation. These results showed that the localized recruitment of the 5' m7G cap proximal to a start codon enabled an enhancement in translation.

sgRNA	Sequence
sg(1)	TTTGCTGGAATCGAGGAATGTGCTT (SEQ ID NO: 278)
sg(2)	GTTGCGGTGCTTTGCTGGAATCGAG (SEQ ID NO: 279)
sg(3)	TTTCGGTCATGTTGCGGTGCTTTGC (SEQ ID NO: 280)
sg(4)	AGGAAGCTCATTTTCGGTCATGTTGC (SEQ ID NO: 281)
sg(5)	CTCGCTGCTCAGGAAGCTCATTTTCG (SEQ ID NO: 282)
sg(6)	CCACCAACACCTCGCTGCTCAGGAA (SEQ ID NO: 283)

10

ADDITIONAL EMBODIMENTS

Embodiment 1: A capped single guide RNA (Capped-sgRNA) comprising from 5' to 3':

- 15 an m7G cap,
 a linker,
 a spacer complementary to a target sequence in a messenger RNA,
 a direct repeat sequence, and
 a Ribonuclease P (RNase P) processing site,
 wherein the direct repeat sequence is capable of binding to a Cas protein.

20

Embodiment 2: The Capped-sgRNA of Embodiment 1, wherein the Cas protein is a nuclease dead Cas (dCas) protein.

Embodiment 3: The Capped-sgRNA of Embodiment 1, wherein the m7G cap comprises one
5 or more chemical modifications relative to the structure of a naturally occurring m7G cap.

Embodiment 4: The Capped-sgRNA of Embodiment 1, wherein the linker comprises about 5 to about 25 nucleotides.

10 Embodiment 5: The Capped-sgRNA of Embodiment 4, wherein the linker comprises about 8 to about 20 nucleotides.

Embodiment 6: The Capped-sgRNA of Embodiment 1, wherein the linker is non-complementary to any messenger RNA sequence.

15 Embodiment 7: The Capped-sgRNA of Embodiment 1, wherein the linker comprises the sequence of GTCAGATCGCCTGGAATT.

Embodiment 8: The Capped-sgRNA of Embodiment 1, wherein the target sequence is
20 proximal to a target start codon of the messenger RNA relative to a 5' m7G cap of the messenger RNA.

Embodiment 9: The Capped-sgRNA of Embodiment 1, wherein the target sequence comprises the target start codon of the messenger RNA.

25 Embodiment 10: The Capped-sgRNA of Embodiment 8, wherein the 5' end of the target sequence is upstream to the target start codon of the messenger RNA.

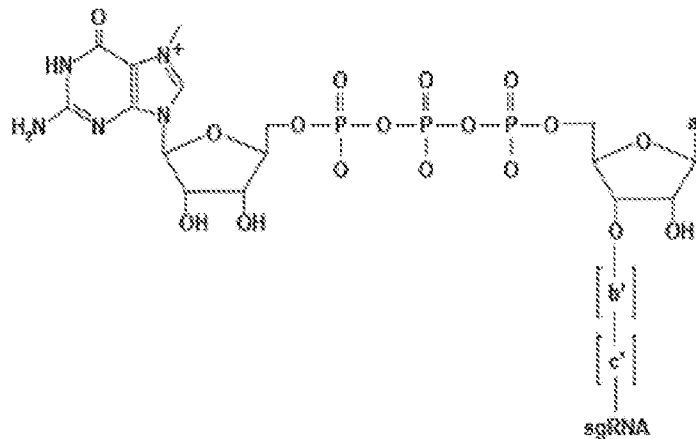
Embodiment 11: The Capped-sgRNA of Embodiment 8, wherein the 5' end of the target
30 sequence is downstream to the target start codon of the messenger RNA.

Embodiment 12: The Capped-sgRNA of Embodiment 1, wherein the spacer is at least 80% complementary to the target sequence in the messenger RNA.

Embodiment 13: The Capped-sgRNA of Embodiment 1, wherein the spacer is at least 90% complementary to the target sequence in the messenger RNA.

Embodiment 14: The Capped-sgRNA of Embodiment 1, further comprising a polyadenylated
5 tail.

Embodiment 15: The Capped-sgRNA of Embodiment 1, having the structure:



wherein a' is a guanosine or adenine, b' is the linker, and c' is the spacer.

10

Embodiment 16: An expression vector encoding the Capped-sgRNA of Embodiment 1.

Embodiment 17: The expression vector of Embodiment 16, further encodes a nuclease dead Cas9.

15

Embodiment 18: A Capped-sgRNA generated by processing the Capped-sgRNA of Embodiment 1 using an RNase P.

Embodiment 19: An expression vector comprising a nucleic acid sequence encoding:

20

a nuclease dead Cas (dCas), and

a capped single guide RNA (Capped-sgRNA) comprising

an m7G cap,

a linker,

a spacer complementary to a target sequence in a messenger RNA, and

25

a direct repeat sequence, wherein the direct repeat sequence is capable of binding to the dCas protein.

Embodiment 20: The expression vector of Embodiment 19, wherein the Capped-sgRNA further comprises a Ribonuclease P (RNase P) processing site.

5 Embodiment 21: The expression vector of Embodiment 19, wherein the Capped-sgRNA further comprises a polyadenylated tail.

Embodiment 22: The expression vector of Embodiment 19, wherein the dCas and the Capped-sgRNA are under the control of the same promoter.

10

Embodiment 23: The expression vector of Embodiment 19, wherein the dCas and the Capped-sgRNA are under the control of different promoters.

Embodiment 24: A method of enhancing protein translation comprising:

15

(a) providing a nuclease dead Cas (dCas) protein, and

a capped single guide RNA (Capped-sgRNA) comprising from 5' to 3':

an m7G cap,

20

a linker,

a spacer complementary to a target sequence in a messenger RNA, and

a direct repeat sequence,

wherein the direct repeat sequence is capable of binding to the dCas protein,

and wherein the target sequence is proximal to a start codon of the messenger

25

RNA relative to a 5' m7G cap of the messenger RNA.

(b) allowing the Capped-sgRNA to bind to the target sequence and the dCas protein to bind to the direct repeat sequence of the Capped-sgRNA, thereby localizing the m7G cap of the Capped-sgRNA to the target sequence.

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other
5 aspects, advantages, and modifications are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A complex comprising:
 - a Cas polypeptide; and
 - a capped-sgRNA comprising
 - (i) an m7G cap or an analog thereof;
 - (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and
 - (iii) a direct repeat capable of binding to the Cas polypeptide.
2. The complex of claim 1, wherein the RNA molecule is a messenger RNA (mRNA).
3. The complex of claim 2, wherein the mRNA has an endogenous m7G cap.
4. The complex of claim 3, wherein the target sequence is downstream of the endogenous m7G cap of the mRNA.
5. The complex of any one of claims 2-4, wherein the mRNA comprises a start codon, and wherein the 5' end of the target sequence is upstream of the start codon.
6. The complex of claim 5, wherein the 5' end of the target sequence is between 1 to 50 nucleotides upstream of the first nucleotide of the start codon.
7. The complex of claim 6, wherein the 5' end of the target sequence is between 1 to 15 nucleotides upstream of the first nucleotide of the start codon.
8. The complex of any one of claims 5-7, wherein the target sequence comprises the start codon.
9. The complex of any one of claims 2-4, wherein the mRNA comprises a start codon, and wherein the 5' end of the target sequence is downstream of the start codon.
10. The complex of claim 9, wherein the 5' end of the target sequence is between 1 to 50 nucleotides downstream of the last nucleotide of the start codon.

11. The complex of claim 10, wherein the 5' end of the target sequence is between 1 to 15 nucleotides downstream of the last nucleotide of the start codon.
12. The complex of claim 10, wherein the 5' end of the target sequence is between 1 to 5 nucleotides downstream of the last nucleotide of the start codon.
13. The complex of any one of claims 1-12, wherein the spacer is at least 80% complementary to the target sequence.
14. The complex of claim 13, wherein the spacer is at least 90% complementary to the target sequence.
15. The complex of any one of claims 1-14, wherein the spacer comprises about 10 to about 40 nucleotides.
16. The complex of claim 15, wherein the spacer comprises about 15 to about 25 nucleotides.
17. The complex of claim 16, wherein the spacer comprises about 20 nucleotides.
18. The complex of any one of claims 1-17, wherein the spacer is connected to the m7G cap or analog thereof via a linker.
19. The complex of claim 18, wherein the linker comprises about 5 to about 25 nucleotides.
20. The complex of claim 19, wherein the linker comprises about 8 to about 15 nucleotides.
21. The complex of any one of claims 18-20, wherein the linker is not complementary to any sequence in the mRNA.
22. The complex of any one of claims 18-21, wherein the linker is conjugated to the m7G cap or analog thereof via polyethylene glycol.

23. The complex of any one of claims 1-22, wherein the Cas polypeptide is a nuclease-deficient Cas (dCas) polypeptide, wherein the dCas comprises an inactivated target cleavage domain and a retained guide cleavage domain.
24. The complex of claim 23, wherein the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d.
25. The complex of claim 24, wherein the direct repeat is capable of binding to a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d.
26. The complex of claim 23, wherein the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas9 (dCas9) polypeptide.
27. The complex of claim 26, wherein the direct repeat is capable of binding to a nuclease-deficient Cas9 (dCas9) polypeptide.
28. A nucleic acid comprising a sequence encoding the capped-sgRNA in the complex of any one of claims 1-27.
29. The nucleic acid of claim 28, further comprising a sequence encoding the Cas polypeptide in the complex of any one of claims 1-27.
30. A nucleic acid comprising a sequence encoding a capped-sgRNA, wherein the capped-sgRNA comprises:
- (i) an m7G cap or analog thereof;
 - (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and
 - (iii) a direct repeat capable of binding to a Cas polypeptide.
31. The nucleic acid of claim 30, wherein the RNA molecule is an mRNA.
32. The nucleic acid of claim 31, wherein the mRNA has an endogenous m7G cap.

33. The nucleic acid of claim 32, wherein the target sequence is downstream of the endogenous m7G cap of the mRNA.
34. The nucleic acid of any one of claims 31-33, wherein the mRNA comprises a start codon, and wherein the 5' end of the target sequence is upstream of the start codon.
35. The nucleic acid of claim 34, wherein the 5' end of the target sequence is between 1 and 50 nucleotides upstream of the first nucleotide of the start codon.
36. The nucleic acid of claim 35, wherein the 5' end of the target sequence is between 1 to 15 nucleotides upstream of the first nucleotide of the start codon.
37. The nucleic acid of any one of claims 34-36, wherein the target sequence comprises the start codon.
38. The nucleic acid of any one of claims 31-33, wherein the mRNA comprises a start codon, and wherein the 5' end of the target sequence is downstream of the start codon.
39. The nucleic acid of claim 38, wherein the 5' end of the target sequence is between 1 to 50 nucleotides downstream of the last nucleotide of the start codon.
40. The nucleic acid of claim 39, wherein the 5' end of the target sequence is between 1 to 15 nucleotides downstream of the last nucleotide of the start codon.
41. The nucleic acid of claim 40, wherein the 5' end of the target sequence is between 1 to 5 nucleotides downstream of the last nucleotide of the start codon.
42. The nucleic acid of any one of claims 30-41, wherein the spacer is at least 80% complementary to the target sequence.
43. The nucleic acid of claim 42, wherein the spacer is at least 90% complementary to the target sequence.

44. The nucleic acid of any one of claims 30-43, wherein the spacer comprises about 10 to about 40 nucleotides.
45. The nucleic acid of claim 44, wherein the spacer comprises about 15 to about 25 nucleotides.
46. The nucleic acid of claim 45, wherein the spacer comprises about 20 nucleotides.
47. The nucleic acid of any one of claims 30-46, wherein the spacer is connected to the m7G cap or analog thereof via a linker.
48. The nucleic acid of claim 47, wherein the linker comprises about 5 to about 25 nucleotides.
49. The nucleic acid of claim 48, wherein the linker comprises about 8 to about 20 nucleotides.
50. The nucleic acid of any one of claims 47-49, wherein the linker is not complementary to any sequence in the mRNA.
51. The nucleic acid of any one of claims 47-50, wherein the linker is conjugated to the m7G cap or analog thereof via polyethylene glycol.
52. The nucleic acid of any one of claims 30-51, further comprising a sequence encoding a RNase P processing site.
53. The nucleic acid of any one of claims 30-52, further comprising a poly-T sequence.
54. The nucleic acid of claim 52, further comprising a poly-T sequence downstream of the sequence encoding a RNase P processing site.
55. The nucleic acid of any one of claims 30-54, further comprising a sequence encoding the Cas polypeptide.

56. The nucleic acid of claim 55, wherein the Cas polypeptide is a nuclease-deficient Cas polypeptide.
57. The nucleic acid of claim 56, wherein the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d.
58. The nucleic acid of claim 57, wherein the direct repeat is capable of binding to a nuclease-deficient Cas13 (dCas13) polypeptide, wherein the dCas13 is dCas13b or dCas13d.
59. The nucleic acid of claim 56, wherein the nuclease-deficient Cas polypeptide is a nuclease-deficient Cas9 (dCas9) polypeptide.
60. The nucleic acid of any one of claims 30-59, further comprising one or more RNA polymerase II promoters.
61. The nucleic acid of any one of claims 55-60, wherein the sequence encoding the capped-sgRNA and the sequence encoding the Cas polypeptide are expressed from the same promoter.
62. The nucleic acid of any one of claims 55-60, wherein the sequence encoding the capped-sgRNA and the sequence encoding the Cas polypeptide are expressed from different promoters.
63. A vector comprising the nucleic acid of any one of claims 30-62.
64. The vector of claim 63, wherein the vector is an AAV vector.
65. A cell comprising the nucleic acid of any one of claims 30-52.
66. A method of regulating translation of an mRNA in a cell, the method comprising contacting the cell with a nucleic acid comprising
- (a) a sequence encoding a Cas polypeptide; and
 - (b) a sequence encoding a capped-sgRNA comprising
 - (i) an m7G cap or analog thereof;

- (ii) a spacer capable of specifically hybridizing with a target sequence in an RNA molecule; and
- (iii) a direct repeat capable of binding to the Cas polypeptide.

67. The method of claim 66, wherein the Cas polypeptide is a nuclease-deficient Cas13 (dCas13) polypeptide.

68. The method of claim 67, wherein the dCas13 polypeptide is dCas13b or dCas13d.

69. The method of claim 68, wherein the dCas13b comprises an inactivated target cleavage domain and a retained guide cleavage domain.

70. The method of claim 69, wherein the dCas13d comprises an inactivated target cleavage domain and a retained guide cleavage domain.

Fig. 1A

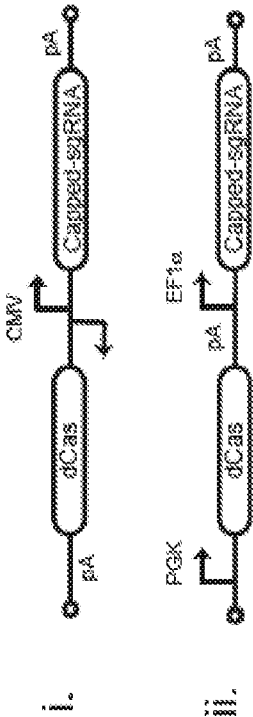
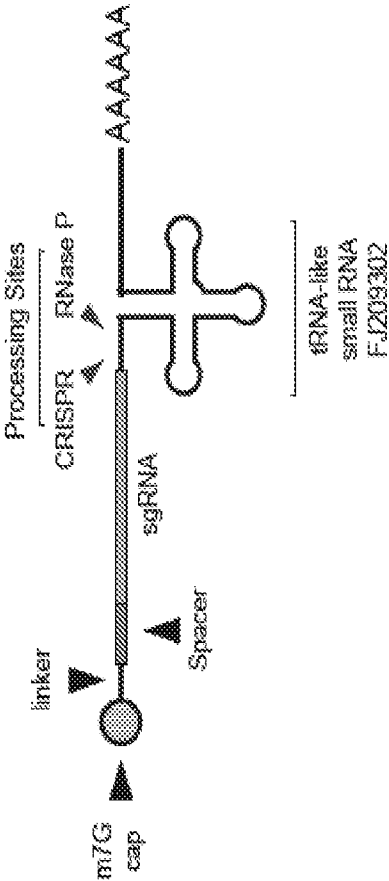
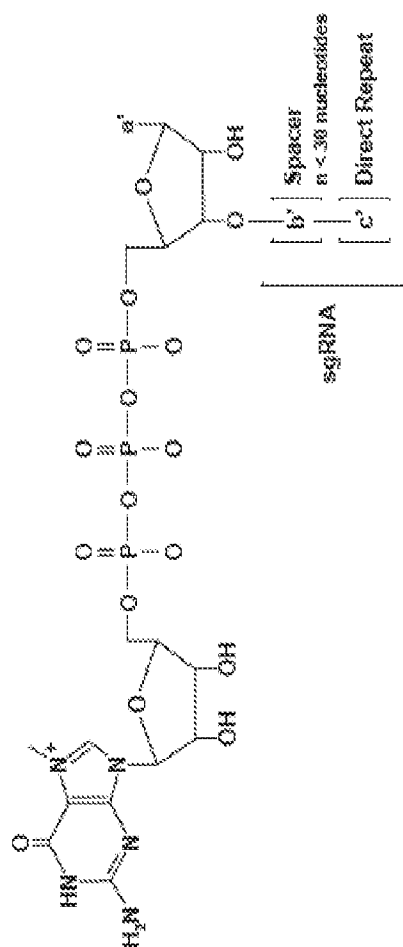


Fig. 1B



Q. 10. $\frac{1}{2}$ of a number is 10. What is the number?



De la

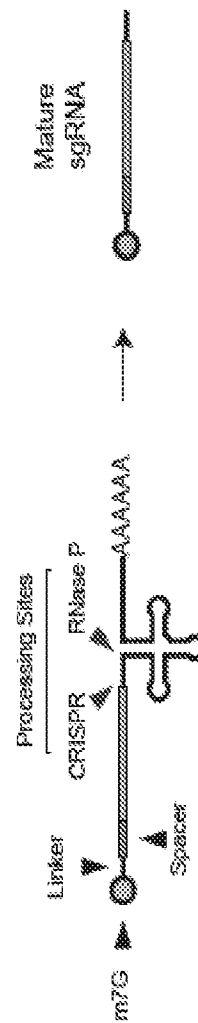


Fig. 1E

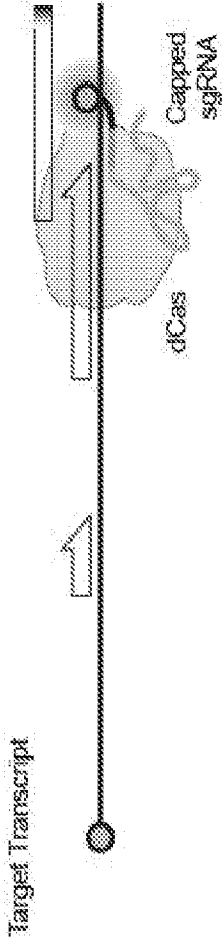


Fig. 1F

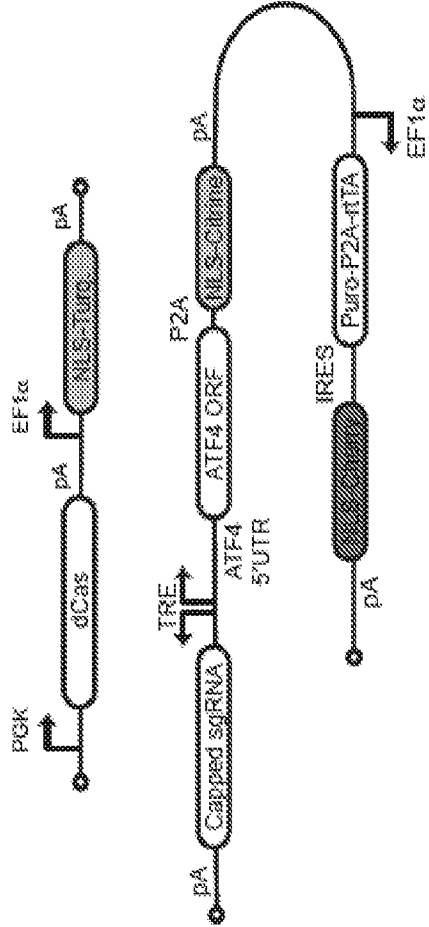
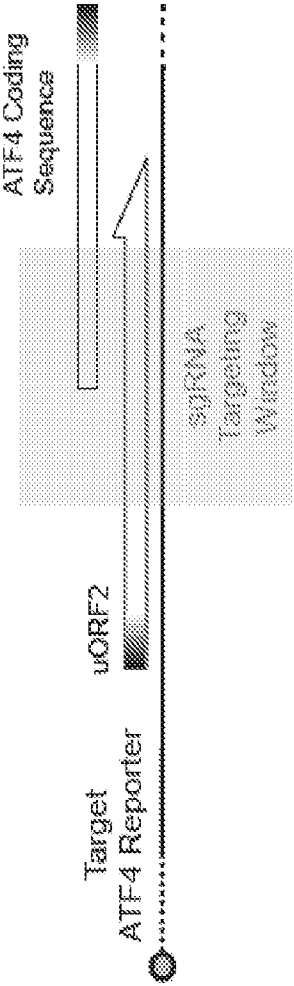
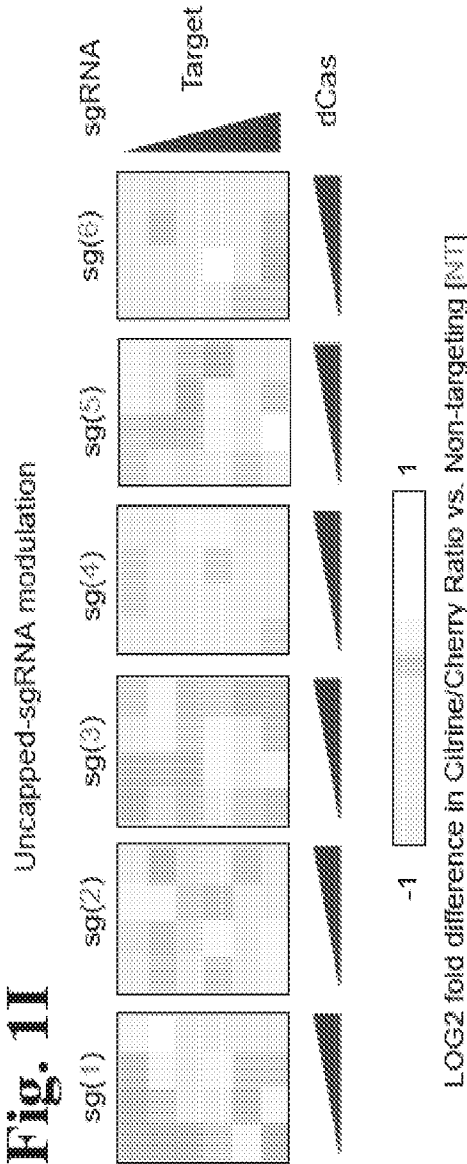
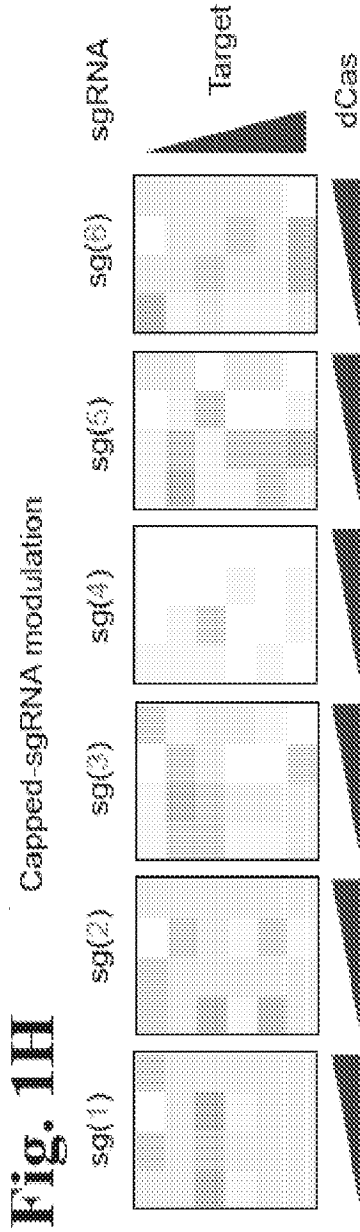


Fig. 1G





INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20/28546

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.: 8, 13-29, 37, 42-65
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:

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 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.:

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.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US20/28546

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 9/22, 15/11, 15/90; C12Q1/68 (2020.01)

CPC - C12N 9/22, 15/11, 15/902, 15/907; C12Q1/68

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	WO 2018/154387 A1 (CRISPR THERAPEUTICS AG) 30 August 2018; paragraphs [0008], [00064]-[00065], [000105], [000108]-[000109], [000172]	1-2, 30-31, 66 ----- 3-4, 5/2-4, 6/5/2-4, 7/6/5/2-4, 9/2-4, 10/9/2-4, 11/10/9/2-4, 12/10/9/2-4, 32-33, 34/31-33, 35/34/31-33, 36/35/34/31-33, 38/31-33, 39/38/31-33, 40/39/38/31-33, 41/40/39/38/31-33, 67-70
Y	(HINNEBUSCH, AG) Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. Microbiology and Molecular Biology Reviews. September 2011; Vol. 75, No. 3; pages 434-467; figure 1; DOI: 10.1128/MMBR.00008-11	3-4, 5/2-4, 6/5/2-4, 7/6/5/2-4, 9/2-4, 10/9/2-4, 11/10/9/2-4, 12/10/9/2-4, 32-33, 34/31-33, 35/34/31-33, 36/35/34/31-33, 38/31-33, 39/38/31-33, 40/39/38/31-33, 41/40/39/38/31-33
Y	WO 2019/060746 A1 (THE BROAD INSTITUTE et al.) 28 March 2019; paragraphs [0006], [0122], [0176]	67-70

☐

Further documents are listed in the continuation of Box C.

☐

See patent family annex.

* Special categories of cited documents:

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

"D" document cited by the applicant in the international application

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

18 June 2020 (18.06.2020)

Date of mailing of the international search report

14 JUL 2020

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

Telephone No. PCT Helpdesk: 571-272-4300