

(19) World Intellectual Property Organization
International Bureau



(10) International Publication Number
WO 2019/018423 A1

(43) International Publication Date
24 January 2019 (24.01.2019)

W!P O | PCT

(51) International Patent Classification:

C12N 9/22 (2006.01) C12N 15/11 (2006.01)
C12N 15/09 (2006.01) C12N 15/85 (2006.01)

(21) International Application Number:

PCT/US20 18/042529

(22) International Filing Date:

17 July 2018 (17.07.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/533,520 17 July 2017 (17.07.2017) US
62/566,815 02 October 2017 (02.10.2017) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(72) Inventors: **GOOTENBERG, Jonathan S.**; c/o 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **ABU-DAYYEH, Omar O.**; c/o 77 Massachusetts Ave., Cambridge, Massachusetts 02139 (US). **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) Agent: **NIX, F. Brent** et al; **JOHNSON, MARCOU & ISAACS, LLC**, P.O. Box 691, Hoschton, Georgia 30548 (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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available):

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

Published:

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

(54) Title: NOVEL TYPE VI CRISPR ORTHOLOGS AND SYSTEMS

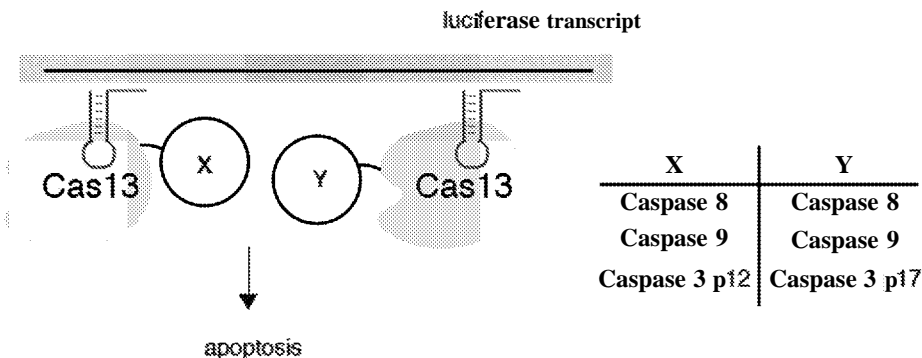


FIG. 1A

(57) Abstract: The present disclosure provides for systems, methods, and compositions for targeting nucleic acids. In particular, the disclosure provides non-naturally occurring or engineered RNA-targeting systems comprising a novel RNA-targeting CRISPR effector protein and at least one targeting nucleic acid component like a guide RNA.

WO 2019/018423 A1

NOVEL TYPE VI CRISPR ORTHOLOGS AND SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Nos. 62/533,520, filed on July 17, 2017, and 62/566,815, filed October 2, 2017. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant numbers MH1 00706 and MH1 10049 awarded by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] The present disclosure generally relates to systems, methods and compositions used for the control of gene expression involving sequence targeting, such as perturbation of gene transcripts or nucleic acid editing, that may use vector systems related to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

BACKGROUND

[0004] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome and transcriptome engineering technologies that employ novel strategies and molecular mechanisms and are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome and transcriptome. This would provide a major resource for new applications in genome engineering and biotechnology.

[0005] The CRISPR-Cas systems of bacterial and archaeal adaptive immunity show extreme diversity of protein composition and genomic loci architecture. The CRISPR-Cas system loci has more than 50 gene families and there is no strictly universal genes indicating

fast evolution and extreme diversity of loci architecture. So far, adopting a multi-pronged approach, there is comprehensive *cas* gene identification of about 395 profiles for 93 Cas proteins. Classification includes signature gene profiles plus signatures of locus architecture. A new classification of CRISPR-Cas systems is proposed in which these systems are broadly divided into two classes, Class 1 with multisubunit effector complexes and Class 2 with single-subunit effector modules exemplified by the Cas9 protein. Novel effector proteins associated with Class 2 CRISPR-Cas systems may be developed as powerful genome engineering tools and the prediction of putative novel effector proteins and their engineering and optimization is important.

[0006] The CRISPR-Cas adaptive immune system defends microbes against foreign genetic elements via DNA or RNA-DNA interference. Recently, the Class 2 type VI single-component CRISPR-Cas effector Cas13 (Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems"; *Molecular Cell* 60:1-13; doi: <http://dx.doi.org/10.1016/j.molcel.2015.10.008>) was characterized as an RNA-guided RNase (Abudayyeh et al. (2016), *Science*, [Epub ahead of print], June 2; "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector"; doi: 10.1126/science.aaf5573). It was demonstrated that C2c2 (e.g. from *Leptotrichia shahii*) provides robust interference against RNA phage infection. Through in vitro biochemical analysis and in vivo assays, it was shown that C2c2 can be programmed to cleave ssRNA targets carrying protospacers flanked by a 3' H (non-G) PAM. Cleavage is mediated by catalytic residues in the two conserved HEPN domains of C2c2, mutations in which generate a catalytically inactive RNA-binding protein. C2c2 is guided by a single guide and can be re-programmed to deplete specific mRNAs in vivo. It was shown that LshC2c2 can be targeted to a specific site of interest and can carry out non-specific RNase activity once primed with the cognate target RNA. These results broaden our understanding of CRISPR-Cas systems and demonstrate the possibility of harnessing C2c2 to develop a broad set of RNA-targeting tools.

[0007] C2c2 is now known as Cas13a. It will be understood that the term "C2c2" herein is used interchangeably with "Cas13a".

[0008] Citation or identification of any document in this application is not an admission that such document is available as prior art to the present invention.

SUMMARY

[0009] There exists a pressing need for alternative and robust systems and techniques for targeting nucleic acids or polynucleotides (e.g. DNA or RNA or any hybrid or derivative thereof) with a wide array of applications, in particular in eukaryotic systems, more in particular in mammalian systems. This invention addresses this need and provides related advantages. Adding the novel RNA-targeting systems of the present application to the repertoire of genomic, transcriptomic, and epigenomic targeting technologies may transform the study and perturbation or editing of specific target sites through direct detection, analysis and manipulation, in particular in eukaryotic systems, more in particular in mammalian systems (including cells, organs, tissues, or organisms) and plant systems. To utilize the RNA-targeting systems of the present application effectively for RNA targeting without deleterious effects, it is critical to understand aspects of engineering and optimization of these RNA targeting tools.

[0010] The CRISPR-Cas13 family was discovered by computational mining of bacterial genomes for signatures of CRISPR systems (Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385-397, doi:10.1016/j.molcel.2015.10.008 (2015)), revealing the single-effector RNA-guided RNase Cas13a/C2c2 (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)) and later the single-effector RNA-guided RNase Cas13b (Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017); Smargon, A. A. *et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618-630 e617, doi:10.1016/j.molcel.2016.12.023 (2017)). Applicants then extended the Cas13 family to include single-effector RNA-guided RNase Cas13c. The Class 2 type VI effector protein C2c2, also known as Cas13a, is a RNA-guided RNase that can be efficiently programmed to degrade ssRNA. C2c2 (Cas13a) achieves RNA cleavage through conserved basic residues within its two HEPN domains, in contrast to the catalytic mechanisms of other known RNases found in CRISPR-Cas systems. Mutation of the HEPN domain, such as (e.g. alanine) substitution, at any of the four predicted HEPN domain catalytic residues converted C2c2 into an inactive programmable RNA-binding protein (dC2c2, analogous to dCas9).

[0011] The programmability and specificity of the RNA-guided RNase Casl3 would make it an ideal platform for transcriptome manipulation. Applicants develop Casl3 for use as a transcript detection tool as well as a mammalian transcript knockdown and binding tool. Applicants extend sequence-specific detection to a method of transcript-based control of cellular mechanisms. In non-limiting examples of the method, transcript detection is linked to induction of apoptosis or to controlling expression of detectable markers.

[0012] Casl3a from *Leptotrichia shahii* (LshCasl3a) is capable of robust RNA cleavage and binding with catalytically inactive versions using programmable crRNAs and that cleavage was dependent on a directly 3'-adjacent motif known as the protospacer flanking site (PFS) with identity H (not guanine) (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)). Upon RNA cleavage, activated LshCasl3a engages in "collateral activity" in which constitutive RNase activity cleaves non-targeted RNAs (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)). This crRNA-programmed collateral activity enables *in vivo* programmed cell death by the bacteria to prevent spread of infection (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)) and has been applied *in vitro* for the specific detection of nucleic acid (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016); East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270-273, doi:10.1038/nature19802 (2016)). Collateral activity was recently leveraged for a highly sensitive and specific nucleic acid detection platform termed SHERLOCK that is useful for many clinical diagnoses (Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Casl3a/C2c2. *Science* **356**, 438-442 (2017)).

[0013] Via screening Casl3a orthologs in bacterial and subsequent biochemical characterization, Applicants select an ortholog optimized for RNA endonuclease activity, the Casl3a from *Leptotrichia wadeii* (LwaCasl3a). LwaCasl3a can be stably expressed in mammalian cells, retargeted to effectively knockdown both reporter and endogenous transcripts in cells, and attains levels of high levels of targeting specificity compared to RNAi without observable collateral activity. Furthermore, Applicants show that catalytically inactive LwaCasl3a (dCasl3a) programmably binds RNA transcripts *in vivo* and can be used

to image transcripts in cells. By engineering a negative-feedback imaging system based upon dCas13a, the formation of stress granules can be tracked in living cells.

[0014] Cas13 is also capable of robust RNA detection. In certain embodiments, Cas13 is converted to an RNA binding protein ("dead Cas13; dCas13) by inactivation of its nuclease activity. Converted to an RNA binding protein, Cas13 is useful for localizing other functional components to RNA in a sequence dependent manner. The components can be natural or synthetic. Applicants have used dCas13 to (i) bring effector modules to specific transcripts to modulate the function or translation, which could be used for large-scale screening, construction of synthetic regulatory circuits and other purposes; (ii) fluorescently tag specific RNAs to visualize their trafficking and/or localization; (iii) alter RNA localization through domains with affinity for specific subcellular compartments; and (iv) capture specific transcripts (through direct pull down of dCas13 or use of dCas13 to localize biotin ligase activity to specific transcripts) to enrich for proximal molecular partners, including RNAs and proteins. Applicants now demonstrate how to use dCas13 to i) organize components of a cell, ii) switch components or activities of a cell on or off, and iii) control cellular states based on the presence or amount of a specific transcript present in a cell. In exemplary embodiments, the invention provides split enzymes and reporter molecules, portions of which are provided in hybrid molecules comprising an RNA-binding CRISPR effector, such as, but not limited to Cas13. When brought into proximity in the presence of an RNA in a cell, activity of the split reporter or enzyme is reconstituted and the activity can then be measured. A split enzyme reconstituted in such manner can detectably act on a cellular component and/or pathway, including but not limited to an endogenous component or pathway, or exogenous component or pathway. A split reporter reconstituted in such manner can provide a detectable signal, such as but not limited to fluorescent or other detectable moiety. In certain embodiments, a split proteolytic enzyme is provided which when reconstituted acts on one or more components (endogenous or exogenous) in a detectable manner. In one exemplary embodiment, there is provided a method of inducing programmed cell death upon detection of an RNA species in a cell. It will be apparent how such a method could be used to ablate populations of cells, based for example, on the presence of virus in the cells.

[0015] In one aspect, the invention provides a method of identifying, measuring, and/or modulating the state of a cell or tissue based on the presence or level of a particular transcript in the cell or tissue. In one embodiment, the invention provides a CRISPR-based control system designed to modulate the presence and/or activity of a cellular system or component,

which may be a natural or synthetic system or component, based on the presence of a selected RNA species of interest. In general, the control system features an inactivated protein, enzyme or activity that is reconstituted when a selected RNA species of interest is present. In an embodiment of the invention, reconstituting an inactivated protein, enzyme or activity involves bringing together inactive components to assemble an active complex.

[0016] Accordingly, in one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a CRISPR protein linked to an inactive first portion of a proteolytic enzyme, wherein the proteolytic enzyme is activated when contacted or reconstituted with a complementary portion of the proteolytic enzyme. According to the invention, the complementary portion of the proteolytic enzyme is provided linked to a second CRISPR protein. Complementry means that taken together, the first portion and the second portion reconstitute function. In one such embodiment, a proteolytic enzyme split in two parts is provided. The enzyme may be split in any fashion such that the pieces of the enzyme possess little or no activity until contacted with one another. The enzyme can be split in multiple parts though a split into two parts is usually preferred, for example to minimize the number of CRISPR protein fusions. In certain embodiments, the parts taken together amount to the whole, i.e., the pieces of the protein or enzyme together make up a whole protein or enzyme. In certain embodiments, the pieces of the protein or enzyme together make up less than a whole protein or enzyme, e.g. where not all of the protein need be present in the reassembled pieces in order for the protein or enzyme to function. In certain embodiments, the pieces of the protein or enzyme together make up more than the whole protein or enzyme, e.g., where the component pieces comprise extra amino acids that contribute to stability and do not block function. In short, the split protein or enzyme can be provided in any configuration that is active once the pieces are reconstituted.

[0017] Usually, RNA binding CRISPR proteins are employed, although DNA-binding CRISPR proteins can be used where the intent is to detect DNA molecules in a cell. For example, the system can be used to detect viral DNA.

[0018] A system of the invention further includes guides for localizing the CRISPR proteins with linked enzyme portions on a transcript of interest that may be present in a cell or tissue. According, the system includes a first guide that binds to the first CRISPR protein and hybridizes to the transcript of interest and a second guide that binds to the second CRISPR protein and hybridizes to the transcript of interest. In most embodiments, it is preferred that the first and second guide hybridize to the transcript of interest at adjacent locations. The locations can be directly adjacent or separated by a few nucleotide, such as

separated by 1 nt, 2 nts, 3 nts, 4 nts, 5 nts, 6 nts, 7 nts, 8 nts, 9 nts, 10 nts, 11 nts, 12 nts, or more. In certain embodiments, the first and second guides can bind to locations separated on a transcript by an expected stem loop. Though separated along the linear transcript, the transcript may take on a secondary structure that brings the guide target sequences into close proximity.

[0019] In an embodiment of the invention, the proteolytic enzyme comprises a caspase. In an embodiment of the invention, the proteolytic enzyme comprises an initiator caspase, such as but not limited to caspase 8 or caspase 9. Initiator caspases are generally inactive as a monomer and gain activity upon homodimerization. In an embodiment of the invention, the proteolytic enzyme comprises an effector caspase, such as but not limited to caspase 3 or caspase 7. Such initiator caspases are normally inactive until cleaved into fragments. Once cleaved the fragments associate to form an active enzyme. The caspase fragments. In one exemplary embodiment, the first portion of the proteolytic enzyme comprises caspase 3 p12 and the complementary portion of the proteolytic enzyme comprises caspase 3 p17.

[0020] In an embodiment of the invention, the proteolytic enzyme is chosen to target a particular amino acid sequence and a substrate is chosen or engineered accordingly. A non-limiting example of such a protease is tobacco etch virus (TEV) protease. Accordingly, a substrate cleavable by TEV protease, which in some embodiments is engineered to be cleavable, serves as the system component acted upon by the protease. In one embodiment, the TEV protease substrate comprises a procaspase and one or more TEV cleavage sites. The procaspase can be, for example, caspase 3 or caspase 7 engineered to be cleavable by the reconstituted TEV protease. Once cleaved, the procaspase fragments are free to take on an active conformation.

[0021] In an embodiment of the invention, the TEV substrate comprises a fluorescent protein and a TEV cleavage site. In another embodiment, the TEV substrate comprises a luminescent protein and a TEV cleavage site. In certain embodiments, the TEV cleavage site provides for cleavage of the substrate such that the fluorescent or luminescent property of the substrate protein is lost upon cleavage. In certain embodiments, the fluorescent or luminescent protein can be modified, for example by appending a moiety which interferes with fluorescence or luminescence which is then cleaved when the TEV protease is reconstituted.

[0022] According to the invention, there is provided a method of providing a proteolytic activity in a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises a first CRISPR protein linked to an inactive first

portion of a proteolytic enzyme, and a second CRISPR protein linked to the complementary portion of the proteolytic enzyme wherein the activity of the proteolytic enzyme is reconstituted when the first portion and the complementary portion of the protein are contacted, and a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA, and a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA. When the target RNA of interest is present, the first and second portions of the proteolytic enzyme are contacted, the proteolytic activity of the enzyme is reconstituted, and a substrate of the enzyme is cleaved.

[0023] According to the invention, there is provided a method of inducing cell death in a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises a first CRISPR protein linked to an inactive first portion of a proteolytic enzyme capable of inducing cell death, a second CRISPR protein linked to the complementary portion of the enzyme wherein the enzyme activity of the proteolytic enzyme is reconstituted when the first portion and the complementary portion of the protein are contacted, and a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA, and a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA. When the target RNA of interest is present, the first and second portions of the proteolytic enzyme are contacted and the proteolytic activity of the enzyme is reconstituted and induces cell death. In one such embodiment of the invention, the proteolytic enzyme is a caspase. In another such embodiment, the proteolytic enzyme is TEV protease, wherein when the proteolytic activity of the TEV protease is reconstituted, a TEV protease substrate is cleaved and / or activated. In an embodiment of the invention, the TEV protease substrate is an engineered procaspase such that when the TEV protease is reconstituted, the procaspase is cleaved and activated, whereby apoptosis occurs. In an embodiment of the invention, a proteolytically cleavable transcription factor can be combined with any downstream reporter gene of choice to yield 'transcription-coupled' reporter systems. In an embodiment, a split protease is used to cleave or expose a degron from a detectable substrate.

[0024] According to the invention, there is provided a method of marking or identifying a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises a first CRISPR protein linked to an inactive first portion of a proteolytic enzyme, a second CRISPR protein linked to the complementary portion of the enzyme wherein the enzyme activity of the proteolytic enzyme is reconstituted when the first portion and the complementary portion of the protein are contacted, a first guide that binds to

the first CRISPR protein and hybridizes to a first target sequence of the RNA, a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA, and an indicator which is detectably cleaved by the reconstituted proteolytic enzyme. The first and second portions of the proteolytic enzyme are contacted when the RNA of interest is present in the cell, whereby the activity of the proteolytic enzyme is reconstituted and the indicator is detectably cleaved. In one such embodiment, the detectable indicator is a fluorescent protein, such as, but not limited to green fluorescent protein. In another such embodiment of the invention, the detectable indicator is a luminescent protein, such as, but not limited to luciferase. In an embodiment, the split reporter is based on reconstitution of split fragments of *Renilla reniformis* luciferase (Rluc). In an embodiment of the invention, the split reporter is based on complementation between two nonfluorescent fragments of the yellow fluorescent protein (YFP).

[0025] A number of applications have been demonstrated for active Cas13. In one aspect, Cas13 was used to targeting a specific transcript for destruction. In addition, Cas13, once primed by the cognate target, was shown to cleave other (non-complementary) RNA molecules in vitro and to inhibit cell growth in vivo. Biologically, this promiscuous RNase activity may reflect a programmed cell death/dormancy (PCD/D)-based protection mechanism of the type VI CRISPR-Cas systems. Accordingly, in an aspect of the invention, it might be used to trigger PCD or dormancy in specific cells—for example, cancer cells expressing a particular transcript, neurons of a given class, cells infected by a specific pathogen, or other aberrant cells or cells the presence of which is otherwise undesirable.

[0026] The invention provides a method of modifying nucleic acid sequences associated with or at a target locus of interest, in particular in eukaryotic cells, tissues, organs, or organisms, more in particular in mammalian cells, tissues, organs, or organisms, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Type VI CRISPR-Cas loci effector protein and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the sequences associated with or at the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break. In a preferred embodiment, the sequences associated with or at the target locus of interest comprises RNA and the effector protein is encoded by a type VI CRISPR-Cas loci. The complex can be formed in vitro or ex vivo and introduced into a cell or contacted with RNA; or can be formed in vivo.

[0027] It will be appreciated that the terms Cas enzyme, CRISPR enzyme, CRISPR protein Cas protein and CRISPR Cas are generally used interchangeably and at all points of reference herein refer by analogy to novel CRISPR effector proteins further described in this application, unless otherwise apparent, such as by specific reference to Cas9. The CRISPR effector proteins described herein are preferably C2c2 effector proteins.

[0028] The invention provides a method of targeting (such as modifying) sequences associated with or at a target locus of interest, the method comprising delivering to said sequences associated with or at the locus a non-naturally occurring or engineered composition comprising a C2c2 loci effector protein (which may be catalytically active, or alternatively catalytically inactive) and one or more nucleic acid components, wherein the C2c2 effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of sequences associated with or at the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break. In a preferred embodiment the C2c2 effector protein forms a complex with one nucleic acid component; advantageously an engineered or non-naturally occurring nucleic acid component. The complex can be formed in vitro or ex vivo and introduced into a cell or contacted with RNA; or can be formed in vivo. The induction of modification of sequences associated with or at the target locus of interest can be C2c2 effector protein-nucleic acid guided. In a preferred embodiment the one nucleic acid component is a CRISPR RNA (crRNA). In a preferred embodiment the one nucleic acid component is a mature crRNA or guide RNA, wherein the mature crRNA or guide RNA comprises a spacer sequence (or guide sequence) and a direct repeat sequence or derivatives thereof. In a preferred embodiment the spacer sequence or the derivative thereof comprises a seed sequence, wherein the seed sequence is critical for recognition and/or hybridization to the sequence at the target locus.

[0029] Aspects of the invention relate to C2c2 effector protein complexes having one or more non-naturally occurring or engineered or modified or optimized nucleic acid components. In a preferred embodiment the nucleic acid component of the complex may comprise a guide sequence linked to a direct repeat sequence, wherein the direct repeat sequence comprises one or more stem loops or optimized secondary structures. In certain embodiments, the direct repeat has a minimum length of 16 nts, such as at least 28 nt, and a single stem loop. In further embodiments the direct repeat has a length longer than 16 nts, preferably more than 17 nts, such as at least 28 nt, and has more than one stem loop or optimized secondary structures. In particular embodiments, the direct repeat has 25 or more

nts, such as 26 nt, 27 nt, 28 nt or more, and one or more stem loop structures. In a preferred embodiment the direct repeat may be modified to comprise one or more protein-binding RNA aptamers. In a preferred embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein. The bacteriophage coat protein may be selected from the group comprising QP, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, Mi1, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕ Cb5, (^Cb8r, (^Cb12r, (^Cb23r, 7s and PRR1. In a preferred embodiment the bacteriophage coat protein is MS2. The invention also provides for the nucleic acid component of the complex being 30 or more, 40 or more or 50 or more nucleotides in length.

[0030] The invention provides cells comprising the type VI effector protein and/or guides and or complexes thereof with target nucleic acids. In certain embodiments, the cell is a eukaryotic cell, including but not limited to a yeast cell, a plant cell, a mammalian cell, an animal cell, or a human cell.

[0031] The invention also provides a method of modifying a target locus of interest, in particular in eukaryotic cells, tissues, organs, or organisms, more in particular in mammalian cells, tissues, organs, or organisms, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a C2c2 loci effector protein and one or more nucleic acid components, wherein the C2c2 effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break. The complex can be formed *in vitro* or *ex vivo* and introduced into a cell or contacted with RNA; or can be formed *in vivo*.

[0032] In such methods the target locus of interest may be comprised within an RNA molecule. Also, the target locus of interest may be comprised within a DNA molecule, and in certain embodiments, within a transcribed DNA molecule. In such methods the target locus of interest may be comprised in a nucleic acid molecule *in vitro*.

[0033] In such methods the target locus of interest may be comprised in a nucleic acid molecule within a cell, in particular a eukaryotic cell, such as a mammalian cell or a plant cell. The mammalian cell may be a non-human primate, bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish or shrimp. The plant cell may be of a crop plant such as cassava, corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree or vegetable. The modification introduced to the cell by the

present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

[0034] The mammalian cell may be a non-human mammal, e.g., primate, bovine, ovine, porcine, canine, rodent, Leporidae such as monkey, cow, sheep, pig, dog, rabbit, rat or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry bird (e.g., chicken), vertebrate fish (e.g., salmon) or shellfish (e.g., oyster, clam, lobster, shrimp) cell. The cell may also be a plant cell. The plant cell may be of a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. The plant cell may also be of an algae, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus *Brassica*; plants of the genus *Lactuca*; plants of the genus *Spinacia*; plants of the genus *Capsicum*; cotton, tobacco, asparagus, carrot, cabbage, broccoli, cauliflower, tomato, eggplant, pepper, lettuce, spinach, strawberry, blueberry, raspberry, blackberry, grape, coffee, cocoa, etc).

[0035] The invention provides a method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Type VI CRISPR-Cas loci effector protein and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break.

[0036] The invention also provides a method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a C2c2 loci effector protein and one or more nucleic acid components, wherein the C2c2 effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break.

[0037] In such methods the target locus of interest may be comprised in a nucleic acid molecule *in vitro*. In such methods the target locus of interest may be comprised in a nucleic acid molecule within a cell. Preferably, in such methods the target locus of interest may be comprised in a RNA molecule *in vitro*. Also preferably, in such methods the target locus of

interest may be comprised in a RNA molecule within a cell. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The cell may be a rodent cell. The cell may be a mouse cell.

[0038] In any of the described methods the target locus of interest may be a genomic or epigenomic locus of interest. In any of the described methods the complex may be delivered with multiple guides for multiplexed use. In any of the described methods more than one protein(s) may be used.

[0039] In further aspects of the invention the nucleic acid components may comprise a CRISPR RNA (crRNA) sequence. Without limitation, the Applicants hypothesize that in such instances, the pre-crRNA may comprise secondary structure that is sufficient for processing to yield the mature crRNA as well as crRNA loading onto the effector protein. By means of example and not limitation, such secondary structure may comprise, consist essentially of or consist of a stem loop within the pre-crRNA, more particularly within the direct repeat.

[0040] In any of the described methods the effector protein and nucleic acid components may be provided via one or more polynucleotide molecules encoding the protein and/or nucleic acid component(s), and wherein the one or more polynucleotide molecules are operably configured to express the protein and/or the nucleic acid component(s). The one or more polynucleotide molecules may comprise one or more regulatory elements operably configured to express the protein and/or the nucleic acid component(s). The one or more polynucleotide molecules may be comprised within one or more vectors. In any of the described methods the target locus of interest may be a genomic or epigenomic locus of interest. In any of the described methods the complex may be delivered with multiple guides for multiplexed use. In any of the described methods more than one protein(s) may be used.

[0041] Regulatory elements may comprise inducible promoters. Polynucleotides and/or vector systems may comprise inducible systems.

[0042] In any of the described methods the one or more polynucleotide molecules may be comprised in a delivery system, or the one or more vectors may be comprised in a delivery system.

[0043] In any of the described methods the non-naturally occurring or engineered composition may be delivered via liposomes, particles including nanoparticles, exosomes, microvesicles, a gene-gun or one or more viral vectors.

[0044] The invention also provides a non-naturally occurring or engineered composition which is a composition having the characteristics as discussed herein or defined in any of the herein described methods.

[0045] In certain embodiments, the invention thus provides a non-naturally occurring or engineered composition, such as particularly a composition capable of or configured to modify a target locus of interest, said composition comprising a Type VI CRISPR-Cas loci effector protein and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In certain embodiments, the effector protein may be a Cas13 loci effector protein.

[0046] The invention also provides in a further aspect a non-naturally occurring or engineered composition, such as particularly a composition capable of or configured to modify a target locus of interest, said composition comprising: (a) a guide RNA molecule (or a combination of guide RNA molecules, e.g., a first guide RNA molecule and a second guide RNA molecule, such as for multiplexing) or a nucleic acid encoding the guide RNA molecule (or one or more nucleic acids encoding the combination of guide RNA molecules); (b) a Type VI CRISPR-Cas loci effector protein or a nucleic acid encoding the Type VI CRISPR-Cas loci effector protein. In certain embodiments, the effector protein may be a Cas13 loci effector protein.

[0047] The invention also provides in a further aspect a non-naturally occurring or engineered composition comprising: (a) a guide RNA molecule (or a combination of guide RNA molecules, e.g., a first guide RNA molecule and a second guide RNA molecule) or a nucleic acid encoding the guide RNA molecule (or one or more nucleic acids encoding the combination of guide RNA molecules); (b) be a Cas13 loci effector protein.

[0048] The invention also provides a vector system comprising one or more vectors, the one or more vectors comprising one or more polynucleotide molecules encoding components of a non-naturally occurring or engineered composition which is a composition having the characteristics as defined in any of the herein described methods.

[0049] The invention also provides a delivery system comprising one or more vectors or one or more polynucleotide molecules, the one or more vectors or polynucleotide molecules comprising one or more polynucleotide molecules encoding components of a non-naturally occurring or engineered composition which is a composition having the characteristics discussed herein or as defined in any of the herein described methods.

[0050] The invention also provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or transcriptome editing, or gene therapy.

[0051] The invention also provides for methods and compositions wherein one or more amino acid residues of the effector protein may be modified e.g., an engineered or non-naturally-occurring effector protein or Casl3. In an embodiment, the modification may comprise mutation of one or more amino acid residues of the effector protein. The one or more mutations may be in one or more catalytically active domains of the effector protein. The effector protein may have reduced or abolished nuclease activity compared with an effector protein lacking said one or more mutations. The effector protein may not direct cleavage of the RNA strand at the target locus of interest. In a preferred embodiment, the one or more mutations may comprise two mutations. In a preferred embodiment the one or more amino acid residues are modified in a Casl3 effector protein, e.g., an engineered or non-naturally-occurring effector protein or Casl3. In particular embodiments, the one or more modified or mutated amino acid residues are one or more of those in Casl3 corresponding to R597, H602, R1278 and H1283 (referenced to Lsh Casl3 amino acids), such as mutations R597A, H602A, R1278A and H1283A, or the corresponding amino acid residues in Lsh Casl3 orthologues.

[0052] In particular embodiments, the one or more modified or mutated amino acid residues are one or more of those in Casl3 corresponding to K2, K39, V40, E479, L514, V518, N524, G534, K535, E580, L597, V602, D630, F676, L709, 1713, R717 (HEPN), N718, H722 (HEPN), E773, P823, V828, 1879, Y880, F884, Y997, L1001, F1009, L1013, Y1093, L1099, L1111, Y1114, L1203, D1222, Y1244, L1250, L1253, K1261, 11334, L1355, L1359, R1362, Y1366, E1371, R1372, D1373, R1509 (HEPN), H1514 (HEPN), Y1543, D1544, K1546, K1548, V1551, 11558, according to Casl3 consensus numbering. In certain embodiments, the one or more modified or mutated amino acid residues are one or more of those in Casl3 corresponding to R717 and R1509. In certain embodiments, the one or more modified or mutated amino acid residues are one or more of those in Casl3 corresponding to K2, K39, K535, K1261, R1362, R1372, K1546 and K1548. In certain embodiments, said mutations result in a protein having an altered or modified activity. In certain embodiments, said mutations result in a protein having an increased activity, such as an increased specificity. In certain embodiments, said mutations result in a protein having a reduced

activity, such as reduced specificity. In certain embodiments, said mutations result in a protein having no catalytic activity (i.e. "dead" Casl3). In an embodiment, said amino acid residues correspond to Lsh Casl3 amino acid residues, or the corresponding amino acid residues of a Casl3 protein from a different species.

[0053] In certain embodiments the one or more modified of mutated amino acid residues are one or more of those in Casl3 corresponding to M35, K36, T38, K39, 157, E65, G66, L68, N84, T86, E88, 1103, N105, E123, R128, R129, K139, L152, L194, N196, K198, N201, Y222, D253, 1266, F267, S280, 1303, N306, R331, Y338, K389, Y390, K391, 1434, K435, L458, D459, E462, L463, 1478, E479, K494, R495, N498, S501, E519, N524, Y529, V530, G534, K535, Y539, T549, D551, R577, E580, A581, F582, 1587, A593, L597, 1601, L602, E611, E613, D630, 1631, G633, K641, N646, V669, F676, S678, N695, E703, A707, 1709, 1713, 1716, R717, H722, F740, F742, K768, 1774, K778, 1783, L787, S789, V792, Y796, D799, F812, N818, P820, F821, V822, P823, S824, F825, Y829, K831, D837, L852, F858, E867, A871, L875, K877, Y880, Y881, F884, F888, F896, N901, V903, N915, K916, R918, Q920, E951, P956, Y959, Q964, 1969, N994, F1000, 110001, Q1003, F10005, K1007, G1008, F1009, N1019, L1020, K1021, 11023, N1028, E1070, 11075, K1076, F1092, K1097, L1099, L1104, L1107, K1113, Y1114, E1149, E1151, 11153, L1155, L1158, D1166, L1203, D1222, G1224, 11228, R1236, K1243, Y1244, G1245, D1255, K1261, S1263, L1267, E1269, K1274, 11277, E1278, L1289, H1290, A1294, N1320, K1325, E1327, Y1328, 11334, Y1337, K1341, N1342, K1343, N1350, L1352, L1355, L1356, 11359, L1360, R1362, V1363, G1364, Y1365, 11369, R1371, D1372, F1385, E1391, D1459, K1463, K1466, R1509, N1510, 11512, A1513, H1514, N1516, Y1517, L1529, L1530, E1534, L1536, R1537, Y1543, D1544, R1545, K1546, L1547, K1548, N1549, A1550, K1553, S1554, D1557, 11558, L1559, G1563, F1568, 11612, L1651, E1652, K1655, H1658, L1659, K1663, T1673, S1677, E1678, E1679, C1681, V1684, K1685, E1689 with reference to the consensus sequence, i.e. based on the alignment of *Leptotrichia wadei* F0279 ("Lew2" or "Lw2") and *Listeria newyorkensis* FSL M6-0635 (also known as *Listeriaceae* bacterium FSL M6-0635 ("Lib" or "LbFSL")). As indicated earlier, in certain embodiments, in the above amino acid residue list, the residues corresponding to R597, H602, R1278 and H1283 (referenced to Lsh Casl3 amino acids) are excluded.

[0054] In certain embodiments, the one or more modified of mutated amino acid residues are one or more conserved charged amino acid residues. In certain embodiments, said amino acid residues may be mutated to alanine.

[0055] In certain embodiments the one or more modified or mutated amino acid residues are one or more of those in Casl3 corresponding to K28, K31, R44, E162, E184, K262, E288, K357, E360, K338, R441 (HEPN), H446 (HEPN), E471, K482, K525, K558, D707, R790, K811, R833, E839, R885, E894, R895, D896, K942, R960 (HEPN), H965 (HEPN), D990, K992, K994 with reference to the consensus sequence, i.e. based on the alignment of the Casl3 orthologues. As indicated earlier, in certain embodiments, in the above amino acid residue list, the residues corresponding to R597, H602, R1278 and H1283 (referenced to Lsh Casl3 amino acids) are excluded.

[0056] The invention also provides for the one or more mutations or the two or more mutations to be in a catalytically active domain of the effector protein. In certain embodiments, the one or more mutations or the two or more mutations may be in a catalytically active domain of the effector protein comprising a HEPN domain, or a catalytically active domain which is homologous to a HEPN domain. The effector protein may comprise one or more heterologous functional domains. The one or more heterologous functional domains may comprise one or more nuclear localization signal (NLS) domains. The one or more heterologous functional domains may comprise at least two or more NLS domains. The one or more NLS domain(s) may be positioned at or near or in proximity to a terminus of the effector protein (e.g., Casl3) and if two or more NLSs, each of the two may be positioned at or near or in proximity to a terminus of the effector protein (e.g., Casl3). The one or more heterologous functional domains may comprise one or more translational activation domains. In other embodiments the functional domain may comprise a transcriptional activation domain, for example VP64. The one or more heterologous functional domains may comprise one or more transcriptional repression domains. In certain embodiments the transcriptional repression domain comprises a KRAB domain or a SID domain (e.g. SID4X). The one or more heterologous functional domains may comprise one or more nuclease domains. In a preferred embodiment a nuclease domain comprises FokI.

[0057] The invention also provides for the one or more heterologous functional domains to have one or more of the following activities: methylase activity, demethylase activity, translation activation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity and nucleic acid binding activity. In certain embodiments of the invention, the one or more heterologous functional domains may comprise epitope tags or reporters. Non-limiting

examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporters include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP).

[0058] At least one or more heterologous functional domains may be at or near the amino-terminus of the effector protein and/or wherein at least one or more heterologous functional domains is at or near the carboxy-terminus of the effector protein. The one or more heterologous functional domains may be fused to the effector protein. The one or more heterologous functional domains may be tethered to the effector protein. The one or more heterologous functional domains may be linked to the effector protein by a linker moiety.

[0059] The invention also provides for the effector protein comprising an effector protein from an organism from a genus comprising *Streptococcus*, *Campylobacter*, *Nitratifactor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium*, *Corynebacter*, *Carnobacterium*, *Rhodobacter*, *Listeria*, *Paludibacter*, *Clostridium*, *Lachnospiraceae*, *Clostridiaridium*, *Leptotrichia*, *Francisella*, *Legionella*, *Alicyclobacillus*, *Methanomethyophilus*, *Porphyromonas*, *Prevotella*, *Bacteroidetes*, *Helcococcus*, *Letospira*, *Desulfovibrio*, *Desulfonatronum*, *Opitutaceae*, *Tuberibacillus*, *Bacillus*, *Brevibacillus*, *Methylobacterium* or *Acidaminococcus*. The effector protein may comprise a chimeric effector protein comprising a first fragment from a first effector protein ortholog and a second fragment from a second effector protein ortholog, and wherein the first and second effector protein orthologs are different. At least one of the first and second effector protein orthologs may comprise an effector protein from an organism comprising *Streptococcus*, *Campylobacter*, *Nitratifactor*, *Staphylococcus*, *Parvibaculum*, *Roseburia*, *Neisseria*, *Gluconacetobacter*, *Azospirillum*, *Sphaerochaeta*, *Lactobacillus*, *Eubacterium*, *Corynebacter*, *Carnobacterium*, *Rhodobacter*, *Listeria*, *Paludibacter*, *Clostridium*, *Lachnospiraceae*, *Clostridiaridium*, *Leptotrichia*, *Francisella*, *Legionella*, *Alicyclobacillus*, *Methanomethyophilus*, *Porphyromonas*, *Prevotella*, *Bacteroidetes*, *Helcococcus*, *Letospira*, *Desulfovibrio*, *Desulfonatronum*, *Opitutaceae*, *Tuberibacillus*, *Bacillus*, *Brevibacillus*, *Methylobacterium* or *Acidaminococcus*.

[0060] In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Cas13p, may originate from, may be isolated from, or may be

derived from a bacterial species belonging to the taxa alpha-proteobacteria, Bacilli, Clostridia, Fusobacteria and Bacteroidetes. In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Casl3p, may originate from, may be isolated from, or may be derived from a bacterial species belonging to a genus selected from the group consisting of *Lachnospiraceae*, *Clostridium*, *Carnobacterium*, *Paludibacter*, *Listeria*, *Leptotrichia*, and *Rhodobacter*. In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Casl3p may originate from, may be isolated from or may be derived from a bacterial species selected from the group consisting of *Lachnospiraceae* bacterium MA2020, *Lachnospiraceae* bacterium NK4A 179, *Clostridium aminophilum* (e.g., DSM 10710), *Lachnospiraceae* bacterium NK4A 144, *Carnobacterium gallinarum* (e.g., DSM 4847 strain MT44), *Paludibacter propionicigenes* (e.g., WB4), *Listeria seeligeri* (e.g., serovar 1/2b str. SLCC3954), *Listeria weihenstephanensis* (e.g., FSL R9-0317 c4), *Listeria newyorkensis* (e.g., strain FSL M6-0635: also "LbFSL"), *Leptotrichia wadei* (e.g., F0279: also "Lw" or "Lw2"), *Leptotrichia buccalis* (e.g., DSM 1135), *Leptotrichia* sp. Oral taxon 225 (e.g., str. F0581), *Leptotrichia* sp. Oral taxon 879 (e.g., strain F0557), *Leptotrichia shahii* (e.g., DSM 19757), *Rhodobacter capsulatus* (e.g., SB 1003, R121, or DE442). In certain preferred embodiments, the Casl3 effector protein originates from *Listeriaceae* bacterium (e.g. FSL M6-0635: also "LbFSL"), *Lachnospiraceae* bacterium MA2020, *Lachnospiraceae* bacterium NK4A 179, *Clostridium aminophilum* (e.g., DSM 10710), *Carnobacterium gallinarum* (e.g., DSM 4847), *Paludibacter propionicigenes* (e.g., WB4), *Listeria seeligeri* (e.g., serovar 1/2b str. SLCC3954), *Listeria weihenstephanensis* (e.g., FSL R9-0317 c4), *Leptotrichia wadei* (e.g., F0279: also "Lw" or "Lw2"), *Leptotrichia shahii* (e.g., DSM 19757), *Rhodobacter capsulatus* (e.g., SB 1003, R121, or DE442); preferably *Listeriaceae* bacterium FSL M6-0635 (i.e. *Listeria newyorkensis* FSL M6-0635) or *Leptotrichia wadei* F0279 (also "Lw" or "Lw2").

[0061] In certain embodiments, a Type VI locus as intended herein may encode Casl, Cas2, and the Casl3p effector protein.

[0062] In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Casl3p, such as a native Casl3p, may be about 1000 to about 1500 amino acids long, such as about 1100 to about 1400 amino acids long, e.g., about 1000 to about 1100, about 1100 to about 1200 amino acids long, or about 1200 to about 1300 amino acids long, or about 1300 to about 1400 amino acids long, or about 1400 to about 1500

amino acids long, e.g., about 1000, about 1100, about 1200, about 1300, about 1400 or about 1500 amino acids long.

[0063] In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Cas13p, comprises at least one and preferably at least two, such as more preferably exactly two, conserved RxxxxH motifs. Catalytic RxxxxH motifs are characteristic of HEPN (Higher Eukaryotes and Prokaryotes Nucleotide-binding) domains. Hence, in certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Cas13p, comprises at least one and preferably at least two, such as more preferably exactly two, HEPN domains. In certain embodiments, the HEPN domains may possess RNase activity. In other embodiments, the HEPN domains may possess DNase activity.

[0064] In certain embodiments, Type VI loci as intended herein may comprise CRISPR repeats between 30 and 40 bp long, more typically between 35 and 39 bp long, e.g., 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 bp long. In particular embodiments, the direct repeat is at least 25 nt long.

[0065] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the effector protein complex as disclosed herein to the target locus of interest. In some embodiments, the PAM may be a 5' PAM (i.e., located upstream of the 5' end of the protospacer). In other embodiments, the PAM may be a 3' PAM (i.e., located downstream of the 5' end of the protospacer). The term "PAM" may be used interchangeably with the term "PFS" or "protospacer flanking site" or "protospacer flanking sequence".

[0066] In a preferred embodiment, the effector protein, particularly a Type VI loci effector protein, more particularly a Cas13p, may recognize a 3' PAM. In certain embodiments, the effector protein, particularly a Type VI loci effector protein, more particularly a Cas13p, may recognize a 3' PAM which is 5H, wherein H is A, C or U. In certain embodiments, the effector protein may be *Leptotrichia shahii* Cas13p, more preferably *Leptotrichia shahii* DSM 19757 Cas13, and the 5' PAM is a 5' H. In certain embodiments, the effector protein may be *Leptotrichia wadei* F0279 (Lw2) Cas13, and the 5'PAM is H, wherein H is C, U or A.

[0067] In certain embodiments, the CRISPR enzyme is engineered and can comprise one or more mutations that reduce or eliminate a nuclease activity. Mutations can also be made at neighboring residues, e.g., at amino acids near those indicated above that participate in the nuclease activity. In some embodiments, only one HEPN domain is inactivated, and in other embodiments, a second HEPN domain is inactivated.

[0068] In certain embodiments of the invention, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In certain embodiments the guide RNA or mature crRNA comprises 19 nts of partial direct repeat followed by 18, 19, 20, 21, 22, 23, 24, 25, or more nt of guide sequence, such as 18-25, 19-25, 20-25, 21-25, 22-25, or 23-25 nt of guide sequence or spacer sequence. In certain embodiments, the effector protein is a Cas13 effector protein and requires at least 16 nt of guide sequence to achieve detectable DNA cleavage and a minimum of 17 nt of guide sequence to achieve efficient DNA cleavage *in vitro*. In particular embodiments, the effector protein is a Cas13 protein and requires at least 19 nt of guide sequence to achieve detectable RNA cleavage. In certain embodiments, the direct repeat sequence is located upstream (i.e., 5') from the guide sequence or spacer sequence. In a preferred embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the Cas13 guide RNA is approximately within the first 5 nt on the 5' end of the guide sequence or spacer sequence.

[0069] In preferred embodiments of the invention, the mature crRNA comprises a stem loop or an optimized stem loop structure or an optimized secondary structure. In preferred embodiments the mature crRNA comprises a stem loop or an optimized stem loop structure in the direct repeat sequence, wherein the stem loop or optimized stem loop structure is important for cleavage activity. In certain embodiments, the mature crRNA preferably comprises a single stem loop. In certain embodiments, the direct repeat sequence preferably comprises a single stem loop. In certain embodiments, the cleavage activity of the effector protein complex is modified by introducing mutations that affect the stem loop RNA duplex structure. In preferred embodiments, mutations which maintain the RNA duplex of the stem loop may be introduced, whereby the cleavage activity of the effector protein complex is maintained. In other preferred embodiments, mutations which disrupt the RNA duplex structure of the stem loop may be introduced, whereby the cleavage activity of the effector protein complex is completely abolished.

[0070] In particular embodiments, the Cas13 protein is an Lsh Cas13 effector protein and the mature crRNA comprises a stem loop or an optimized stem loop structure. In particular embodiments, the direct repeat of the crRNA comprises at least 25 nucleotides comprising a stem loop. In particular embodiments, the the stem is amenable to individual base swaps but activity is disrupted by most secondary structure changes or truncation of the crRNA.

Examples of disrupting mutations include swapping of more than two of the stem nucleotides, addition of a non-pairing nucleotide in the stem, shortening of the stem (by removal of one of the pairing nucleotides) or extending the stem (by addition of one set of pairing nucleotides). However, the crRNA may be amenable to 5' and/or 3' extensions to include non-functional RNA sequences as envisaged for particular applications described herein.

[0071] The invention also provides for the nucleotide sequence encoding the effector protein being codon optimized for expression in a eukaryote or eukaryotic cell in any of the herein described methods or compositions. In an embodiment of the invention, the codon optimized nucleotide sequence encoding the effector protein encodes any Cas13 discussed herein and is codon optimized for operability in a eukaryotic cell or organism, e.g., such cell or organism as elsewhere herein mentioned, for instance, without limitation, a yeast cell, or a mammalian cell or organism, including a mouse cell, a rat cell, and a human cell or non-human eukaryote organism, e.g., plant.

[0072] In certain embodiments of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the Cas13 effector proteins. In preferred embodiments at least one or more C-terminal or N-terminal NLSs are attached (and hence nucleic acid molecule(s) coding for the Cas13 effector protein can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In certain embodiments of the invention, at least one nuclear export signal (NES) is attached to the nucleic acid sequences encoding the Cas13 effector proteins. In preferred embodiments at least one or more C-terminal or N-terminal NESs are attached (and hence nucleic acid molecule(s) coding for the Cas13 effector protein can include coding for NES(s) so that the expressed product has the NES(s) attached or connected). In a preferred embodiment a C-terminal and/or N-terminal NLS or NES is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. In a preferred embodiment, the codon optimized effector protein is Cas13 and the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 16 nucleotides, such as at least 17 nucleotides, preferably at least 18 nt, such as preferably at least 19 nt, at least 20 nt, at least 21 nt, or at least 22 nt. In certain embodiments, the spacer length is from 15 to 17 nt, from 17 to 20 nt, from 20 to 24 nt, eg. 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, from 27-30 nt, from 30-35 nt, or 35 nt or longer. In certain embodiments of the invention, the codon optimized effector protein is Cas13 and the direct repeat length of the guide RNA is at least 16 nucleotides. In certain embodiments, the

codon optimized effector protein is Cas13 and the direct repeat length of the guide RNA is from 16 to 20 nt, e.g., 16, 17, 18, 19, or 20 nucleotides. In certain preferred embodiments, the direct repeat length of the guide RNA is 19 nucleotides.

[0073] The invention also encompasses methods for delivering multiple nucleic acid components, wherein each nucleic acid component is specific for a different target locus of interest thereby modifying multiple target loci of interest. The nucleic acid component of the complex may comprise one or more protein-binding RNA aptamers. The one or more aptamers may be capable of binding a bacteriophage coat protein. The bacteriophage coat protein may be selected from the group comprising QP, F2, GA, fr, JP501, MS2, M12, R17, BZ13, JP34, JP500, KU1, Mi1, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕ Cb5, (λ)Cb8r, (λ)Cb12r, (λ)Cb23r, 7s and PRR1. In a preferred embodiment the bacteriophage coat protein is MS2. The invention also provides for the nucleic acid component of the complex being 30 or more, 40 or more or 50 or more nucleotides in length.

[0074] Accordingly, it is an object of the invention not to encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. § 112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product. It may be advantageous in the practice of the invention to be in compliance with Art. 53(c) EPC and Rule 28(b) and (c) EPC. Nothing herein is to be construed as a promise.

[0075] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[0076] In a further aspect, the invention provides a eukaryotic cell comprising a nucleotide sequence encoding the CRISPR system described herein which ensures the generation of a modified target locus of interest, wherein the target locus of interest is

modified according to in any of the herein described methods. A further aspect provides a cell line of said cell. Another aspect provides a multicellular organism comprising one or more said cells.

[0077] In certain embodiments, the modification of the target locus of interest may result in: the eukaryotic cell comprising altered (protein) expression of at least one gene product; the eukaryotic cell comprising altered (protein) expression of at least one gene product, wherein the (protein) expression of the at least one gene product is increased; the eukaryotic cell comprising altered (protein) expression of at least one gene product, wherein the (protein) expression of the at least one gene product is decreased; or the eukaryotic cell comprising an edited transcriptome.

[0078] In certain embodiments, the eukaryotic cell may be a mammalian cell or a human cell.

[0079] In further embodiments, the non-naturally occurring or engineered compositions, the vector systems, or the delivery systems as described in the present specification may be used for RNA sequence-specific interference, RNA sequence specific modulation of expression (including isoform specific expression), stability, localization, functionality (e.g. ribosomal RNAs or miRNAs), etc.; or multiplexing of such processes.

[0080] In further embodiments, the the non-naturally occurring or engineered compositions, the vector systems, or the delivery systems as described in the present specification may be used for RNA detection and/or quantification in a sample, such as a biological sample. In certain embodiments, RNA detection is in a cell. In an embodiment, the invention provides a method of detecting a target RNA in a sample, comprising (a) incubating the sample with i) a Type VI CRISPR-Cas effector protein capable of cleaving RNA, ii) a guide RNA capable of hybridizing to the target RNA, and iii) an RNA-based cleavage inducible reporter capable of being non-specifically and detectably cleaved by the effector protein, (b) detecting said target RNA based on the signal generated by cleavage of said RNA-based cleavage inducible reporter.

[0081] In an embodiment the Type VI CRISPR-Cas effector protein comprises a Cas13 effector protein. In an embodiment, the RNA-based cleavage inducible reporter construct comprises a fluorochrome and a quencher. In certain embodiments, the sample comprises a cell-free biological sample. In other embodiments, the sample comprises or a cellular sample, for example, without limitation a plant cell, or an animal cell. In an embodiment of the invention, the target RNA comprises a pathogen RNA, including, but not limited to a target RNA from a virus, bacteria, fungus, or parasite. In an embodiment, the guide RNA is

designed to detect a target RNA which comprises a single nucleotide polymorphism or a splice variant of an RNA transcript. In an embodiment, the guide RNA comprises one or more mismatched nucleotides with the target RNA. In certain embodiments, the guide RNA hybridizes to a target molecule that is diagnostic for a disease state, such as, but not limited to, cancer, or an immune disease.

[0082] The invention provides a ribonucleic acid (RNA) detection system, comprising a) a Type VI CRISPR-Cas effector protein capable of cleaving RNA, b) a guide RNA capable of binding to a target RNA, and c) an RNA-based cleavage inducible reporter capable of being non-specifically and detectably cleaved by the effector protein. Further, the invention provides a kit for RNA detection, which comprises a) a Type VI CRISPR-Cas effector protein capable of cleaving RNA, and b) an RNA-based cleavage inducible reporter capable of being non-specifically and detectably cleaved by the effector protein. In certain embodiments, the RNA-based cleavage inducible reporter construct comprises a fluorochrome and a quencher.

[0083] In further embodiments, the the non-naturally occurring or engineered compositions, the vector systems, or the delivery systems as described in the present specification may be used for generating disease models and/or screening systems.

[0084] In further embodiments, the non-naturally occurring or engineered compositions, the vector systems, or the delivery systems as described in the present specification may be used for: site-specific transcriptome editing or perturbation; nucleic acid sequence-specific interference; or multiplexed genome engineering.

[0085] Also provided is a gene product from the cell, the cell line, or the organism as described herein. In certain embodiments, the amount of gene product expressed may be greater than or less than the amount of gene product from a cell that does not have altered expression or edited genome. In certain embodiments, the gene product may be altered in comparison with the gene product from a cell that does not have altered expression or edited genome.

[0086] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0087] The novel features of the invention are set forth with particularity in the appended claims. A understanding of the features and advantages of the present invention will be

obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0088] Figs. 1A-1B. Inducible apoptosis. (Fig. 1A) The cartoon depicts caspase activation by dimerization. Caspase 8 and Caspase 9 exemplify initiator caspases, which are found as monomers at physiological concentrations and may dimerize to become active. The cartoon depicts dimerization wherein caspases are maintained in proximity by association with Cas13 complexes bound to a luciferase transcript. Caspase 3 exemplifies effector caspases which may be found as stable, but inactive, dimers at physiological concentration. Activation of these caspases depends on proteolytic cleavage which allows the active site to rearrange. The cartoon depicts the Caspase 3 fragments p12 and p17 maintained in proximity by association with Cas13 complexes bound to a luciferase transcript. (Fig. 1B) The cartoon depicts caspase activation with an engineered tobacco etch virus (TEV) protease. Inactive N-terminal and C-terminal fragments of TEV protease are provided. TEV protease activity is reconstituted by maintaining the N-terminal and C-terminal fragments in proximity through association with Cas13 complexes bound to a luciferase transcript.

[0089] Fig. 2. Guide proximity. Guides for positioning Cas13 complexes on a luciferase transcript are depicted. (SEQ ID Nos. 167-169)

[0090] Fig. 3. Inducible apoptosis. Guides depicted in Fig. 2 were used to locate Cas13 complexes bearing functional domains to induce apoptosis along a luciferase transcript. Guide pairs 1-6 indicate the seed guide paired with each of guides 1-6. Caspase 8 and Caspase 9: caspase activity is induced when caspase 8 or caspase 9 enzymes attached to Cas13 are maintained in proximity by Cas13 complex formation on a luciferase transcript. SNIPPER Caspase 7 and SNIPPER Caspase 3: caspase activity is induced when Cas13 complexes bearing TEV N-terminal and C-terminal are maintained in proximity, activating the TEV protease activity leading to cleavage and activation of caspase 7 or caspase 3 pro-proteins. Split Caspase 3: The activity of split caspase 3 is reconstituted when the fragments are maintained in proximity by attachment to Cas13 complexes with a luciferase transcript.

[0091] Figs. 4A-4D. Comparison of dimerization and TEV protease ("SNIPPER") approaches. Apoptosis induced by TEV-dependent activation of caspase 7 or caspase 3, or by dimerization of caspase 8 or caspase 9 was compared. Guide pairs 1-6 indicate the seed guide paired with each of guides 1-6.

[0092] Figs. 5A-5I. Comparison of caspase variants. (Figs. 5A-5C) Cell death data is normalized to cell survival and shown relative to the non-targeting condition for all four

caspase variants (Fig. 5A) as and SNIPPER variants separately (Figs. 5B, 5C). (D-F) Raw cell death data relative to the non-targeting condition is shown, demonstrating which guide pairs yield the most effective cell death. (Figs. 5G-5I) Caspase variants are compared by cell death ratio for all four caspase variants (Fig. 5G) as and SNIPPER variants separately (Figs. 5H, 5I).

[0093] Figs. 6A-6L. (Figs. 6A-6K) Sequence alignment of Cas13 orthologs. (Fig. L) Sequence alignment of HEPN domains. (Additional SEQ ID Nos. 170 and 171)

[0094] Figs. 7A-70. Alignment of sequences of Cas13 orthologs of Figs. 6A-6L with consensus sequence indicated.

[0095] Figs. 8A-8C. Alignment of *Leptotrichia wadei* F0279 Cas13 ("Lew2C2c2") and *Listeria newyorkensis* FSL M6-0635 Cas13 ("LibC2c2").

[0096] Figs. 9A-9B. RNA binding by truncations of dCas13b. Various N-terminal and C-terminal truncations of dCas13b are depicted. RNA binding is indicated where there is ADAR-dependent RNA editing as measured by restoration of luciferase signal, comparing activity using targeting and non-targeting guides. Amino acid positions correspond to amino acid positions of *Prevotella* sp. P5-125 Cas13b protein.

[0097] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

[0098] General Definitions

[0099] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in *Molecular Cloning: A Laboratory Manual*, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); *Molecular Cloning: A Laboratory Manual*, 4th edition (2012) (Green and Sambrook); *Current Protocols in Molecular Biology* (1987) (F.M. Ausubel et al. eds.); the series *Methods in Enzymology* (Academic Press, Inc.); *PCR 2: A Practical Approach* (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); *Antibodies, A Laboratory Manual* (1988) (Harlow and Lane, eds.); *Antibodies A Laboratory Manual*, 2nd edition 2013 (E.A. Greenfield ed.); *Animal Cell Culture* (1987) (R.I. Freshney, ed.); Benjamin Lewin, *Genes IX*, published by Jones and Bartlett, 2008 (ISBN 0763752223); Kendrew et al. (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a*

Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471 185710); Singleton et al., Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[00100] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[00101] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[00102] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[00103] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[00104] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[00105] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues,

cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[00106] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[00107] Reference is made to U.S. Provisional Application Nos. 62/351,662 and 62/51,803, filed on June 17, 2016, 62/376,377, filed August 17, 2016, 62/410,366, filed October 19, 2016, 62/432,240, filed December 9, 2016, 62/471,792, filed March 15, 2017, and 62/484,786 filed April 12, 2017, and PCT Application PCT/US2017/038154, filed June 19, 2017. Reference is also made to U.S. Provisional Application Nos. 62/471,710, filed March 15, 2017, 62/432,553, filed December 9, 2016, 62/456,645, filed February 8, 2017, and 62/471,930, filed March 15, 2017.

[00108] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

[00109] In general, a CRISPR-Cas or CRISPR system as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence

(encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). When the CRISPR protein is a Cas13 protein, a tracrRNA is not required.

[00110] In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[00111] In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. The term "targeting sequence" means the portion of a guide sequence having sufficient complementarity with a target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10-30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00112] In a classic CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95% complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82%, or 81%, or 80% complementarity between the sequence and the guide, with it

advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[00113] In certain embodiments, modulations of cleavage efficiency can be exploited by introduction of mismatches, e.g. 1 or more mismatches, such as 1 or 2 mismatches between spacer sequence and target sequence, including the position of the mismatch along the spacer/target. The more central (i.e. not 3' or 5') for instance a double mismatch is, the more cleavage efficiency is affected. Accordingly, by choosing mismatch position along the spacer, cleavage efficiency can be modulated. By means of example, if less than 100 % cleavage of targets is desired (e.g. in a cell population), 1 or more, such as preferably 2 mismatches between spacer and target sequence may be introduced in the spacer sequences. The more central along the spacer of the mismatch position, the lower the cleavage percentage.

[00114] The methods according to the invention as described herein comprehend inducing one or more nucleotide modifications in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s).

[00115] For minimization of toxicity and off-target effect, it will be important to control the concentration of Cas mRNA or protein and guide RNA delivered. Optimal concentrations

of Cas mRNA or protein and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci.

[00116] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence, but may depend on for instance secondary structure, in particular in the case of RNA targets.

[00117] The nucleic acid molecule encoding a Cas is advantageously codon optimized. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a Cas is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene

expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at www.kazusa.or.jp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" *Nucl. Acids Res.* 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas correspond to the most frequently used codon for a particular amino acid.

[00118] In certain embodiments, the methods as described herein may comprise providing a Cas transgenic cell in which one or more nucleic acids encoding one or more guide RNAs are provided or introduced operably connected in the cell with a regulatory element comprising a promoter of one or more gene of interest. As used herein, the term "Cas transgenic cell" refers to a cell, such as a eukaryotic cell, in which a Cas gene has been genomically integrated. The nature, type, or origin of the cell are not particularly limiting according to the present invention. Also the way how the Cas transgene is introduced in the cell is may vary and can be any method as is known in the art. In certain embodiments, the Cas transgenic cell is obtained by introducing the Cas transgene in an isolated cell. In certain other embodiments, the Cas transgenic cell is obtained by isolating cells from a Cas transgenic organism. By means of example, and without limitation, the Cas transgenic cell as referred to herein may be derived from a Cas transgenic eukaryote, such as a Cas knock-in eukaryote. Reference is made to WO 2014/093622 (PCT/US 13/74667), incorporated herein by reference. Methods of US Patent Publication Nos. 20120017290 and 201 10265198 assigned to Sangamo Biosciences, Inc. directed to targeting the Rosa locus may be modified to utilize the CRISPR Cas system of the present invention. Methods of US Patent Publication No. 20130236946 assigned to Collectis directed to targeting the Rosa locus may also be modified to utilize the CRISPR Cas system of the present invention. By means of further example reference is made to Piatt et. al. (*Cell*; 159(2):440-455 (2014)), describing a Cas9 knock-in mouse, which is incorporated herein by reference. The Cas transgene can further comprise a Lox-Stop-polyA-Lox(LSL) cassette thereby rendering Cas expression inducible by Cre recombinase. Alternatively, the Cas transgenic cell may be obtained by introducing the Cas transgene in an isolated cell. Delivery systems for transgenes are well known in the art. By means of example, the Cas transgene may be delivered in for instance eukaryotic cell

by means of vector (e.g., AAV, adenovirus, lentivirus) and/or particle and/or nanoparticle delivery, as also described herein elsewhere.

[00119] It will be understood by the skilled person that the cell, such as the Cas transgenic cell, as referred to herein may comprise further genomic alterations besides having an integrated Cas gene or the mutations arising from the sequence specific action of Cas when complexed with RNA capable of guiding Cas to a target locus, such as for instance one or more oncogenic mutations, as for instance and without limitation described in Piatt et al. (2014), Chen et al., (2014) or Kumar et al.. (2009).

[00120] In some embodiments, the Cas sequence is fused to one or more nuclear localization sequences (NLSs) or nuclear export signals (NESs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs. In some embodiments, the Cas comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs at or near the carboxy-terminus, or a combination of these (e.g. zero or at least one or more NLS or NES at the amino-terminus and zero or at one or more NLS or NES at the carboxy terminus). When more than one NLS or NES is present, each may be selected independently of the others, such that a single NLS or NES may be present in more than one copy and/or in combination with one or more other NLSs or NESs present in one or more copies. In a preferred embodiment of the invention, the Cas comprises at most 6 NLSs. In some embodiments, an NLS or NES is considered near the N- or C-terminus when the nearest amino acid of the NLS or NES is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO: 1); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK) (SEQ ID NO:2); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 3) or RQRRNELKRSP (SEQ ID NO:4); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRS SGPYGGGGQYFAKPRNQQGGY (SEQ ID NO: 5); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 6) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 7) and PPKKARED (SEQ ID NO: 8) of the myoma T protein; the sequence POPKKKPL (SEQ ID NO: 9) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 10) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 11) and PKQKKRK (SEQ ID NO: 12) of the

influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: 13) of the Hepatitis virus delta antigen; the sequence REKKKFLRR (SEQ ID NO: 14) of the mouse Mxl protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 15) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: 16) of the steroid hormone receptors (human) glucocorticoid. Non-limiting examples of NESs include an NES sequence LYPERRRILT (SEQ ID No. 17) (ctgtaccctgagcggctcggcgatcctgacc) (SEQ ID No. 18). In general, the one or more NLSs or NESs are of sufficient strength to drive accumulation of the Cas in a detectable amount in respectively the nucleus or the cytoplasm of a eukaryotic cell. In general, strength of nuclear localization/export activity may derive from the number of NLSs/NESs in the Cas, the particular NLS(s) or NES(s) used, or a combination of these factors. Detection of accumulation in the nucleus/cytoplasm may be performed by any suitable technique. For example, a detectable marker may be fused to the Cas, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI) or cytoplasm. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of CRISPR complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by CRISPR complex formation and/or Cas enzyme activity), as compared to a control not exposed to the Cas or complex, or exposed to a Cas lacking the one or more NLSs or NESs. In certain embodiments, other localization tags may be fused to the Cas protein, such as without limitation for localizing the Cas to particular sites in a cell, such as organelles, such as mitochondria, plastids, chloroplast, vesicles, golgi, (nuclear or cellular) membranes, ribosomes, nucleolus, ER, cytoskeleton, vacuoles, centrosome, nucleosome, granules, centrioles, etc.

[00121] In certain aspects the invention involves vectors, e.g. for delivering or introducing in a cell Cas and/or RNA capable of guiding Cas to a target locus (i.e. guide RNA), but also for propagating these components (e.g. in prokaryotic cells). As used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to

which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00122] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171 156 A1, the contents of which are herein incorporated by reference in their entirety.

[00123] The vector(s) can include the regulatory element(s), e.g., promoter(s). The vector(s) can comprise Cas encoding sequences, and/or a single, but possibly also can comprise at least 3 or 8 or 16 or 32 or 48 or 50 guide RNA(s) (e.g., sgRNAs) encoding sequences, such as 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(s) (e.g., sgRNAs). In a single vector there can be a promoter for each RNA (e.g.,

sgRNA), advantageously when there are up to about 16 RNA(s) ; and, when a single vector provides for more than 16 RNA(s) , one or more promoter(s) can drive expression of more than one of the RNA(s), e.g., when there are 32 RNA(s) , each promoter can drive expression of two RNA(s) , and when there are 48 RNA(s) , each promoter can drive expression of three RNA(s) . By simple arithmetic and well established cloning protocols and the teachings in this disclosure one skilled in the art can readily practice the invention as to the RNA(s) for a suitable exemplary vector such as AAV, and a suitable promoter such as the U6 promoter. For example, the packaging limit of AAV is -4.7 kb. The length of a single U6-gRNA (plus restriction sites for cloning) is 361 bp. Therefore, the skilled person can readily fit about 12-16, e.g., 13 U6-gRNA cassettes in a single vector. This can be assembled by any suitable means, such as a golden gate strategy used for TALE assembly (<http://www.genome-engineering.org/taleffectors/>). The skilled person can also use a tandem guide strategy to increase the number of U6-gRNAs by approximately 1.5 times, e.g., to increase from 12-16, e.g., 13 to approximately 18-24, e.g., about 19 U6-gRNAs. Therefore, one skilled in the art can readily reach approximately 18-24, e.g., about 19 promoter-RNAs, e.g., U6-gRNAs in a single vector, e.g., an AAV vector. A further means for increasing the number of promoters and RNAs in a vector is to use a single promoter (e.g., U6) to express an array of RNAs separated by cleavable sequences. And an even further means for increasing the number of promoter-RNAs in a vector, is to express an array of promoter-RNAs separated by cleavable sequences in the intron of a coding sequence or gene; and, in this instance it is advantageous to use a polymerase II promoter, which can have increased expression and enable the transcription of long RNA in a tissue specific manner. (see, e.g., nar.oxfordjournals.org/content/34/7/e53.short, www.nature.com/mt/journal/v16/n9/abs/mt2008144a.html). In an advantageous embodiment, AAV may package U6 tandem gRNA targeting up to about 50 genes. Accordingly, from the knowledge in the art and the teachings in this disclosure the skilled person can readily make and use vector(s), e.g., a single vector, expressing multiple RNAs or guides under the control or operatively or functionally linked to one or more promoters—especially as to the numbers of RNAs or guides discussed herein, without any undue experimentation.

[00124] The guide RNA(s) encoding sequences and/or Cas encoding sequences, can be functionally or operatively linked to regulatory element(s) and hence the regulatory element(s) drive expression. The promoter(s) can be constitutive promoter(s) and/or conditional promoter(s) and/or inducible promoter(s) and/or tissue specific promoter(s). The promoter can be selected from the group consisting of RNA polymerases, pol I, pol II, pol III,

T7, U6, HI, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EFla promoter. An advantageous promoter is the promoter is U6.

[00125] Aspects of the invention relate to the identification and engineering of novel effector proteins associated with Class 2 CRISPR-Cas systems. In a preferred embodiment, the effector protein comprises a single-subunit effector module. In a further embodiment the effector protein is functional in prokaryotic or eukaryotic cells for *in vitro*, *in vivo* or *ex vivo* applications. An aspect of the invention encompasses computational methods and algorithms to predict new Class 2 CRISPR-Cas systems and identify the components therein.

[00126] In one embodiment, a computational method of identifying novel Class 2 CRISPR-Cas loci comprises the following steps: detecting all contigs encoding the CasI protein; identifying all predicted protein coding genes within 20kB of the casI gene, more particularly within the region 20 kb from the start of the casI gene and 20 kb from the end of the casI gene; comparing the identified genes with Cas protein-specific profiles and predicting CRISPR arrays; selecting partial and/or unclassified candidate CRISPR-Cas loci containing proteins larger than 500 amino acids (>500 aa); analyzing selected candidates using PSI-BLAST and HHPred, thereby isolating and identifying novel Class 2 CRISPR-Cas loci. In addition to the above mentioned steps, additional analysis of the candidates may be conducted by searching metagenomics databases for additional homologs.

[00127] In one aspect the detecting all contigs encoding the CasI protein is performed by GenemarkS which a gene prediction program as further described in "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." John Besemer, Alexandre Lomsadze and Mark Borodovsky, Nucleic Acids Research (2001) 29, pp 2607-2618, herein incorporated by reference.

[00128] In one aspect the identifying all predicted protein coding genes is carried out by comparing the identified genes with Cas protein-specific profiles and annotating them according to NCBI Conserved Domain Database (CDD) which is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function

relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). In a further aspect, CRISPR arrays were predicted using a PILER-CR program which is a public domain software for finding CRISPR repeats as described in "PILER-CR: fast and accurate identification of CRISPR repeats", Edgar, R.C., BMC Bioinformatics, Jan 20;8: 18(2007), herein incorporated by reference.

[00129] In a further aspect, the case by case analysis is performed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). PSI-BLAST derives a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein-protein BLAST. This PSSM is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences. Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

[00130] In another aspect, the case by case analysis is performed using HHpred, a method for sequence database searching and structure prediction that is as easy to use as BLAST or PSI-BLAST and that is at the same time much more sensitive in finding remote homologs. In fact, HHpred's sensitivity is competitive with the most powerful servers for structure prediction currently available. HHpred is the first server that is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment databases, like Pfam or SMART. This greatly simplifies the list of hits to a number of sequence families instead of a clutter of single sequences. All major publicly available profile and alignment databases are available through HHpred. HHpred accepts a single query sequence or a multiple alignment as input. Within only a few minutes it returns the search results in an easy-to-read format similar to that of PSI-BLAST. Search options include local or global alignment and scoring secondary structure similarity. HHpred can produce pairwise query-template sequence alignments, merged query-template multiple alignments (e.g. for transitive searches), as well as 3D structural models calculated by the MODELLER software from HHpred alignments.

[00131] The term "nucleic acid-targeting system", wherein nucleic acid is DNA or RNA, and in some aspects may also refer to DNA-RNA hybrids or derivatives thereof, refers collectively to transcripts and other elements involved in the expression of or directing the activity of DNA or RNA-targeting CRISPR-associated ("Cas") genes, which may include sequences encoding a DNA or RNA-targeting Cas protein and a DNA or RNA-targeting guide RNA comprising a CRISPR RNA (crRNA) sequence and (in some but not all systems)

a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence, or other sequences and transcripts from a DNA or RNA-targeting CRISPR locus. In general, a RNA-targeting system is characterized by elements that promote the formation of a DNA or RNA-targeting complex at the site of a target DNA or RNA sequence. In the context of formation of a DNA or RNA-targeting complex, "target sequence" refers to a DNA or RNA sequence to which a DNA or RNA-targeting guide RNA is designed to have complementarity, where hybridization between a target sequence and a RNA-targeting guide RNA promotes the formation of a RNA-targeting complex. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[00132] In an aspect of the invention, novel RNA targeting systems also referred to as RNA- or RNA-targeting CRISPR/Cas or the CRISPR-Cas system RNA-targeting system of the present application are based on identified Type VI Cas proteins which do not require the generation of customized proteins to target specific RNA sequences but rather a single enzyme can be programmed by a RNA molecule to recognize a specific RNA target, in other words the enzyme can be recruited to a specific RNA target using said RNA molecule.

[00133] In an aspect of the invention, novel DNA targeting systems also referred to as DNA- or DNA-targeting CRISPR/Cas or the CRISPR-Cas system RNA-targeting system of the present application are based on identified Type VI Cas proteins which do not require the generation of customized proteins to target specific RNA sequences but rather a single enzyme can be programmed by a RNA molecule to recognize a specific DNA target, in other words the enzyme can be recruited to a specific DNA target using said RNA molecule.

[00134] The nucleic acids-targeting systems, the vector systems, the vectors and the compositions described herein may be used in various nucleic acids-targeting applications, altering or modifying synthesis of a gene product, such as a protein, nucleic acids cleavage, nucleic acids editing, nucleic acids splicing; trafficking of target nucleic acids, tracing of target nucleic acids, isolation of target nucleic acids, visualization of target nucleic acids, etc.

[00135] As used herein, a Cas protein or a CRISPR enzyme refers to any of the proteins presented in the new classification of CRISPR-Cas systems.

Cas13 Nuclease

[00136] The Class 2 type VI effector protein Cas13 is a RNA-guided RNase that can be efficiently programmed to degrade ssRNA. Cas13 effector proteins of the invention include, without limitation, the following 21 ortholog species (including multiple CRISPR loci: *Leptotrichia shahii*; *Leptotrichia wadei* (Lw2); *Listeria seeligeri*; *Lachnospiraceae* bacterium MA2020; *Lachnospiraceae* bacterium NK4A179; [*Clostridium*] *aminophilum* DSM 10710;

Carnobacterium gallinarum DSM 4847; Carnobacterium gallinarum DSM 4847 (second CRISPR Loci); Paludibacter propionicigenes WB4; Listeria weihenstephanensis FSL R9-0317; Listeriaceae bacterium FSL M6-0635; Leptotrichia wadei F0279; Rhodobacter capsulatus SB 1003; Rhodobacter capsulatus R121; Rhodobacter capsulatus DE442; Leptotrichia buccalis C-1013-b; Herbinix hemicellulosilytica; [Eubacterium] rectale; Eubacteriaceae bacterium CFDDCI004; Blautia sp. Marseille-P2398; and Leptotrichia sp. oral taxon 879 str. F0557. Twelve (12) further non-limiting examples are: Lachnospiraceae bacterium NK4A144; Chloroflexus aggregans; Demequina aurantiaca; Thalassospira sp. TSL5-1; Pseudobutyrvibrio sp. OR37; Butyrvibrio sp. YAB3001; Blautia sp. Marseille-P2398; Leptotrichia sp. Marseille-P3007; Bacteroides ihuae; Porphyromonadaceae bacterium KH3CP3RA; Listeria riparia; and Insoliti spirillum peregrinum.

[00137] Casl3 may be any member in the Casl3 family. For example, Casl3 may be Casl3a, Casl3b, Casl3c, Casl3d, or other member in the Casl3 family. Exemplary orthologs of Casl3 (e.g., Casl3a, Casl3b, Casl3c, and Casl3d) include those described in Tables 1, 2, 3, 4a, 4b, 5, and 6 in WO/2018/107129, filed June 14, 2018, which is incorporated herein by reference. In some embodiments, Casl3d may be the Casl3d effectors described in Yang WX et al. (2018) (Mol Cell. 2018 Apr 19;70(2):327-339.e5), which is incorporated herein by reference.

[00138] Non-limiting examples of protein and direct repeat sequences of Casl3 orthologs include the following. The Casl3 proteins may be codon optimized for expression in mammalian cells.

<p>Leptotrichia shahii (SEQ ID Nos. 19 and 20)</p>	<p>CCACCCCAA TATCGAAGG GGACTAAA AC</p>	<p>MGNLFGHKRWYEVDRDKKDFKIKRKVKVKRNYDGN KYILNINENNKEKIDNNKFIRKYINYKKNNDNILKEF TRKFHAGNILFKLKGKEGIIRIENDDFLETEEVLYI EAYGKSEKLKALGITKKKIIDEAIRQGITKDDKKIEIK RQENEEIEIDIRDEYTNKTLNDCSILRIIENDELETK KSIYEIFKNINMSLYKIIKIIENETEKVFENRYEEHL REKLLKDDKIDVILTNFMEIREKIKSNLEILGFVKFYL NVGGDKKSKNKKMLVEKILNINVDLTVEDIADFVI KELEFWNITKRIEKVKKVNNEFLEKRRNRITYIKSYV LLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIEK ILAEFKIDELIKKLEKELKKGNCDEIFGIFKKHYKVN FD SKKF SKKSDEEKEL YKII YRYLKGRIEKIL VNE QK VRLKKMEKIEIEKILNE SILSEKILKRVKQ YTLEHIM Y LGKLRHNDIDMTT VNTDDF SRLHAKEELDLELITFF A STNMELNKIF SRENINNDENIDFFGGDREKNYVLDK KILNSKIKIIRDLD FIDNKN NITN NFIRKFTKIGTNERN RILHAISKERDLQGTQDDYNKVINIIQNLKISDEEVSK ALNLDV VFKDKNIITKINDIKI SEENNDIK YLP SF S KVLPEILM.YRNNPKNEPFDTIETEKIVLNALIYVNKE LYKKLILEDDEENESKNIFLQELKKT LGNIDEIDENII ENYYKNAQISASKGNNKA IKKYQKKVIECYIGYLRK NYEELFDF SDFKMNIQEIKKQIKDIKDNKT YERITVK</p>
---	--	--

		<p>TSDKTIVTNDDFEYIISIFALLNSNAVINKIRNRFFATS VWLNT SEYQNHDLDEIMQLNTRLNECITENWNLNL EEFIQKMKEIEKDFDDFKIQTKKEIFNNYEDIKNNIL TEFKDDINGCDVLEKKLEKIVIFDDETKFEIDKKSNI QDEQRKLSNINKKDLKKKVDQYIKDKDQEIKSKILC RIIFNSDFLKK YKKEIDNLIEDME SENENKFQEIYYPK ERKNELYIYKKNLFLNIGNPNFDKIYGLISNDIKMAD AKFLFNIDGKNIRKNKISEIDAILKNLNDKLNYSKE YKEKYIKKLENDFFAKNIQNKNYKSFEKDYNRVS EYKKIRDLVEFNLYLNKIESYLIDINWKLAIQMARFER DMHYIVNGLRELGIKL SGYNTGI SRA YPKRNGSDGF YTTTAYYKFFDEESYKKFEKICYGFGIDLSENSEINKP ENESIRNYISHFYIVRNPADYSIAEQIDRVSNLLSYST RYNNSTYASVFEVFKKDVNLDYDELKKKFKLIGNN DILERLMKPKKVSVLELESYNSDYIKNLIHELLTKIEN TND TL</p>
<p>Leptotrichia wadei (Lw2) (SEQ ID Nos. 21 and 22)</p>	<p>GATTAGAC TACCCAAA AACGAAGG GGACTAAA AC</p>	<p>MK VTK V DGI sFKK YIEEGKL VK ST SEENR T SERL SEL LSIRLDIYKPNPDNASEEENRIRRENKFF SNK VLHL KDSVLYLKNRKEKNAVQDKNYSEEDISEYDLKNKN SF SVLKKILLNED VNSEELEIFRKD VE AKLNKINSLK YSFEENKANYQKINENNVEKVGKSKRNIIYDYRE SAKRNDYINNVQEAFDKLYKKEDIEKLFFLIENSKKH EKYKIREYYHKIIGRKNNDKENFAKIIYEEIQNVNNIKE LIEKIPDMSELKKSQVFYKYLDKEELNDKNIKYAF CHFVEIEMSQLKNYVYKRLSNISNDKIKRIFEYQNL KKLIENKLLNKLD TYVRNCGK YNYYLQ VGEI AT SDF IARNRQNE AFLRNIGVS SVAYF SLRNILETENENDIT GRMRGKTVKNNKGEEKYVSGEVDKIYNENKQNEV KENLKM FY SYDFNMDNKNEIEDFFANIDE AIS SIRHG IVHFNLELEGKDIF AFKNI AP SEISKKMF QNEINEKKL KLKIFKQLNSANVFNYYEKDVIIKYLKNTKFNFNfC NIPFVPSFTKLYNKIEDLRNTLKFFWSVPKDKKEKDA QIYLLKNIYYGEFLNKFVKNSKVFFKITNEVIKINKQ RNQKTGHYK YQKFENIEKT VPVE YL AIQ SREMINN QDKEEKNTYIDFIQQIFLKGFDYLNKNNLKYIESNN NNDNNDIFSKIKIKDNKEK YDKILKNYEKHNRNKE IPHEINEFVREIKLGKILKYTENLNMFYLLKLLNHKE LTNLKGSLEK YQSANKKE TF SDELELINLLNLDNRR VTEDFELEANEIGKFLDFNENKIKDRKELKFFDTNKI YFDGENIHKHRAFYNIKKYGMLNLEKIADKAKYKIS LCELKEYSNKKNEIEKNYTMQQNLHRKYARPKKDE KFNDEDYKEYEKAIGNIQKYTHLKNKVEFNELNLLQ GLLLKILHRLVGYTSIWERDLRFRLKGEFPENHYIEEI FNFDNSKNVYKSGQIVEKYINFYKEL YKDNVEKRS IYSDKKVKKLKQEKKDL YIRNYIAHFNYIPHAESLL EVLENLRKLLSYDRKLNKAIMKSIVDILKEYGFVATF KIGADKKIEIQTLESEKIVHLKNLKKKMLMTRNSEE LCELVKVMFEYKALEKRPAATKKAGQAKKKKGSYP</p>

		YDVDPDYAYPYDVPDYAYPYDVPDYA*
<p><i>Listeria seeligeri</i> (SEQ ID Nos. 23 and 24)</p>	<p>GTAAGAGA CTACCTCTA TATGAAAG AGGACTAA AAC</p>	<p>MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKTMRI TKVEVDRKKVLISRDKNGGKLVYENEMQDNTEQIM HHKKS SFYKSVVNKTICRPEQKQMKKLVHGLLQEN SQEKIKVSDVTKLNI SNFLNHRFKKSLY YFPENSPDK SEYRIEINLSQLLEDSLKKQQGTFICWESFSKDMEL YINWAENYISSKTKLIKK SIRNNRIQSTESRS GQLMD RYMKDILNKNKPFDIQ SVSEKYQLEKLT SALKATFK EAKKNDKEINYKLLK STLQNHERQIIEELKENSELNQF NIEIRKHLETYFPIKKTNRKVGDIRNLEIGEIQKIVNH RLKNKIVQRILQEGKLASYEIESTVNSNSLQKIKIEEA FALKFINACLFASNNLRNMVYPVCKKDILMIGEFKN SFKEIKHKKFIRQWSQFFSQEITVDDIELASWGLRGAI APIRNEIHLKKHSWKKFFNNPTFKVKKSKIINGKTK DVTSEFLYKETLFDYFYSELDSVPELIINKMESSKIL DYYSSDQLNQVFTIPNFELSLTSAVPFAPSFKRVYL KGFYQNQDEAQP DYNLKLNIYNEKAFNSEAFQAQ YSLFKMVYYQVFLPQFTTNDL FKSSVDFILTLNKE RKGYAKAFQDIRKMNKDEK PSEYMSYIQSQLMLYQ KKQEEKEKINHFEKFINQVFIKGFNSFIEKNRLTYICH PTKNTVPENDNIEIPFHTDMDD SNAIFWLMCKLLDA KQLSELRNEMIKFSCSLQSTEEISTFTKAREVIGLALL NGEKGCNDWKELFDDKEAWKKNMSLYVSEELLQS LPYTQEDGQTPVINRSIDL VKK YGTETILEKLFSSSDD YKVSADIAKLHEYDVTEKIAQQESLHKQWIEK PGL ARDSAWTKKYQNVINDISNYQWAKTKVELTQVRHL HQLTIDLLSRLAGYMSIADRDFQF SSNYILERENSEY RVTSWILLSENKNKNKYNDYELYNLKNASIKVSSKN DPQLKVDLQQLRLTLEYLELFDNRLKEKRNNISHFN YLNQQLGNSILELFD DARDVLSYDRKLKNAVSKSLK EILSHGMEVTFKPLYQTNHHLKIDKLQPKKIHLGE KSTVSSNQVSNEYCQLVRTLLTMK</p>
<p>Lachnospiraceae bacterium MA2020 (SEQ ID Nos. 25 and 26)</p>	<p>GTATTGAGA AAAGCCAG ATATAGTTG GCAATAGA C</p>	<p>MQISKVNHHKHAVGQKDRERITGFIYNDPVGDEKSL EDVVAKRANDTKVLFNVFNTKDL YD SQE SDKSEKD KEISKGAKFVAKSFNSAITILKKQNKIYSTLTSQQVI KELKDKFGGARIYDDIEEALTETLKK SFRKENVRN SIKVLINAAGIRSSLSKDEEELIQEYFVKQLVEEYTK TKLQKNVVKSIKNQNMVIQPDSDSQVLSLSESRREK QSSAVS SDTLVNCKEKDVLKAFLTDYAVLDEDERNS LLWKLRLNLVNL YFYGSESIRDYSYTKESVWKEHD EQKANKTLFIDEICHITKIGKNGKEQKVL DYEENRSR CRKQNINYRSALNYAKNNTSGIFENEDSNHFWIULI ENEVERLYNGIENGEEFKFETGYI SEKVWKA VINHLS IKYIALGKAVYNYAMKELSSPGDIEPGKIDDSYINGIT SFDYEIIKAEESLQRDISMNVVFATNYLACATVDTDK DFLLFSKEDIRSC TKKDG NLCKNIMQFWGGYSTWK NFC EEYLKDDKD ALELLYSLKSMLYSMRNS SFHFST ENVDNGSWDTELIGKLFEEDCNRAARIEKEKFYNNN</p>

		<p>LHMFYS SSLLEKVLERL YSSHHERASQVP SFNRV FV RKNFP SSLSEQRITPKF TD SKDEQIWQ SAVY YLCKEI Y YNDFLQ SKEA YKLFREGVKNL DKNDFNNQK AADS FKQAVVYYGKAIGNATLSQVCQAIMTEYNRQNN DG LKKKSAYA EKQNSNKYKHYPFLKQVLQSAFW EYL DENKEIYGFISAQIFDCSNVEIKAEDFIANYSSQ QYKK LVDKVKKTPELQKWYTLGRLFNPRQANQFLGSIRNY VQFVKDIQRRAKENGNPIRNYEVLSDSI IKILEMC TKLNGTTSNDIFIDYFRDEDEYAEYISQFVNFGDVHS GAALNAFCNSESEGKKNGIYYDGTNP IVNRNWV LCK LYGSPDLI SKIISRVNENMIFIDFFiKQEDLIRE YQIKGI CSNKKEQQDLRTFQVLKNRVELRDIVEYSEIFNELYG QLIKWC YLRERDLMYFQLGFHYLCLNNAS SKEADYI KTNVDDRNISGAILYQI AAMYTNGLPVY YKKDDM YV ALKSGKKASDELNSNEQTSKkrNYFLKYGNNILGDK KDQLYLAGLELFENVAEHENIIIFRNEIDHFHYFYDR DRSM LDLYSEVFDRFFTYDMKLRKNV VNM LYNILL DUNIV SSFVFETGEKK VGRGD SEVIKP SAKIRLRANN GVS SDVFTYKVGSKDELKIATLP AKNEEFLLNVARLI YYPDME AVSENMVREGVVK VEK SNDKKGKI SRGSN TRSSNQSKYNNKSKNRMNYSMGSIFEKMDLKFD</p>
<p>Lachnospiraceae bacterium NK4A 179 (SEQ ID Nos. 27 and 28)</p>	<p>GTTGATGAG AAGAGCCC AAGATAGA GGGCAATA AC</p>	<p>MKISKVREENRGAKLTVNAKTAVVSENRSQEGILYN DPSRYGKSRKNDEDRDRYIE SRLKSSGKL YRIFNEDK NKRETDELQ WFLSEIVKKFNRRNGLVLS DMLSVDDR AFEKA FEKYAELSYTNRRNKVSGSPAFETCGVDAAT AERLKGIISETNFNRIKNNIDNKVSEDIIDRIIAKYLK KSLCRERVKRGLKLLMNAFDLPYSDPDIDVQRDFI DYVLEDFYHVRAKSQVSRSIKMNMPVQPEGDGKF AITVSKGGTESGNKRS AEKEAFKKFLSDYASLDERV RDDMLRRMRRLV VLYFYGSDDSKLSDVNEKFDVW EDHAARRVDNREFIKLPLENKL ANGKTDKDAERIRK NTVKELYRNQNIGCYRQAVKAVEEDNNGRYFDDK MLNMFFIHRIEYGV EKIYANLKQVTEFKARTGYLSE KIWKDLrNYISIKYIAMGKAVVNYAMDELNASDKKE IELGKI SEEYL SGISSFDYELIKAEEMLQRET AVYVAF AARHLSSQTVELDSENSDFLLLKPKGTMDKNDKNK LASNNILNFLKDKETLRDTILQYFGGHSLWTDFFPDK YLAGGKDDVDFLTDLKDVIY SMRND SFHYATENHN NGKWNKELISAMFEHETERMTVVMKDKFY SNNLP MFYKNDDLKLLIDLYKDNVERASQVPSFNKVFVR KNFPALVRDKDNLGIELDLKADADKGENELKFYNA LYMFKEIYYNAFLNDKNVRERFITKATKVADNYD RNKERNLKDRIK SAGSDEK KKLREQLQN YIAENDFG QRIKNIVQVNPDYTLAQICQLIMTEYNQQNNGCMQK KSAARKDrNKDSYQHYKMLLLVNL RKAFLEFIKENY AFVLKPYKHDLCDKADFVPDFAKYVKPYAGLISRV AGSSELQKWYIVSRFLSPAQANHMLGFLHSYKQYV WDIYRRASETGTETrNHSIAEDKIAGVDITDVDAVIDL SVKLCGTIS SEISDYFKDDEVYAEYIS SYLDFEYDGG</p>

		<p>NYKDSLNRFCNSDAVNDQKVALYYDGEHPKLN IL SKLYGERRFLEKITDRVSRSDI VE YYKLK QTKGIFDSEDEQKNIKKFQEMKNIVEFRDL MDYSEIA DELQGQLINWIYLRERDLMNFQ LGYHYACLNND SN KQATYVTLDYQGKKNR KINGAILYQICAMYINGLPL YYVDKDSSE WTVSDGKESTGAKIGEFYRYAKSFENT SDCYASGLEIFENISEHDNITELRNYIEHFR YYYSSFDR SFLGIYSEVDFRFFTYDLKYR KKNVPTILYNILLQHFV NVRFEFVSGKKM IGIDKKDRKIAKEKECARITIREKN GYSE QFTYKLNKGTVYVDARDKRYLQSIIRLLF YP EKVNMDEMIEVKEKKKP SDNNTGKGY SKRDRQQD RKEYDKYKEKKKKEGNFL SGMGGNINWDEIN AQLK N</p>
<p>[Clostridium] aminophilum DSM 10710 (SEQ ID Nos. 29 and 30)</p>	<p>GTCTATTGC CCTCTATAT CGGGCTGTT CTCCAAAC</p>	<p>MKFSKVDHTRS AVGIQKATD SVHGMLYTD PPKQEV NDLDKRFDQLNVKAKRL YNVFNQ SKAEEDDDEKRF GKVVKKNRELKDLLFH REVSRYN SIGNAKYNYGI KSNPEEIVSN LGMVE SLKGERDPQK VISKLLLYLRK GLKPGTDGLRMILEASCGLRKLSGDEKEL KVFLQTL DEDFEKKTFFKKNLIRSIE NQNMAVQPSNEGDP IIGITQ GRFNSQK NEEKSAIERMMSMYADLNEDHREDVLRK LRRLNVLYFNVDTEKTEEPTLPGEVDTNP VFVWHD HEKGKENDRQFATFAKILTEDRE TRKKEKLAVKEAL NDLKSAIRDHNIMAY RCSIKVTEQDKDGLFFEDQRIN RFWIHH IESAVERILASINPEKLYKLRIGYLGEK VWK DLLNYLSIKYIAVVGKAVFHFAMED LGKTGQDIELGK LSNSVSGGLTSFDYE QIRADETLQRQLSVEVAFAN NLFRAV VGQTGKKIEQSKSEENEEDFLLWKA EKIAE SIKKEGEGNTLKSILQFFGGASS WDLNHFCAAYGNE SSALGYETKFADDL RKAIYSLRNETFHFTTLNKGSFD WNA KLIGDMF SHEAATGIAVERTRFYSNNL PMFYRE SDLKRIMDHL YNTYHPRASQ VPSFNSVFVRKNFRLF LSNTLNTNT SFDTEVYQKWE SGVYYLFKEIYNSFL PSGDAHHLFFEGLRRIRKEADNLPVIG KEAKKRNAV QDFGRRCDLKNL SL SAIC QMIMTE YNEQNNGNRK VKSTREDKR KPDIF QHYKMLLLRTLQE AF AIYIR REE FKFIFDLPKTL YVMKPV EEF LPNWK SGMFDSL VERV KQSPDLQR WYVLCKFLNGRLLNQLSGVIRSYIQFAG DIQRRAKANHNRLYMDNTQRVEYYSN VLEVVD FCI KGT SRF SNVFSDFRDE DAYADYLDNYLQFKDEKIA EVS SFA ALKTFCNEEEVKAGIYMDGENPVMQR NIV MAKLFGPDEVLKNVVPKVTREE IEE YYQLEKQIAPY RQNGYCK SEED QKLLRFQRIKNRVEFQTITEF SEIIN ELLGQLISW SFLRERDLL YFQLGFHY LCLHNDTEKP AEYKEISREDGTVIR NAILHQVAAMYVGGLPVYTLA DKKLA AFEKGEADCKLSISKDTAGAGKKIKD FFRYS KYVLIKDRMLTDQNQKYTIY LAGLELFENTDEHDNI TDVRKYVDH FKYYSATSDENAMSILDYSEIHDRFF T YDMKYQKNVANMLENILLRHFV LIRPEFFTGSKKVG EGKKITCKARA QIEIAENGMRSEDFTYKLSDGKKNIS</p>

		TCMIAARDQKYLNTVARLLYPHEAKKSIVDTREK KNNKKTNRGDGTFNKQKGTARKEKDNGPREFNDT GFSNTPFAGFDPFRNS
Carnobacterium gallinarum DSM 4847 (SEQ ID Nos. 31 and 32)	ATTAAAGAC TACCTCTAA ATGTAAGA GGACTATAA C	MRITKVKIKLDNKLYQVTMQKEEKYGTLKLNEE SR KSTAEILRLKKA SFNK SFHSKTINS QKENKNATIKKN GDYISQIFEKL VGVD TNKNIRKPKMSLTDLKDLPKK DLALFIKRKFKNDDIVEIKNLDLISLFYNALQKVPGE HFTDES WADFCQEMMPYREYKNKFIERKIILLANSIE QNKGF SINPETF SKRKRVLHQWAIEVQERGDF SILDE KLSKLA EIYNFKKMCKRVQDELNDLEK SMKKGKNP EKEKEA YKKQKNFKIKTIWKDYKYKTHIGLIEKIKEN EELNQFNIEIGKYFEHYFPIKKERCTE DEPYL NSETI ATTVNYQLKNALISYLMQIGKYKQFGLENQVLDSK KLQEIGIYEGFQTKFMDACVFATSSLNIIEMRSGDI LGKREFKEAIATSSFVNYHHFFPYFPFELKGMKDRES ELIPFGEQTEAKQMQNIWALRGSVQQIRNEIFHSFDK NQKFNL PQLDKSNFEFDASENSTGKSQSYIETDYKFL FEAEKNQLEQFFIERIKS SGALEYYP LKSLEKLF AKK EMKFSLGSQV VAFAPS YKKLVKKGHSYQTATEGTA NYLGLSYYNRYELKEE SFQAQYYLLKLIYQYVFLPN FSQGNSPA FRET V KAILRINKDEARKKMKNKFLR KYAFEQVREMEFKETPDQYMSYLQSEMREEKVRKA EKNDKGFEKNITMNF EKLLMQIFVKGFDVFLTTFAG KELLLSSEEKVIKETEI SL SKKINEREKTLKASIQVEH QLVATNSAISYWLFC KLLDSRHLNELRNEMIKFKQS RIKFNHTQH AELIQNLLPIVELTIL SNDYDEKND SQN VDVSA YFEDKSL YETAPYVQTDDRTRV SFRPILKLE KYHTKSLIEALLKDN PQFRVAATDIQEW MHKREEIG ELVEKRKNLHTEWAEGQQTLGAEKREEYRDYCKKI DRFNWKANK VTLTYL SQLHYLITDLLGRM VGF SAL FERDLVYF SRSF SELGGETYHISDYKNLSGVLRLNAE VKPIKIKNIKVIDNEENPYK GNEPEVKPFLDRLHAYL ENVIGIKAVHGKIRNQT AHL SVLQLELSMIESMNNLR DLMAYDRKLKNAVTKSMIKILDKHGMILKLKIDEN HKNFEIESLIPKEIIHLKDKAIKTNQVSEEYCQLVLAL LTTNPGNQLN
Carnobacterium gallinarum DSM 4847 (SEQ ID Nos. 33 and 34)	AATATAAAC TACCTCTAA ATGTAAGA GGACTATAA C	MRMTKVKINGSPVSMNRSKLNHGLVWNGTTNTVNI LTKKEQSFAASFLNKT LVKADQVKGYK VLAENIFIIF EQLEKSNSEKPSVYLNNIRRLKEAGLKRFFK SKYHEE IKYTSEKNQSVPTKLNLIPLFFNAVDRIQEDKFDEKN WSYFCKEMSPYLDYKKS YLNRKKEILANSIQQRGF SMPTAEENLLSKRKQLFQQWAMKFQESPLIQQNNF AVEQFNKEFANKINELAAVYNVDELCTAITEKLMNF DKDKSNKTRNFEIKKLWKQHPHNKDKALIKLFNQE GNEALNQFNIELGKYFEHYFPKTGKKESAESYYLNP QTIKTVGYQLRNAFVQYLLQV GKLHQYNKGVLD S QTLQEIGMYEGFQTKFMDACVFA SSSLRNIIQATTNE DILTREKFKKELEKNVELKHDLFFKTEIVEERDENPA

		<p>KKIAMTPNELDLWAIRGAVQVRNQHQQINKRHE PNQLKVGSGFENGDLGNVSYQKTIYQKLFDAEIKDIEI YFAEKIKS SGALEQYSMKDLEKLF SNKEL TL SLGGQ VVAFAPSYKKLYKQGYFYQNEKTIELEQFTDYDFS DVFKANYYLKLIYHYVFLPQF SQANNKLFKDTVHY VIQQNKELNTTEKDKKNNKKIRKYAFEQVKLMKNE SPEKYMQYLQREMQEERTIKEAKKTNEEKPNYNFE KLLIQIFIKGFDTFLRNFDLNLNPAEELVGTVKEKAE GLRKRKERIAKILNVDEQIKTGDEEIAFWIFAKLLDA RHLSELRNEMIKFKQSSVKKGLIKNGDLIEQMOPILE LCILSNDSE SMEKESFDKIEVFLEKVELAKNEPYMQE DKLTVPVKFRFMKQLEKYQTRNFIE NLVIENPEFKVSE KIVLWHEEKEKIADLVDRKTKLHEEWASKAREIEE YNEKIKKNKSKKLDKPAEFAKFAEYKIICEAIENFNR LDHKVRLTYLKNLHYLMIDLMGRMVGF SVLFRDF VYMGRSYSALKKQSIYLN DYDTFANIRDWEVNENK HLFGTSS SDLTFQETAEFKNLKKPMENQLK ALLGVT NHSFEIRNNIAHLHVLNRNDGKGEGVSLLSMNDLRK LMSYDRKLNKNAVTKAIKILDKHGMILKLTNNDHTK PFEIESLKPCKIIHLEKSNHSFPMDQVSQEYCDLVKK MLVFTN</p>
<p>Paludibacter propionigenes WB4 (SEQ ID Nos. 35 and 36)</p>	<p>CTTGTGGAT TATCCCAA ATTGAAGG GAACTACA AC</p>	<p>MRVSKVKVKDGGKDKMVLVHRKTTGAQLVYSGQP VSNETSNIPEKKRQSF DLSTLNKTIKFDTAKKQKL NVDQYKIVEKIFKYPKQELPKQIKAEIILPFLNHKFQ EPVKYWKNGKEE SFNLTLIVEAVQAQDKRKLQPY YDWKTWYIQTK SDLLKKS IENNRIDL TENL SKRKA LLAWETEFTASGSIDLTHYHKVYMTDVLCKMLQDV KPLTDDKGKINTN AYHRGLKKALQNHQPAIFGTREV PNEANRADNQLSIYHLEVVKYLEHYFPIKTSKRRNT ADDIAHYLKAQTLKTTIEKQLVN AIRANIIQQGKTNH HELKADTTSNLIRIKTNEAFVLNLTGTCAFAANNIR NMVDNEQTNDILGKGDFIKSLLKDNTNSQLYSFFFG EGLSTNKA EKETQLWGIRGAVQQIRNNVNHYYKDA LKTVFNISNFENPTITDPKQQTNYADTIYKARFINELE KIPEAFAQQLKTGGAVSYTIENTLKSLLTTFQFSLCR STIPFAPGFKKVFNGGINYQNAKQDES FYELMLEQY LRKENFAEESYNARYFMLKLIYNNLFLPGFTTDRKA FADSVGFVQMKNKQAEKVNPRKKEAYAFEAVRP MTAADSIADY MAYVQSELMQEQNKKEEKVAEETRI NFEKFVLQVFIKGFDSFLRAKEFDVQMPQPQLTAT ASNQQKADKLNQLEASITADCKLTPQYAKADDATHI AFYVFCLLDA AHL SNLRNELIKFRE SVNEFKFHLL EIIEICLLSADVPTDYRDLYSSEADCLARLPFIEQG ADITNWSDLFVQSDKHSPVIHANIELSVKYGTTKLE QIINKDTQFKTTEANFTAWNTAQKSIEQLIKQREDHH EQWVKAKNADDKEKQERKREKSNFAQKFIEKHGD DYLDICDYINTYNWLDNKMHFVHLNRLHGLTIELLG RMAGFVALFDRDFQFFDEQQIADEFKLHG FVNLHSI DKKLNEVPTKKIKEIYDIRNKIIQINGNKINESVRANL</p>

		<p>IQFIS SKRNYYNNAFLHVSND EIKEKQMYDIRNHIAH FNYLTKDAADFSLIDLINELRELLHYDRKLKNAVSK AFIDLFDKHGMILKLNADHKLKVESLEPKKIYHL GSSAKDKPEYQYCTNQVMMAYCNMCRSLLEMKK</p>
<p>Listeria weihenstephanen sis FSL R9-0317 (SEQ ID Nos. 37 and 38)</p>	<p>GATTTAGAG TACCTCAA AATAGAAGA GGTCTAAAA C</p>	<p>MLALLHQEVP SQKLUNLKS LNTESLTKLFPKPFQNM ISYPPSKGAEHVQFCLTDIAVPAIRDLDEIKPDWGIFF EKLKPYTDWAESYIHYKQTTIQKSIEQNKIQSPDSPR KLVLQKYVTAFLNGEPLGLDLVAKKYKLADLAESF KVVLDNEDKSANYKIKACLQQHQRNILDELKEDPEL NQYGIEVKKYIQRYPFIKRAPNRSKHARADFLKKELI ESTVEQQFKNAVYHYVLEQGKMEAYELTDPKTKDL QDIRSGEAFSFKFINACAFASNNLKMILNPECEKDILG KGDFKKNLPNSTTQSDVVKKMIPFFSDEIQNVNDFE AIWAIRGSIQQIRNEVYHCKKHSWKSILKIKGFEP NNMKYTDSDMQKLMKDIAKIPDFIEEKLKSSGIIRF YSHDKLQSIWEMKQGFSLTTNAPFVPSFKRVYAKG HDYQT SKNRYYDLGLTTFDILEYGEEDFRARYFLTK LVYYQQFMPWFTADNNAFRDAANFVLRNLKNRQQ DAKAFINIREVEEGEMPRDYMGYVQGGIAIHEDSTE DTPNHFEKFISQVFIKGFDSHMRSADLKFIKNPRNQG LEQSEIEEMSFDIKVEPSFLKNKDDYIAFWTFCKMLD ARHLSELRNEMIKYDGHILTGEQEIIGLALLGVDSREN DWKQFFSSEREYEKIMKGYVGEELYQREPYRQSDG KTPILFRGVEQARKYGTETVIQRLFDASPEFKVSKCN ITEWERQKETIETIERRKELHNEWKNPKKPNNAF FKEYKECCDAIDAYNWHKNKTTLVYVNELHLLIEI LGRYVGYVAIADRDFQCMANQYFKHSGITERVEYW GDNRLKSIKKLDTFLKKEGLFVSEKNARNHIAHLNY LSLKSECTLLYLSERLREIFKYDRKLKNAVSKSLIDIL DRHGMSVVFANLKENKHRLVIKSLEPKLRHLGK KIDNGYIETNQVSEEYCGIVKRLEI</p>
<p>Listeriaceae bacterium FSL M6-0635 (SEQ ID Nos. 39 and 40)</p>	<p>GATTTAGAG TACCTCAA AAAAAAGA GGACTAAA AC</p>	<p>MKITKMRVDGRTIVMERTSKEGQLGYEGIDGNKTTE IIFDKKESFYKSILNKTVRKPDEKEKNRRKQAINKA INKEITELMLAVLHQEVP SQKLUNLKS LNTESLTKLFP KPFQNMISYPPSKGAEHVQFCLTDIAVPAIRDLDEI KPDWGIFFEKLKPYTDWAESYIHYKQTTIQKSIEQNK IQSPDSPRKLVLQKYVTAFLNGEPLGLDLVAKKYKL ADLAESFKLVLDNEDKSANYKIKACLQQHQRNILDE LKEDPELNQYGIEVKKYIQRYPFIKRAPNRSKHARA DFLKKELIESTVEQQFKNAVYHYVLEQGKMEAYEL TDPKTKDLQDIRSGEAFSFKFINACAFASNNLKMILN PECEKDILGKGNFKKNLPNSTTRSDVVKKMIPFFSDE LQNVNDFEAIWAIRGSIQQIRNEVYHCKKHSWKSIL KIKGFEPNNMKYADSDMQKLMKDIAKIPEFIEE KLKSGVRFYRHDELQSIWEMKQGFSLTTNAPFV PSFKRVYAKGHDYQT SKNRYYNLDLTTFDILEYGEE DFRARYFLTKLVYYQQFMPWFTADNNAFRDAANFV LRLNKNRQQDAKAFINIREVEEGEMPRDYMGYVQG</p>

		<p>QIAIHEDSIEDTPNHFEKFISQVFIKGFDRHMRSANLK FIKNPRNQGLEQ SEIEEMSFDIKVEP SFLKNKDDYIAF WIFCKMLDARHLSERNEMIKYDGHLTGEQEIIIGLA LLGVD SRENDWKQFF SSEREYEKIMKGYVVEEL YQ REPYRQSDGKTPILFRGVEQARKYGTETVIQRLFDA NPEFKVSKCNLAEWERQKETIEETIKRRKELHNEWA KNPKKPQNNAFFKEYKECCDAIDAYNWHKNKTTLA YVNELHLLIEILGRY VGYVAIADRDF QCMANQ YFK HSGITERVEYWGDNRLKSIKKLDTFLKKEGLFVSEK NARNHIAHLN YL SLK SECTLL YL SERLREIFK YDRKL KNAVSKSLIDILDRHGMSVVFANLKENKHRLVIKSL EPKCLRHLGGKKIDGGYIETNQVSEEYCGIVKRLL M</p>
<p>Leptotrichia wadei F0279 (SEQ ID Nos. 41 and 42)</p>	<p>GATATAGAT AACCCCAA AAACGAAG GGATCTAAA AC</p>	<p>MKVTK VDGI sFKk YIEEGKL VK STSEENR T SERL SEL LSIRLDIYIKNPDNASEEENRIRRENKFF SNKVLHL KDSVLYLKNRKEKNAVQDKNYSEEDISEYDLKNKN SF SVLKKILLNED VNSEELEIFRKD VEAKLNKINSLK YSFEENKANYQKINENNVKVGKSKRNIIDYYRE SAKRNDYINNVQEAFDKLYKKEDIEKLFLLIENSKKH EKYKIREYYHKIIGRKNKDNKFNFAKIIEEIQNVNNIKE LIEKIPDMSELKKSQVFYKYLDKEELNDKNIKYAF CHFVEIEMSQLLKNYVYKRLSNISNDKIKRIFEYQNL KKLLENKLLNKLD TYVRNCGK YNYYLQ VGEI AT SDF IARNR QNE AFLRNIGVS SVAYF SLRNILETENENDIT GRMRGKT VKNNGEEKYVSGEVDKIYNENKQNEV KENL KMFY SYDFNMDNKNEIEDFFANIDE AIS SIRHG IVHFNLELEGKDIF AFKNI AP SEISKKMF QNEINEKKL KLIKIFKQLNSANVFNYEKDVIKYLKNTKFNFNfC NIPFVPSFTKLYNKIEDLRNTLKFFWSVPKDKKEEKDA QIYLLKNIYYGEFLNKFVKNSKVFFKITNEVIKINKQ RNQKTGHYK YQKFENIEKT VPVEYLAIIQ SREMINN QDKKEEKNTYIDFIQQIFLKGFDYLNKNNLKYIESNN NNDNNDIFS KIKIKKDNKEK YDKILKNYEKHNRNKE IPHEINEFVREIKLGKILKYTENLNMFYLLKLLNHKE LTNLKGSLEK YQSANKKE TF SDELELINLLNLDNRR VTEDFELEANEIGKFLDFNENKIKDRKELKKFDTNKI YFDGENIIKHRAFYNIKKYGMLNLEKIADKAKYKIS LKELKEYSNKKNEIEKNYTMQQNLHRKYARPKKDE KFNDYDYKEYEKAIGNIQKYTHLKNKVEFNELNLLQ GLLLKILHRLVGYTSIWERDLRFRLKGEFFPENHYIEEI FNFDNSKNVKYKSGQIVEKYINFYKELYKDNVEKRS IYSDKKVKKLKQEKDLYIRNYIAHFNYIPHAESLL EVLENLRKLLSYDRKLKNAIMKSIVDILKEYGFVATF KIGADKKIEIQTLESEKIVHLKNLKKKLMTRNSEE LCELVKVMFEYKALE</p>
<p>Rhodobacter capsulatus SB</p>	<p>GCCTCACAT CACCGCAA GACGACGG</p>	<p>MQIGK VQGRTI SEFGDP AGGLKRKI STDGKNRKELP AHL SSDPK ALIGQ WISGIDKI YRKPDSRK SDGK AIHSP TPSKMQFD ARDDLGE AFWKL VSE AGL AQD SDYDQF</p>

<p>1003 (SEQ ID Nos. 43 and 44)</p>	<p>CGGACTGA AC</p>	<p>KRRLHP YGDKF QPAD SGAKLKFE ADPPEPQ AFHGR WYGAMSKRGNDAKELAAALYEHHLVDEKRIDGQP KRNPKTDFK APGLVVARALGIES SVLPRGMARLARN WGEETIQTYFVVDVAASVKEVAKAAVSAAQAFDPP RQVSGRSLSPKVGAFALAEHLERVTGSKRCSFDPAAG PSVLALHDEVKKTYYKRLCARGKNAARAFPADKTEL LALMRHTHENRVRNQMVVMGRVSEYRGQQAGDLA QSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMK AWIDPMGKI VNTEKNDRDLT AA VNIRQ VISNKEMV AEAMARRGIYFGETPELDRLGAEGNEGFV FALLRYL RGCNRNQT FHLGARAGFLKEIRKELEKTRWGKAKEA EHVVLTDKT VAAIRAIIDND AKALGARLL ADL SGAF VAHYASKEHF STLYSEIVKAVKD APEVS SGLPRLKL LLKRADGVRGYVHGLRDRTRKHAFATKLPPPPAPREL DDPATKARYIALLRLYDGPFRAYASGITGTALAGPA ARAKE AAT ALA QSVNVTK AYS DVMEGRT SRLRPPN DGETLREYLSALTGETATEFRVQIGYESDSENARKQ AEFIENYRRDMLAFMFEDYIRAKGFDWILKIEPGAT AMTRAPVLPIDTRGQYEHWQAALYLV MHFVPAS DVS NLLHQLRKWEALQGYEL VQDGDATDQADAR REALDLVKRFRDVL VLFKLTGEARFEGRAAPFDLKP FRALFANPATFDRLFMATPTTARPAEDDPEGDGASE PELRVARTLRGLRQIARYNHMAVLSDLFAKHKVRD EEVARLAEIEDETQEK SQIVAAQELRTDLHDK VMKC HPKTISPEERQSYAAAIKTIEEHRFLVGRVYLG DHLR LHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGA GADWAVTIAGAANTDARTQTRKDLAHFNVLDRAD GTPDLTALVNRAREMMA YDRKRKNAVPR SILDMLA RLGLTLKWQMKDHLQDATITQAAIKHLDKVR LTV GGPAAVTEARF SQDY LQMVA AVFNGS VQNP KPRRR DDGD AWHKPPKP ATAQSPDQKPPNK AP SAGSRLP PPQVGEVYEGVVVKVIDTGS LGFLAVEGVAGNIGLH ISRLRRIRED AII VGRRYRFRVEI YVPPK SNT SKLNAA DLVRID</p>
<p>Rhodobacter capsulatus R121 (SEQ ID Nos. 45 and 46)</p>	<p>GCCTCACAT CACCGCCAA GACGACGG CGGACTGA AC</p>	<p>MQIGK VQGRTI SEFGDPAGGLKRKI STDGKNRKELP AHLSSDPKALIGQWISGIDKIYRKPD SRKSDGKAIHSP TPSKMQFDARDDLGE AFWKL VSEAGLAQD SDYDQF KRRLHP YGDKF QPAD SGAKLKFE ADPPEPQ AFHGR WYGAMSKRGNDAKELAAALYEHHLVDEKRIDGQP KRNPKTDFK APGLVVARALGIES SVLPRGMARLARN WGEETIQTYFVVDVAASVKEVAKAAVSAAQAFDPP RQVSGRSLSPKVGAFALAEHLERVTGSKRCSFDPAAG PSVLALHDEVKKTYYKRLCARGKNAARAFPADKTEL LALMRHTHENRVRNQMVVMGRVSEYRGQQAGDLA QSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMK AWIDPMGKI VNTEKNDRDLT AA VNIRQ VISNKEMV AEAMARRGIYFGETPELDRLGAEGNEGFV FALLRYL RGCNRNQT FHLGARAGFLKEIRKELEKTRWGKAKEA EHVVLTDKT VAAIRAIIDND AKALGARLL ADL SGAF</p>

		<p>VAHYASKEHFSTLYSEIVKAVKDAPEVS SGLPRLKL LLKRADGVRGYVHGLRDTRKHAFATKLP PPPAPREL DDPATKARYIALLRLYDGPFRAYASGITGTALAGPA ARAKEAATALAQSVNVTKAYS DVM EGRSSRLRPPN DGETLREYLSALTGETATEFRVQIGYESDSENARKQ AEFIENYRRDMLAFMFEDYIRAKGFDWILKIEPGAT AMTRAPVLPEPIDTRGQYEHWQAALYLV MHFVPAS DVSNLLHQLRKWEALQGKYELVQDGDATDQADAR REALDLVKRFRDVLVFLKTGEARFEGRAAPFDLKP FRALFANPATFDRLFMATPTTARPAEDDPEGDGASE PELRVARTLRGLRQIARYNHMAVLSDLFAKHKVRD EEVARLAEIEDETQEK SQIVAAQELRTDLHDKVMKC HPKTISPEERQSYAAAIKTIEEHRFLVGRVYLG DHLR LHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGA GADWAVTIAGAANTDARTQTRKDLAHFNVLDRAD GTPDLTALVNRAREMMA YDRKRKNAVPRSILDMLA RLGLTLKWQMKDHL LQDATITQAAIKHLDKVRLTV GGPAAVTEARFSDYLQMVAAVFNGSVQNP KPRRR DDGDAWHKPPKPATAQS QPDQKPPNKAP SAGSRLP PPQVGEVYEGVVVKVIDTGS LGFLAVEGVAGNIGLH ISRLRRIREDAll VGRRYRFRVEIYVPPK SNT SKLNAA DLVRID</p>
<p>Rhodobacter capsulatus DE442 (SEQ ID Nos. 47 and 48)</p>	<p>GCCTCACAT CACCGCAA GACGACGG CGGACTGA AC</p>	<p>MQIGKVQGRTISEFGDPAGGLKRKISTDGKNRKELP AHLSSDPKALIGQWISGIDKIYRKPDSRKSDGKAIHSP TPSKMQFDARDDLGEAFWKL VSEAGLAQSDYDQF KRRLHPYGDKFQPAD SGAKLKFEADPPEPQAFHGR WYGAMSKRGNDAKELAAALYEHLHVDEKRIDGQP KRNPKTDKFAPGLVVARALGIES SVLPRGMARLARN WGEIIIQTYFVVDVAASVKEVAKAAV SAAQAFDPP RQVSGRSLSPKVG FALAEHLERVTGSKRCSFDPAAG PSVLALHDEVKKTYKRLCARGKNAARAF PADKTEL LALMRHTHENVRNQMVRMGRVSEYRGQQAGDLA QSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMK AWIDPMGKIVNTEKNDRDLTAAVNIRQVISNKEMV AEAMARRGIYFGETPELDRLGAEGNEGFV FALLRYL RGCNRNQTFFHLGARAGFLKEIRKELEKTRWGKAKEA EHVVLTDKTVA AIRAIIDNDAKALGARLLADL SGAF VAHYASKEHFSTLYSEIVKAVKDAPEVS SGLPRLKL LLKRADGVRGYVHGLRDTRKHAFATKLP PPPAPREL DDPATKARYIALLRLYDGPFRAYASGITGTALAGPA ARAKEAATALAQSVNVTKAYS DVM EGRSSRLRPPN DGETLREYLSALTGETATEFRVQIGYESDSENARKQ AEFIENYRRDMLAFMFEDYIRAKGFDWILKIEPGAT AMTRAPVLPEPIDTRGQYEHWQAALYLV MHFVPAS DVSNLLHQLRKWEALQGKYELVQDGDATDQADAR REALDLVKRFRDVLVFLKTGEARFEGRAAPFDLKP FRALFANPATFDRLFMATPTTARPAEDDPEGDGASE PELRVARTLRGLRQIARYNHMAVLSDLFAKHKVRD EEVARLAEIEDETQEK SQIVAAQELRTDLHDKVMKC</p>

		<p>HPKTISPEERQSYAAAIKTIEEHRFLVGRVYLGDHRL LHRLMMDVIGRLIDYAGAYERDTGTFLINASKQLGA GADWAVTIAGAANTDARTQTRKDLAHFNVLD RAD GTPDLTALVNRAREMMA YDRKRKNAVPRSILDM LA RLGLTLKWQMKDHL LQDATITQAAIKHLDKVRLTV GGPAAVTEARF SQDY LQMVA AVFNGSVQNP KPRRR DDGDAWHKPPK PATAQS QPDQKPPNKAP SAGSRLP PPQVGEVYEGVVVKVIDT GSLGFLAVEGVAGNIGLH ISRLRRIRED AII VGRRYRFRVEI YVPPK SNT SKLNAA DLVRID</p>
<p>Leptotrichia buccalis C-1013- b (SEQ ID Nos. 49 and 50)</p>	<p>GGATTTAGA CCACCCCAA AAATGAAG GGGACTAA AACA</p>	<p>MKVTKVGGISUKKYTSEGRLVKSESEENRTDERLSA LLNMRLDMYIKNP SSETKENQKRIGK LKKFF SNKM VYLDKNTLSLKN GKKENIDRE YSETDILE SDVRDKK NFAVLKKIYLNENVNSEELEVFRNDIKK KLNKINSLK YSFEKNKANYQKINENNIEK VEGKSKRNIIYDYRES AKRDAYVSNVKEAFDKLYKEEDIAKL VLEIENLTKL EKYKIREFYHEIIGRKNDKENFAKIIYEEIQNVNMMK ELIEKVPDMSELKKSQVFYKY YLDKEELNDKNIKYA FCUFVEIEMSQLKNYVYKRLSNISNDKIKRIFEYQN LKKLIENKLLNKLD TYVRNCGKYNYYLQDGEIATSD FIARNRQNEAFLRNIIIGVS SVA YFSLRNILETENENDI TGRMRGKTVKNNKGEEKYVS GEVDKIYENENK KNE VKENLKM FY SYDFNMDNKNEIEDFF ANIDEAIS SIRH GIVHFNLELEGKDIF AFKNIAP SEISKMFQNEINEKK LKLKIFRQLNSANVFRYLEKYKILNYLKRTRFEFVNK NIPFVPSFTKL YSRIDDLKNSLGIYWKTPKTND DDKT KEIIDAQIYLLKNIYYGEFLNYFMSNNGNFFEISKEIIE LNKNDKRNLKTGF YKLQKFEDIQEKIPKE YLANIQSL YMINAGNQDEEEKDTYIDFIQKIFLKGFM TYLANNG RLSLIYIGSDEETNT SLAEKKQEFDKFLKK YEQNNNI KIPYEINEFLREIKLGNILKYTERLNM FYLILKLLNHK ELTNLKGSLEK YQSANKEE AFSDQLELINLLNLDNN RVTEDFELEADEIGKFLDFNGNKVKDNKELKKFDTN KIYFDGENIIKHRAFYNIKKYGMLN LLEKIADKAGY KISIEELKKYSNKKNEIEKNHKMQENLHRKYARPRK DEKFTDED YE SYKQAIENIEE YTHLKNKVEFNELNLL QGLLLRILHRLVGYTSIWERDLRFRLKGEFPENQYIE EIFNFENKKNV KYKGGQIVEKYIKFYKELHQNDEVK INKYS SANIKVLKQEKKDL YIRNYIAHFNYIPHA EISL LEVLENLRKLLSYDRK LKNV MKSVVDILKEYGFV ATFKIGADKKIGIQTLESEKIVHLK NLKKKLM TDR NSEELCKLVKIMFEYKMEEKKSEN</p>
<p>Herbinix hemicellulosilyti ca (SEQ ID Nos. 51</p>	<p>GTAACAATC CCCGTAGAC AGGGGAAC TGCAAC</p>	<p>MKLTRRRISGNSVDQKITAAFYRDMSQGLLYYDSED NDCTDKVIESMDFERSWRGRILKNGEDDKNPFYMF VKGLVGSNDKIVCEPIDVSDPDNLDILINKNL TGFG RNLKAPD SNTLENLIRKI QAGIPEEEVLP ELKKIKE MIQKDIVNRKEQLL KSIKNNRIPF SLEGSKL VPSTKK MKWLFKLIDVPNKTFNEKMLEKYWEIYDYDKLKAN</p>

<p>and 52)</p>		<p>ITNRLDKTDDKKARSISRAVSEELREYHKNLRTNYNR FVSGDRPAAGLDNGGSAKYNPDKEEFLFLKEVEQY FKK YFPVK SKHSNK SKDKSLVDK YKNYCSYK VVKK EVNRSIINQLVAGLIQQGKLLYFYFYNNTWQEDFLN SYGLSYIQVEEAFKKSVMTSLSWGINRLTSFFIDDSN TVKFDDITTKKAKEAIESNYFNKLRTCSRMQDHFKE KLAFFYPVYVKDKKDRPDDDIENLIVLVKNAIESVS YLRNRTFHFES SLELLKELDDKNSGQNKID YSVA AEFIKRDIEENLYDVFREQIRSLGIAEYKADMISDCF KTCGLEFALYSPKNSLMPAFKNVYKRGANLNKAYI RDKGPKETGDQGQNSYKALEEYRELTWYIEVKNND QSYNAYKNLLQLIYYHAFLPEVRENEALITDFINRTK EWNRKETEERLNTKNNKKHKNFDENDDITVNTYRY ESIPDYQGESLDDYLKVLQRKQMARAKEVNEKEEG NNNYIQFIRDVVVWAFGAYLENKLKNYKNELQPPL S KENIGLNDTLKELFPEEKVKSPFNIKCRFSISTFIDNK GKSTDNTSAEAVKTDGKEDEKDKKNIKRKDLLCFY LFLRLDENEICKLQHFIK YRC SLKERRFPGNRTKL EKETELLAELEELMELVRFTMPSIPEISAKAESGYDT MIKK YFKDFIEKKVFKNPKT SNL YYHSD SKTPVTRK YMALLMRSAPLHLYKDIFKGYLITKKECLEYIKLS NIIKDYQNSLNELHEQLERIKLKSEKQNGKDSL YLDK KDFYKVKEYVENLEQVARYKHLQH KINFESLYRIFR IHVDIAARMVGYTQDWERDMHFLFKALVYNGVLEE RRFEAIFNNDDNNDGRIVKKIQNNLNKNRELVSM LCWNKKNLKNNEFGAIIWKRNP IAHLNHFTQTEQN SK SSLESLeNSLRILLAYDRKRQNAVTKTrNDLLLND YHI RIKWEGRVDEGQIYFNIKEKEDIENEPIIHLKHLHKK DCYIYKNSYMFDKQKEWICNGIKEEVYDKSILKCIG NLFKFDYEDKNKSSANPKHT</p>
<p>[Eubacterium] rectal e (SEQ ID Nos. 53 and 54)</p>	<p>TGTGAAAGT AGCCCGATA TAGAGGGC AATAACGT</p>	<p>MLRRDKE VKKLYNVFNQIQVGTKPKKWNNDEKL SP EENERRAQ QKNIKMKN YKWREACSKY VESSQRIIND VIFY SYRKAKNKLRYMRKNEDILKKMQEAEKLSKF S GGKLEDFVAYTLRKS L VVSKYDTQEFDSLAA MVVF LECIGKNNI SDHEREIVCKLLELIRKDF SKLDPNVKGS QGANIVRSVRNQNMIVQPQGD RFLFPQVYAKENET VTNKNVEKEGLNEFLN YANLDDEKRAE SLRKL RRI LDVYF SAPNHYEKDM DITLSDNIEKEKFNVWEKHEC GKKETGLFVDIPDVLMEAEAENIKLDA VVEKRERKV LNDRVRKQNIICYRYTRAVVEKYNSNEPLFFENNAI NQYWIHHIEN AVERILKNCKAGKLFKLRKGYLAEK VWKDAINLISIKYIALGKAVYNFALDDIWKDKKNKE LGI VDERIRNGIT SFDYEMIK AHENLQREL AVDIAF S VNNLARAVCDMSNLGNKESDFLL WKRNDIADKLL NKDDMAS VSAVLQFFGGKS SWDTNIFKDAYKGGKK YNYEVRFIDDLRKAICYCARNENFHFKTALVNDEKW NTELFGKIFERETEFCLNVEKDRFY SNNLYMFYQVS ELRNMLDHLYSRSVSRAAQVPSYNSVIVRTAFPEYIT NVLGYQKPSYDADTLGKWYSACY YLLKEIYYNSFL</p>

		<p>QSDRALQLFEKSVKTLWSWDDKKQQRRAVDNFKDHFS DIKSACTSLAQVCQIYMTEYNQQNNQIKKVRS SND SI FDQPVYQHYKVLLKKAIANAFADYLKNNKDLFGFI GKPFKANEIREIDKEQFLPD WTSRKYE ALCIEVSGSQ ELQKWYIVGKFLNARSLNLMVGSMSRSYIQYVTDIKR RAASIGNELHVSVHDVEKVEKWVQVIEVCSLLASRT SNQFEDYFNDKDDYARYLKS YVDFS NVDMPS EYSA LVDFSNEEQSDLYVDPKNPKVNRNIVHSKLFAADHI LRDIVEPVSKDNIEEFYSQKAEIAYCKIKGKEITAE EQ KAVLKYQKLKNRVELRDI VE YGEIINELLGQLINW SF MRERDLYFQLGFHYDCLRNSDKKPEGYKNIKVDE NSIKDAILYQIIGMYVNGVTVYAPEKDGDKLKEQCV KGGVGVKVSFAHRYSKYLGLNEKTLYNAGLEIFEV VAEHEDIINLRNGIDHFKY YLGDYRSMLS IYSEV FDR FFTYDIKYQKNVLNLLQNILLRFINVIVE PILESGFKTI GEQTKPGAKLSIRSISDTFQYKVKGGTLITDAKDER YLETIRKILYYAENEEDNLKKS VVVTNADKYEKNKE SDDQNKQKEKKNKDNKGKKNEETKSDAEKNNNER LSYNPFANLNFKLSN</p>
<p>Eubacted aceae bacterium CHKCI004 (SEQ ID Nos. 55 and 56)</p>	<p>GTAGATAGC CCGATATAG AGGGCAAT AAAC</p>	<p>MKISKE SUKRTAVAVMEDRVGGVVYVPGGS GIDLS NNLKKRSM DTKSLYNVFNQIQAGTAPSEYEWKDYL SEAENKKREAQKMIQKANYELRRECEYAKKANLA VSRIIF SKKPKKIF SDDDIISUMKKQRLSKFKGRMEDF VLIALRKS L VVSTYNQEVFDSRKAATVFLKNIGKKNI SADDERQIKQLMALIRED YDKWNPDKD SSDKKESS GTKVIRSIEHQNMVIQPEKNKLSL SKISNVGKKTCTK QKEKAGLDAFLKEYAQIDENSRMEYLKCLRLLD YFAAPSSYIKGAAVSLPENINF SSELNVWERHEAAK KVNINFVEIPESLLNAEQNNNKINKVEQEHSLEQLRT DIRRRNITCYHFANALAADERYHTLFFENMAMNQF WIHHMENAVERILKKNVGTFLFKLRIGYL SEKVWK DMLNLLSIKYIALGKAVYHFALDDIWKADIWKDAS DKNSGKINDLTLKGIS SFDYEMVKAQEDLQREMAV GVAFSTNNLARVTCKMDDLSD AESDFLLWNKEAIR RHVKYTEKGEIL SAILQFFGGRSLWDESLFEKAYS DS NYELKFLDDLKRAIYAARNETFHFKTA AIDGGSWNT RLFGLFEKEAGLCLNVEKNKFYSNNLVLFYKQEDL RVFLDKL YGKEC SRAAQIPSYNTILPRK SF SDFMKQL LGLKEPVYGSAILDQWYSACY YLFKEVYYNLFQDS SAKALFEKAVKALKGADKKQEKAVE SFRKRYWEIS KNASLAEICQSYITEYNQQNNKERK VRSANDGMFNE PIYQHYKMLLKEALKMAFASYIKNDKELKFVYKPT KLFEVSQDNFLPNWNSEKYNTLISEVKNSPDLQKWY IVGKFMNARMLNLLL GSMRSYLQYVSDIQKRAAGL GENQLHL SAENVGQVKKWIQVLEVCLLSVRISDKF TDYFKDEEEYASYLKEYVDFEDSAMPDYSALLAFS NEGKIDLYVDASNPKNRNIQAKLYAPDMVLKKV VKKISQDECKEFNEKKEQIMQFKNKGDEVSWEEQQ KILEYQKLKNRVELRDLSE YGELINELLGQLINW SYL</p>

		RERDLL YFQLGFHY SCLMNESKKPDAYKTIRRGTVS IENAVLYQIIAMYINGFPVYAPEKGEKLPQCKTGSAG QKIRAFQCWASMVEKKKYELYNAGLELFEVVKEHD NIIDLNRNKIDHFKYYQGNSILAL YGEIFDRFFTYDM KYRNNVLNHLQNILLRHNVIKPIISKDKKEVGRGKM KDRAAFLLEEVS SDRFTYKVKEGERKID AKNRLYLE TVRDILYFPNRAVNDKGEDVIICSKKAQDLNEKKAD RDKNHDK SKDTNQQKEGKNQEEK SENKEPYSDRMT WKPFAGIKLE
Blautia sp. Marseille-P2398 (SEQ ID Nos. 57 and 58)	ATCTAATGA GAACATCCC AAGATAAC GGGAAATA AC	MKISKVDHVKSGIDQKLSSQRGMLYKQPQKKYEGK QLEEHVRNLSRKAKALYQVFPVSGNSKMEKELQIIN SFIKNILLRLDSGKTSEEIVGYINTYSVASQISGDHIQE LVDQHLKESLRKYTCVGDKRIYVPDIIVALLKSKFNS ETLQYDNSSELKILIDFIREDYLKEKQIKQIVHSIENNST PLRIAIEINGQKRLIPANVDNPKKSYIFEFLKEYAQSDP KGQESLLQHMRYLILLYLYGPDKITDDYCEEIEAWN FGSIVMDNEQLFSEEASMLIQDRIYVNQQIEEGRQSK DTAKVKKNKSKYRMLGDKIEHSINESVVKHYQEAC KAVEEKDIPWIKYISDHVMSVYSSKNRVDLTKLSLP YLAKNTWNTWISFIAMKYVDMGKGVYHFAMSDVD KVGKQDNLIIGQIDPKF SDGIS SFD YERIKAEDDLHRS MSGYIAFAVNNFARAICSDDEFKKNRKEDVLTVGLD EIPLYDNVKKRLLQYFGGASNWDDSIIDIIDDKDLVA CIKENLYVARNVNFHFAGSEKVQKKQDDILEEIVRK ETRDIGKHRYKVFYSNNVAVFYCEDEDIKLMNHL YQ REKPYQAQIPSYNKVISKTYLPDLIFMLLKGKNRTKI SDPSIMNMFRGTFYFLLKEIYYNDFLQASNLKEMFC EGLKNNVKNKKSEKPYQNFMRRFEELENMGMDFG EICQQIMTDYEQQNKQKKKTATAVMSEKDKKIRTL DNDTQKYKHFRTLLYIGLREAFIYKDEKNKEWYE FLREPVKREQPEEKEFVNKWKLNQYSDCSELILKDS LAAAWYVVAHFINQAQLNHLIGDIKNYIQFISDIDRR AKSTGNPVSESTEIQIERYRKILRVLEFAKFFCGQITN VLTDYQDENDFSTHVGHYVKFEKKNMEPAHALQ AFSNSLYACGKEKKKAGFYDGMNPVNRNITL AS MYGNKKLLENAMNPVTEQDIRKYYSLMAELD SVLK NGAVCKSEDEQKNLRHFQNLKNRIELVDVLTLSLV NDLVAQLIGWVYIRERDMMYLQLGLHYIKLYFTDS VAEDSYLRTL DLEEGSIADGAVLYQIASLYSFLNPM YVKPNKSSVYCKKHVNSVATKFDIFEKEYCNGDET VIENGLRLFENINLHKDMVKFRDYLAHFKYFAKLDE SILELYSKAYDFFFSYNIKLKKSVS YVLTNVLLSYFIN AKLSFSTYKS SGNKTVQHRTTKISVVAQTDYFTYKL RSIVKNKNGVESIENDDRRCEVVNIAARDKEFVDEV CNVINYNSDK
Leptotrichia sp. oral taxon 879	ATAGACCAC CCCAATATC GAAGGGGA	MGNLFGHKRWYEVDRDKKDFKIKRKYKVKRNYDGN KYILNINENNNKEKIDNNKFIGEFVNYKKNNNVLKE FKRKFHAGNILFKLKGKEEIIIRIENDDFLETEEVVLY

<p>str. F0557 (SEQ ID Nos. 59 and 60)</p>	<p>CTAAAAC</p>	<p>IEVYGKSEKLEKALEITKKKIIDEAIRQGITKDDKKEIK RQENEEIEIDIRDEYTNKTLNDCSILRIENDELETK KSIYEIFKNINMSLYKIIKIIENETEKVFENRYEEHL REKLLKDNKIDVILTNFMEIREKIKSNLEIMGFVKFY LNVSGDKKSENKKMFVEKILNTNVDLTVEDIVDFI VKELKFWNITKRIEKVKKFNNEFLENRRNRITYIKSY VLLDKHEKFKIERENKKDKIVKFFVENIKNNSIKEKIE KILAEFKINELIKKLEKELKKGNCDEIFGIFKKHYKV NFD SKKFSNK SDEEKEL YKIIYRYLKGRIEKIL VNEQ KVRLKKMEKIEIEKILNESILSEKILKRVKQYTLHEIM YLGKLRHNDIVKMT VNTDDF SRLHAKEELDLELITF FASTNMELNKIFNGKEKVTDFFGFNLNGQKITLKEK VPSFKLNILKKNLFINNENNIDEKLSHFYSFQKEGYL LRNKILHNSYGNIQETKNLKGGEYENVEKLIKELKVS DEEISKSLSDVIFEGKVDIINKINSLKIGEYKDKKYL PSF SKIVLEITRKFREINKDKLFDIE SEKIILNAVKYVN KILYEKIT SNEENEFLKTLPKL VKK SNNKKENKNLL SIEEYYKNAQV SSSKGDKKAIKKYQNKVTNAYLEYL ENTFTEIIDF SKFNLNYDEIKTKIEERKDNK SKIID SIS TNINITNDIEYIISIFALLNSNTYINKIRNRFFATSVWLE KQNGTKE YD YENII SILDEVLLINLLRENNITDILDK NAIDAKIVE NDETYIKNYIFE SNEEKLKKRLFCEELV DKEDIRKIFEDENFKFK SFIKKNEIGNFKINF GILSNLE CNSEVEAKKIIGKNSKKLESFIQNIIDEYKSNIRTLF SS EFLEKYKEEIDNL VEDTESENKNKFEKIYYPKEHKNE LYIYKKNLFLNIGNPNFDKI YGLISKDIKNVDTKILFD DDIKKNKISEIDAILKNLNDKLNKNGYSNDYKAKYVVK LKENDDDFF AKNIQENENYS SFGFEKDYNKVSEYKKI RDLVEFNLYLNKIESYLIDINWKLAIQMARFERDMHYI VNGLRELGIKLSGYNTGISRAYPKRNGSDGFYTTTA YYKFFDEE SYKKFEKIC YGFGIDLSENSEINKPENE SI RNYISHFYIVRNPFADYSIAEQIDRVSNLLSYSTRYNN STYASVFEVFKKDVNLDYDELKKKFRLLIGNNDILER LMKPKKVSVLELESYNSDYIKNLIIELLTKIENTNDTL</p>
<p>Bergeyella zoohelcum WP_002664492 (SEQ ID Nos. 61 and 62)</p>	<p>GTTGGAAC GCTCTCATT TTGGAGGGT AATCACAAC</p>	<p>MENKTSLGNNIYYNPFKPKQDKSYFAGYFNAAMENT DSVFRELKRLKGKE YTSENFFDAIFKENISLVE YER YVKLLSDYFPMARLLDKKE VPIKERKENFKKNFKGII KA VRDLRNFYTHKEHGE VEITDEIF GVLDEMLKSTV LTVKVVKTDTKTEILKKSIEKQLDILCQKKLEYL RDTARKIEEKRRNQRRERGEKELVAPFKYSDKRDDLI AAIYNDAFDVYIDKKKDSLKESKAKYNTKSDPQQE EGDLKIPI SKNGVVFLSLFLTKQEIHAFAKSKIAGFKA TVIDEATVSEATVSHGKNSICFMATHEIFSHLAYKKL KRKVRTAEINYGEAENAEQLSVYAKETLMMQMLDE LSKVPDVVYQNL SEDVQKTFIED WNE YLKENNGDV GTMEEEQVIHPVIRKRYEDKFNYFAIRFLDEFAQFPT LRFQVHLGNYLHDSRPKENLISDRRIKEKITVFGRLS</p>

		<p>ELEHKKALFIKNTETNEDREHYWEIFPNPNYDFPKEN ISVNDKDFPIAGSILDREKQPVAGKIGIKVKLLNQQY VSEVDKAVKAHQLKQRKASKPSIQNIIEEIVPINESNP KEAIVFGGQPTAYLSMNDIHSILYEFFDKWEKKKEK LEKKGEKELRKEIGKELEKKIVGKIQAQIQQIIDKDT NAKILKPYQDGNST AIDKEKLIKDLKQEQNILQKLL DEQTVREKEYNDFIAYQDKNREINKVRDRNHKQYL KDNLKRKYPEAPARKEVLYYREKGGKVA VWLANDI KRFMPTDFKNEWKGEQHSLLQKSLAYYEQCKEELK NLLPEKVFQHLPFKLGGYFQQKYL YQFYTCYLDKRL EYISGLVQQAENFKSENKVFKKVENECFKFLKKQNY THKELDARVQSILGYPIFLERGFMDKPTIIGKTKFK GNEALFADWFRYYKEYQNFQTFYDTENYPLVELEK KQADRKRKTKIYQQKKNVFTLLMAKHIFKSVFKQ DSIDQF SLEDLYQSREERLGNQERARQTGERNTNYI WNKTVDLKLCDGKITVENVKLKNVGDVFIKYEYDQR VQAFKYEENIEWQ AFLIKE SKEEENYPYVVEREIEQ YEKVREELLKEVHLIEEYILEKVKDKKILKKGDNQ NFKYYILNGLLKQLKNEDVESYKVFNLNTEPEDVNI NQLKQEATDLEQKAFVLT YIRNKFAHNQLPKKEFW DYCQEKYKIEKEKTYAEYFAEVFKKEKEALIK</p>
<p>Prevotella intermedia WP_036860899 (SEQ ID Nos. 63 and 64)</p>	<p>GTTGCATCT GCCTGCTGT TTGCAAGGT AAAAACAA C</p>	<p>MEDDKKTTDSIRYELKDKHFWAAFLNLARHNVYIT VNHINKILEEGEINRDGYETTLKNTWNEIKDINKKDR LSKLIIKHFPFLEAATYRLNPTDTTKQKEEKQAEAQS LESLRKSFFVFIYKLRDLRNHYSHYKHSKSLERPFE EGLLEKMYNIFNASIRLVKEDYQYNKDINPDEDFKH LDRTEEEFN YYFTKDNENITE SGLLFFVSLFLEKKD AIWMQQKLRGFKDNRENKKKMTNEVFCRSRMLLP KLRLQSTQTQDWILLDMLNELIRCPKSLYERLREED REKFRVPIEIADEDYDAEQEPFKNTLVRHQDRFPYFA LRYFDYNEIFTNLRFQIDLGT YHFSIYKKQIGDYKES HHLTHKLYGFERIQEFTKQNRPDWRKFKVTFNSFE TSKEPYIPETTPHYHLENQKIGIRFRNDNDKIWPSLKT NSEKNEK SKYKLDK SFQAE AFLSVHELLPMMFYYLL LKTENTDNDNEIETKKKENKNDKQEKHKIEEIIENKI TEIYALYDTFANGEIKSIDELEEYCKGKDIEIGHLPKQ MIAILKDEF1KVMATE AERKQEEMLVDVQK SLES NQINEEIEENVERKNS SLKSGKIASWL VNDMMRFQPV QKDNKPLNNSKANST EYQLLQRTLAFSGSEHERL APYFKQTKLISSNPHPFKDT EWEKCNNILSFYRSY LEAKKNFLES LKPEDWEKNQYFLK LKPKTKPKTLV QGWKNGFNLPRGIFTEPIRKWFMKHRENITVAELKR VGLVAKVIPLFF SEEYKDSVQPFYNYITFVGNINKP DEKNFLNCEERRELLRKKKDEFKKMTDKEKEENPS YLEFKSWNKFERELRLVRNQDIVTWLLCMELFNKK KIKELNVEKIYLNINTNTTKKEKNTEEKNGEENIK EKNNILNRIMPRLPIK VYGRENFSKNKKKIRRNTF FTVYIEEKGTLLKQGNFKALERDRRLGGLF SFVKT PSKAE SKSNTISKLRVEYELGEYQKARIEI IKDMLALE</p>

		<p>KTLIDKYNSLDTDNFNKMLTDWLELKGEPAKASQ NDVDLLIAVRNAF SHNQYPMRNRIAFANINPFLSSA NTSEEKGLGIANQLKDKTHKTIEKIIIEKPIETKE</p>
<p>Prevotella buccae WP_004343973 (SEQ ID Nos.65 and 66)</p>	<p>GTTGCATCT GCCTTCTTT TTGAAAGGT AAAAACAA C</p>	<p>MQKQDKLFVDRKKNAIFAFPKYITIMENKEKPEPIYY ELTDKHFVA AFLNLARHNVYTTINHINRRLEIAELK DDGYMMGIKGSWNEQAKKLDKKVRLRDLIMKHFP FLE AAA YEMTNSK SPNNKEQREKEQ SEALSLNNLKN VLFIFLEKLQVLRNYYSHYKYSEESPKPIFETSLLKN MYKVF DANVRLVKRDYMHENIDMQRDFTHLNRK KQVGRTKNIIDSPNFHYHFADKEGNMTIAGLLFFVSL FLDKKDAIWMQKKLKGFKDGRNLREQMTNEVFCRS RISLPKLENVQTKDWMQLDMLNEL VRCPKSLYE RLREKDRESFKVPFDIF SDDYNAEEEPFKNTLVRHQD RFPYFVLR YFDLNEIFEQLRFQIDLGT YHF SIYNKRIG DEDEVRLH LTHHL YGFARIQDFAPQNQPEWRKLVK DLDFHET SQEPYISK TAPHYHLENEKIGIKFCSAHNN LFPSLQTDKTCNGRSKFN LGTQFTAEAFLSVHELLP MMFYLLLT KDYSRKESADKVEGIIRKEISNIYAIYD AFANNErNSIADL TRRLQNTNILQGHLPKQMISILKG RQKDMGKEAERKIGEMIDDTQRRLDLLCKQTNQKI RIGKRNAGLLKSGKIADWL VNDMMRFQP VQKDQN NIPrNNSKANST EYRMLQRALALFGSENFRLKAYFN QMNLVGN DNPHFPLAETQWEHQTNILSFYRNYLEA RKKYLKGLKPQNWKQYQHFLILKVQKTNRNTLV TG WKNSFNLPRGIFTQPIREWFEKHNSKRIYDQILSFD RVGFVAKAIPL YFAEEYKDNVQPFYDYPFNIGNRLK PKKRQFLDKKERVELWQKNKELFKNYPSEKKKTDL AYLDFL SWKKFERELRLIKNQDI VTWLMFKELFNM ATVEGLKIGEIH LRDIDTNTANEESNNILNRIMPMLK PVKTYETDNKGNILKERPLATFYIETETKVLKQGNF KALVKDRRLNGLFSF AETDLNLEEHPISKLSVDLEL IKYQTTRISIFEMTLGLEKKLIDKYSTLPTDSFRNMLE RWLQCKANRPELKNYVNSLIAVRNAF SHNQYPMYD ATLFAEVKKFTL FPSVDTKKIELNIAPQLLEIVGKAIK EIEKSENKN</p>
<p>Alistipes sp. ZOR0009 WP_047447901 (SEQ ID Nos. 67 and 68)</p>	<p>GCTGTTATA TCCTTACCT TTGTAAGGG AAGTACAG C</p>	<p>MSNEIGAFREHQFAYAPGNEKQEEATFATYFNLALS NVEGMMFGEVESNPDKIEKSLDTPAILRQIASFIW LSKEDHPDKAYSTEEVKVIVTDLVRRLCF YRNYF SH CFYLDTQYFYSDDELVDTTAIGEKL PYNFHHFITNRLF RYSLPEITLFRWNEGERKYEILRDGLIFFCCLFLKRGQ AERFLNELRFFKRTDEEGRIKRTIFTKYCTRESHKHIG IEEQDFLIFQDIIGDLNRV PKVCDGVVDLSKENERYIK NRET SNESDENKARYRLLIREKDKFPY YLMRYIVDF GVLPCITFKQNDYSTKEGRGQFH YQDAAVAQEERC YNFVVRNGNVYYSYMPQAQNVVRISELQGTISVEEL RNMVYASrNGKDVNKSVEQYLYHLHLLY EKILTISG QTIKEGRVDVEDYRPLLDKLLLRPASNGEELRREL R KLLPKRVCDLLSNRFDCSEGVSAVEKRLKAILLRHE</p>

		<p>QLLSQNPALHIDKIKSVIDYLYLFFSDDEKFRQQPTE KAHRGLKDEEF QMYHYLVGD YDSF1PLALWKELE A SGRLKPEMRKLT SATSLHGL YMLCLKGT VEWCRKQ LMSIGKGTAKVEAIADRVGLKLYDKLKEYTPEQLER EVKLVVMHGYAAAATPKPKAQAAIPSKLTELRFYSF LGKREMSFAAFIRQDKKAQKLWLRNFYTVENIKTLQ KRQAAADAACKLYNLVGEVERVHTNDKVLVLA QRYRERLLNVGSKCAVTLDNPERQQKLADVYEVQN AWL SIRFDDLDF TLTHVNL SNLRKAYNLIPRKHIL AF KEYLDNRVKQKLCEECRNVRKEDLCTCCSPRYSN LTSWLKENHSESSIEREAATMMLLD VERKLL SFLLD ERRK APIEYGFIPF SALVKECRL ADAGLCGIRND VL HDNVISYADAIGKLSAYFPKEASEAVEYIRRTKEVRE QRREELMANSSQ</p>
<p>Prevotella sp. MA2016 WP_03 6929 175 (SEQ ID Nos. 69 and 70)</p>	<p>GTTGTAGAA GCTTATCGT TTGGATAGG TATGACAAC</p>	<p>MSKECKKQRQEKKRRLQKANF SISLTGKHVFGAYF NMARTNFVKTINYILPIAGVRGNYSQINKMLHAL FLIQAGRNEELTTEQKQWEKKLRLNPEQQTKFQKLL FKF1FPVLGPMMAADVADFKAYLNKKKSTVQTEDET AMLKGVSLADCLDIICLMADTLTECRNFYTHKDPYN KPSQLADQYLHQEMIAKKLDKVVVASRRILKDREG LSVNEVEFLTGDHLHQEVLKDEFNAKVKDGKVM KTFVEYDDFYFKISGKRLVNGYTVTTKDDKPVNVN TMLPALSDFGLLYFCVLFLSKPYAKLFIDEVRLFEYS PFDDKENMIMSEMLSIYRIRTPRLHKIDSHDSKATLA MDIFGELRRCPMELYNLLDKNAGQPFHDEVKHPNS HTPDVSKRLRYDDRFPALRYIDETELFKRIRFQLQ LGSFRYKFDKENCIDGRVRRRIQKEINGYGRMQE VADKRMWKWGDLIQKREERSVKLEHEELYINLDQF LEDTADSTPYVTDRRP AYNIIANRIGLYWEDSQNPK QYKVFDENGMYIPELVVTEDEKKAIPKMPAPRCALSV YDLPAMLFYELREQQDNEFPQAEQVIEYEDDYRK FFKAVAEGKLPFKRPKEFRDFLKKKEYPKLRMADIP KKLQLFLC SHGLCYNNKPETVYERLDRLTLQFILEER ELHIQNRLEH YQKDRDMIGNKDNQ YGKKSFSVVRH GALARYLAQSMMEWQPTKLKDKKEKGHDKLTGLNY NVLTAYLATYGHQPVPPEGFTPTLEQVLINAHLIGG SNPHPFINKVLALGNRNIEELYLHYLEEELKHRSRIQ SLSSNPSDKALSALPFIHHRMRYHERTSEEMMALA ARYTTIQLPDGLFTPYILEILQKHYSNDLQNALSQ DVPVKLNPTCNAAYLITLFYQTVLKDNAQPFYLSDK TYTRNKDGEKAESF SFKRAYELF SVLNNNKDTPFP EMIPFLTSDEIQLERLSAKLLDGDGNPVPEVGEKGP ATDSQGNITWKRRIYSEVDDYAEKLTDRDMKISFKG EWEKLPRWKQDKIIKRRDETRRQMRDELLQRMPRY IRDIKDNERTLRRYKTQDM VLFLLAEKMTNIISEQS SEFNWKQMRLSKVCNEAFLRQTLTFRVPVTVGETTI YVEQENMSLKNYGEFYRFLTDDRLMSLLNIVETLK PNENGDLVIRHTDLMSELAAYDQYRSTIFMLIQSIEN LIITNNAVLDDPDADGFVREDLPKRNNFASLLELIN</p>

		QLNVELTDDERKLL VAIRNAF SHNSYNIDF SLIKDV KHLPE VAKGILQHLQ SMLGVEITK
Riemerella anatipestifer WP_004919755 (SEQ ID Nos. 71 and 72)	GTTGGGACT GCTCTCACT TTGAAGGGT ATTCACAAC	MEKPLLPNVYTLKHKFFWGAFLNIARHNAFITICHIN EQLGLKTP SNDDKIVD VVCET WNNILNNDHDLKK S QLTELILKHFPFLTAMCYHPPKKEGKKKGHQKEQQK EKESEAQSQAELNPSKLI EALEILVNQLHSLRNYYS HYKHKKPDAEKDIFKHL YKAFDASLRMVKEDYKAH FTVNLTRDFAHLNRKGKNKQDNPDFNRYRFEKDG FTE SGLLFF TNLFLDKRD AYWMLKK VSGFKA SFIKQ REKMTTE VFCRSRILLPKLRLE SRYDFINQMLLDML S ELSRCPKLLYEKLSEENKKHFQVEADGFLDEIEEEQN PFKDTLIRHQDRFPYFALRYLDL NESFKSIRFQVDLG TYHYCIYDKKIGDEQEKRHLTRTLL SFGRLQDF TEIN RPQEWKALTKDL DYKETSNOQPFISKTTPHYHITDNKI GFRLGT SKEL YP SLEIKDGANRIAK YPYN SGF VAHAF ISVHELLPLMFYQHLT GKSEDLLKETVRHIQRIYKDF EERINTIEDLEKANQGRLPLGAFPKQMLGLLQNKQ PDL SEKAKIKIEKLI AETKLL SiTRLN TKL S SPKLGKR REKLIK TGVL ADWL VKDFMRF QPVA YDAQNQPIK S SKANSTEFWFIRRALALYGGEKNRLEGYFKQTNLIG NTNPHPFLNKFNWKACRNLVDFYQQYLEQREKFL AIKNQPWEPYQYCLLLKIPKENRKNLVKGWEQGGIS LPRGLF TEAIRETL SEDLML SKPIRKEIKKHGRVGF S RAITLYFKEKYQDKHQSFYNLSYKLEAKAPLLKREE HYEYWQNKQPQSPTESQRLELHTSDRWKDYLLYKR WQHLEKRLRYRNQDVMLWMLTLELTKNHFKELN LNYHQLKLENLAVNVQEADAKLNPLNQTLPMVLPV KVYPATAFGEVQYHKTPIRT VYIREEHTK ALKMGNF KALVKDRRLNGLF SFIKEENDTQKHPISQLRLRRELE IYQSLRVD AFKETL SLEEKLLNKHT SLSSLENEFRAL LEEWKKEYAAS SMVTDEHIAFIAS VRNAFCHNQ YPF YKEALHAPIPLFTVAQPTTEEKDGLGIAEALLKVLRE YCEIVKSI
Prevotella aurantiaca WP_025000926 (SEQ ID Nos. 73 and 74)	GTTGTATCT GCCTTCTGT TTGAAAGGT AAAAACAA C	MEDDKKTTGSISYELKDKHFWAAFLNLARHNVYITI NHINKLLEIREIDNDEKVLDIKTLWQKGNKDLNQA RLRELMTKHFPFLETAIYTKNKEDKKEVKQEKQAEA QSLESKDCFLFLDKLQEARNYYSHYKYSEFSKEPE FEEGLLEKM YNIFGNNIQL VIND YQFINKDINPDEDFK HLDRKGQFKYSFADNEGNITESGLLFFVSLFLEKKD AIWMQQKLNKGFKNLENKKKMTHEVFCRSRILMPK LRLESTQTQDWILLDMLNELIRCPKSLYERLQGDDR EKFKVPFDPADEDYNAEQEPFKNTLIRHQDRFPYFV LRYFD YNEIFKNLRFQIDLGT YHF SIYKKLIGGQKED RHLTHKLYGFERIQEFAKQNRPD EWKAIVKDLDTYE TSNKRYISETTPHYHLENQKIGIRFRNGNKEIWPSLK TNDENNEKSKYKLDKQYQAEAFLSVHELLPMMFY LLKKEKPNND EINASIVEGFIKREIRNIFKLYDAFAN GEINNIDDLEKYCADKGIPKRHLPKQMVAILYDEHK

		<p>DMVKEAKRKQKEMVKDTKLLATLEKQTQKEKED DGRNVKLLK SGEIARWL VNDMMRFQPVQKDNEGK PLNNSKANSTEYQMLQRSALYNNEEKPTRYFRQV NLIESNNPHFPLKWKWECCNNILTFYYSYLTKKIEF LNKLPEDWKKNQYFLKKEPKTNRETLVQGWKN GFNLPRGIFTEPIREWFKRHQNSKEYEKVEALDRV GLVTKVIPLFFKEEYFKDKEENFKEDTQKEINDCVQP FYNFPYNVGNHHPKEKDFLHREERIELWDKKKDKF KGYKEKIK SKKLTEKDKEEFRS YLEF QSWNKFEREL RLVRNQDIVTWLLCKELIDKLIKIDELNIEELKKLRLN NIDTDT AKKEKNNILNRVMPMELP VTVYEIDD sFKI VKDKPLHTIYKEAETKLLKQGNFKALVKDRRLNGL FSFVKTNSEAESKRNPISKLRVEYELGEYQEARIEIQ DML ALEEKLINK YKDLPTNKFSEMLNS WLEGKDE A DKARF QND VDFLI A VRN AF SHNQ YPMHNKIEF ANIK PFSLYTANNSEEKGLGIANQLKDKTKETTDKIKKIEK PIETKE</p>
<p>Prevotella saccharolytica WP_05 1522484 (SEQ ID Nos. 75 and 76)</p>	<p>GTTGTGTCT ACCTCCTTT TTGAGAGGT AAAAACAG C</p>	<p>MEDKPFWAFFNLARHNVYLTVNHINKLLDLEKLY DEGKHKEIFEREDIFNISDDVMNDANSNGKKRKLDI KKIWDDLDTDLTRKYQLRELILKHFPFIQPAIIGAQT KERTTIDKDKRSTSTSNDSLKQTGEDINDLLSLSNV KSMFFRLLQILEQLRNYYSHVKHSKSATMPNFDEDL LNWMRYIFID SVNKVKEDYSSNSVIDPNT SF SHLIYK DEQGKIKPCRYPFTSKDGSINAFGLLFFVSLFLEKQD SIWMQKKIPGFKKASENYMKMTNEVFCRNHILLPKI RLET VYDKD WMLLDMLNE VVRCPL SL YKRLTP AAQ NKFKVPEK SSDNANR QEDDNPFSRILVRHQNRFP YF VLRFFDLNEVFTTLRFQINLGCYHFAICKKQIGDKKE VHHLIRTL YGF SRLQNFTQNTRPEEWNTLVKTTEP SS GNDGKT VQGVPLPYISYTIPTPHYQIENEKIGIKIFDGD AVDTDIWPSVSTEKQLNPKDKYTLTPGFKADVFLSV HELLPMMFYQLLLCEGMLKTDAGNAVEKVLIDTR NAIFNLYDAFVQEKINTITDLENYLQDKPILIGHLPKQ MIDLLKGHQRDMLKAVEQKKAMLIKDTERRLLKLD KQLKQETDVAAKNTGTLKNGQIADWL VNDMMRF QPVKRDKEGNPINC SKANSTE YQMLQRAF AFYATD S CRLSRYFTQLHLIHSDNSHLFLSRFEYDKQPNLIAFY AAYLKAKLEFLNELQPQNWASDNYFLLLRAPKNDR QKL AEGWKNGFNLPRGLF TEKIKT WFNEFKTIVDI S DCDIFKNRVGQVARLIPVFFDKKFKDHSQPFYRYDF NVGNVSKPTEANYLSKKGKREELFKSYQNKFKNNIPA EKTKE YRE YKNF SLWKKFERELRLIKNQDILIWL KMLFDEKIKPKKDILEPRI AVS YIKLD SLQTNT STAGS LNALAKVVPMTL AIHID SPKPKGK AGNNEKENKEF T VYIKEEGTKLLKWGNFKTLLADRRIKGLFSYIEHDDI DLKQHPLTKRRVDLELDLYQTCRIDIFQQTGLEAQ LLDKYSDLNTDNFYQMLIGWRKKEGIPRNIKEDTDF LKD VRN AF SHNQ YPD SKKI AFRRIRKFNPKELILEEE</p>

		EGLGIATQMYKEVEKVVNRIKRIELFD
<p>Prevotella intermedia</p> <p>WP_061868553</p> <p>(SEQ ID Nos. 77 and 78)</p>	<p>GTTGCATCT</p> <p>GCCTGCTGT</p> <p>TTGCAAGGT</p> <p>AAAAACAA</p> <p>C</p>	<p>MEDDKKTTD SIRYELKDKHFWAAFLNLARHNVYIT</p> <p>VNHINKILEEDEINRDGYENTLENSWNEIKDINKKDR</p> <p>LSKLIKHFPFLEATTYRQNPTDTTKQKEEKQAEAQS</p> <p>LESLKKSFFVFIYKLRDLRNHYSHYKHSKSLERPKEFE</p> <p>EDLQNKMYNIFDVS IQFVKEDYKHNTDINPKKDFKH</p> <p>LDRKRKGGKFHYSFADNEG NITESGLLFFVSLFLEKKD</p> <p>AIWVQKKLEGFKC SNKSYQKMTNE VFCRSRMLLPK</p> <p>LRLESTQTQDWILLDMLNELIRCPKSLYERLQGVNR</p> <p>KKFYVSFDPADEDYDAEQEPFKNTLVRHQDRFPYFA</p> <p>LRYFDYNEVFANLRFQIDLGT YHF SIYKKLIGGQKED</p> <p>RHLTHKLYGFERIQEFDKQNRPDEWKAIVKDSDTFK</p> <p>KKEEKEEEKPYISETTPHYHLENKKIGIAFKNHNIWP</p> <p>STQTELTNNKRKKYNLGT SIKAEAFLSVHELLPMMF</p> <p>YYLLKTENTKNDNKVGGK KETKKQGKHKIEAIIES</p> <p>KIKDIYALYDAFANGEINSEDELKEYLKGKDIKIVHL</p> <p>PKQMIAILKNEHKDMAEKAEAKQEKMLATENRLK</p> <p>TLDKQLKGKIQNGKRYNSAPKSGEIASWL VNDMMR</p> <p>FQPVQKDENGE SLNNSKANS TEYQLLQRTL AFFGSE</p> <p>HERLAPYFKQTKLIES SNPHPFLNDTEWEKCSNILSF</p> <p>YRSYLKARKNFLESLKPEDWEKNQYFLMLKEPKTN</p> <p>RETLVQGWKNGFNLPRGFFTEPIRKWFMEHWKSIK</p> <p>VDDLKRVGLVAKVTPLFFSEKYKDSVQPFYNYPFNV</p> <p>GDVNKPKEDFLHREERIELWDKKKDKFKGYKAKK</p> <p>KFKEMTDKEKEEHRSYLEFQSWNKFERELRLVRNQ</p> <p>DIVTWLLCTELIDKLKIDELNIKELKKLRLKDINTDT</p> <p>AKKEKNNILNRVMPMELPVTYKVNKGGYIKNKP</p> <p>LHTIYIKEAETKLLKQGNFKALVKDRRLNGLF SFVK</p> <p>TPSEAESSENPISKL RVEYELGKYQNARLDIIEDMLA</p> <p>LEKKLIDKYNSLDTDNFHNMLTGWLELKGEAKKAR</p> <p>FQNDVKLLT AVRNAF SUNQYPMYDENLFGNIERFSL</p> <p>SSSNIIESKGLDIAAKLKEEVSKAAKKIQNEEDNKKE</p> <p>KET</p>
<p>Capnocytophaga canimorsus</p> <p>WP_013997271</p> <p>(SEQ ID Nos. 79 and 80)</p>	<p>GTTGGA ACT</p> <p>GCTCTCATT</p> <p>TTGGAGGGT</p> <p>AATCACAAC</p>	<p>MKNIQRLGKGNEF SPFKKEDKFYFGGFLNLANNNIE</p> <p>DFFKIITRF GIVITDENKKPKETF GEKILNEIFKKDI SI</p> <p>VDYEKWVNIFADYFPFTKYL SL YLEEMQFKNRVICF</p> <p>RDVMKELLKTVEALRN FYTHYDHEPIKIEDRVFYFL</p> <p>DKVLLDVSLTVKNKY LKTDKTKEFLNQHIGEELKEL</p> <p>CKQRKDYLVGKGKRIDKESEIINGIYNNAFKDFICKR</p> <p>EKQDDKENHNS VEKILCNKEPQNKKQKS SATVWEL</p> <p>CSKSSSKYTEK SFPNRENDKHCL E VPI SQKGIVFLLSF</p> <p>FLNKGEIYALTSNIKGF KAKITKEEPVTYDKNSIRYM</p> <p>ATHRMFSFLAYKGLKR KIRT SEINY NEDGQASSTYE</p> <p>KETLMLQMLDELNKVPD VVYQNLSE DVQKTFIEDW</p> <p>NEYLKENNGDVGTM EEEQVIHPVIRKRYEDKFN YFA</p> <p>IRFLDEF AQFP TLR FQVHLGN YLCDKRTKQICDTTTE</p> <p>REVKKKITVFGR LSELENKKAIFLNEREEIKGWE VFP</p> <p>NPSYDFPKENISVNYKDFPIVGSILDREKQPVS NKIGI</p>

		<p>RVKIADELQREIDKAIKEKKLRNPKNRKANQDEKQK ERLVNEIVSTNSNEQGEPVWIGQPTAYLSMNDIHSV LYEFLINKISGEALETKIVEKIETQIKQIIGKDATTKIL KPYTNANSNSINREKLLRDLEEQEQILKTLLEEQQQR EKDKKDKK SKRKHEL YP SEKGKVA VWLANDIKRF MPKAFKEQWRGYHHSLLQK YLAYYEQ SKEELKNLL PKEVFKHFPFKLKG YFQQQYLNQFYTDYLRRLSYV NELLNINQNFKNKDALKATEKECFKFFRKQNYIINP INIQISILVYPIFLKRGFLDEKPTMIDREKFKENKDTE LADWFMHYKNYKEDNYQKFYAYPLEKVEEKEKFK RNKQINKQKKNVDVYTLMMVEYIIQKIFGDKFVEENP LVLKGFQSKAERQQNNTAATTQERNLNGILNPK DIKIQGITVKG VKLKDIGNFRKYEIDQRVNTFLDYE PRKEWMA YLPNDWKEKEKQGQLPPNNVIDRQISKY ETVRSKILLKDVQELEKIISDEIKEEHRHDLKQGKYY NFKYYILNGLLRQLKNENVENYKVFKLNTNPEKVNI TQLKQEATDLEQKAFVLT YIRNKFAHNQLPKKEFW DYCQEKYGKIEKEKTYAEYFAEVFKREKEALIK</p>
<p>Porphyromonas galae WP_039434803 (SEQ ID Nos. 81 and 82)</p>	<p>GTTGGATCT ACCCTCTAT TTGAAGGGT ACACACAA C</p>	<p>MTEQSERPYNGTYTLEDKHFVA AFLNLARHNAYI TLTHIDRQLAYS KADITNDQDVL SFKALWKNFDNDL ERKSRLRSLILKHF SFLEGAAYGKKLFE SKSSGNKSS KNKELTKKEKEELQANALSLDNLKSILFDLQKLKD FRNYYSHYRHSGSSELPLFDGNMLQRLYNVFDVSV QRVKIDHEHNDEVDPHYHFNHLVRKGGKDRYGHN DNP SFKHDFVDGEGMVTEAGLLFFVSLFLEKRD AIW MQKKIRGFKGGTET YQQMTNEVFCRSRISLPKLE SLRMDDWMLLDMLNELVRCPKPLYDRLREDDRAC FRVPVDILPDEDDTDGGGEDPFKNTLV RHQDRFPYF ALRYFDLKKVFTSLRFHIDLGT YHFAYKMKMIGEPE DRHLTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYF ETGDKPYISQTSPHYHIEKGKIGLRFMPEGQHLWSP EVGTTTRTGRSKYAQDKRLTAEAFLSVHELMPMMFY YFLLREKYSEEVSAERVQGRIKRVIEDVYAVYDAFA RDEINTRDELACLADKGIRRGHLPRQMIAILSQEHK DMEEKIRKKLQEMMADTDHRLDMLDRQTRKIRIG RKNAGLPKSGVIADWLVRDMMRFQPVAKDASGKP LNNSKANSTEYRMLQRALALFGGEKERLTPYFRQM NLTGGNNPHFPHETR WESHTNILSFYRSYLARKA FLERIGRSDRVENRPFLLLKEPKTDRQTLVAGWKGE FHLPRGIFTEAVRDCLIEMGHDEVASYKEVGFMAKA VPLYFERACEDRVQPFYDSPFN VGNLSLKPCKGRFLS KEERAEW ERGKERFRDLEAWSYSAARRIEDAFAGI EYASPGNKKKIEQLLRDLSLWEAFESKLKVRADRIN LAKLKEILEAQEHYHDFKSWQKFERELRLVKNQ DIITWMMCRDLMEENKVEGLDTGTLYLKD IRPNVQ EQGSLNVLNRVKPMRLP VVVYRADSRGHVITKEEAP LATVYIEERDTKLLKQGNFKSFVKDRRLNGLF SFVD TGGLAMEQYPISKLRVEYELAKYQTARVCVFELTLR LEE SLLTRYPHLPDES FREMLESWSDPLLAKWPELH</p>

		GKVRLLIAVRNAF SHNQYPMYDE AVFSSIRKYDPSS PDAIEERMGLNIAHRLSEEVKQAKETVERIIQA
Prevotella sp. P5-125 WP_044065294 (SEQ ID Nos. 83 and 84)	GTTGTGGAA GGTCCAGTT TTGAGGGGC TATTACAAC	MNIPALVENQKKYFGTYSVMAMLNAQTVLDHIQKV ADIEGEQNNENLWFHPVMSHLYNAKNGYDKQPE KTMFIIERLQSYFPFLKIMAENQREYSNGKYKQNRV EVNSNDIFEVLKRAFGVLKMYRDLTNHYKTYEKL NDGCEFLTSTEQPLSGMINNYTVALRNMNERYGY KTEDLAFIQDKRFKFKDAYGKKKSQVNTGFFLSLQ DYNGDTQKKLHL SGVGIALLICLFLDKQYINIFL SRL PIFSSYNAQSEERRIIIRSFGINSIKLPKDRIHSEKSNKS VAMDMLNEVKRCPDELFTTLSAEKQSRFRIISDDUN EVLMKRS SDRFVPLLLQYIDYGKLFDFHIRFHVNMGK LRYLLKADKTCIDGQTRVRVIEQPLNGFGRLEEAE MRKQENGTFGNSGIRIRDFENMKRDDANPANYPYIV DTYTHYLENNKVEMFINDKED SAPLLPVIEDDRYV VKTIPSCRMS TLEIPAMAFIMFLFGSKKTEKLIVDVH NRYKRLFQAMQKEEVTAENIASFGIAESDLPQKILDL ISGNAHGKDVDAFIRLTVDDMLTDTERRIKRFKDDR KSIRSADNKMGRGFKQISTGKLADFLAKDIVLFQPS VNDGENKITGLNYRIMQSAIAVYDSGDDYEAKQF KLMFEKARLIGKGTTEPHFPLYKVFARSIPANA VEFY ERYLIERKFYLTGLSNEIKKGNRVDVPFIRRDQNKW KTPAMKTLGRIYSEDLPVELPRQMFDNEIKSULKSLP QMEGIDFNANVTYLIAEYMKRVLDDDFQTFYQWN RNYRYMDMLKGEYDRKGS LQHCFTSVEEREGLWK ERASRTERYRKQASNKIRSNRQMRNASSEEIETILDK RLSNSRNEYQKSEKVIRRYRVQDALLFLLAKKTLTE LADFDGERFKLKEIMPDAEKGILSEIMPMSFTFEKGG KKYTITSEGMLKKNYGDFVFLASDKRIGNLLELVGS DIVSKEDIMEEFNKYDQCRPEIS SIVFNLEKWAFTY PELSARVDREEKVD FKSILKILLNNKNINKEQSDILRK IRNAFDHNNYPDKGVVEIKALPEIAMS IKKAFGEYAI MK
Flavobacterium branchiophilum WP_014084666 (SEQ ID Nos. 85 and 86)	GTTGTAAC GCCCTTATT TTGAAGGGT AAACACAA C	MENLNKILDKENEICI SKIFNTKGIAAPITEKALDNIKS KQKNDLNKEARLHYF SIGHSFKQIDTKKVFDYVLIEE LKDEKPLKFITLQKDFFTKEFSIKLQKLINSIRNINNH YVTINFNDINLNKIDSNVFHFLKESFELAIIEKYYKVN KKYPLDNEIVLFLKELFIKIDENTALLNYFTNL SKDEA IEYILTFTITENKIWNINNEHNILNIEKGKYLTFEAMLF LITIFLYKNEANHLLPKLYDFKNNKSKQELFTFFSKK FTSQDIDAEEGHLIKFRDMIQYLNHYPTAWNNDLKL ESENKNKIMTTKLIDSIIEFELNSNYP SFATDIQFKKE AKAFLFASNKKRNQTSFSNKSYNEEIRHNPHIKQYR DEIASALTPISFNVKEDKFKIFVKKHVLEEFYFNSIGY EKFLEYNDFTEKEKEDFGLKLYSNPKTNKLIERIDNH KLVKSHGRNQDRFMDFSMRFLAENNYFGKDAFFKC YKFYDTQEQDEF LQSNENDDVKFHKGKVTTYIKY EEHLKNYSYWD CPFVEENNSMSVKISIGSEEKILKIQ

		<p>RNLMIYFLENALYNENVENQGYKLVNNYYRELKKD VEESIASLDLIKSNPDFKSKYKKILPKRLLHNYAPAK QDKAPENAFETLLKKADFREEQYKLLKKAHEKN KEDFVKRNKGKQFKLHFIRKACQMMYFKEYNTLK EGNAAFEKKDPVIEKRKNKEHEFGHHKNLNITREEF NDYCKWMFAFNGND SYKKYLRDLFSEKUFFDNQE YKNLFESSVNLEAFYAKTKELFKKWIETNKPTNEN RYTLENYKNLILQKQVFINVYHFSKYLIDKNLLSEN NVIQYKSLENVEYLISDFYFQSKLSIDQYKTCGKLFN KLKSNKLEDCLLYEIAAYNYIDKKNVHKIDIQKILTSKI ILTINDANTPYKISVPFNKLERYTEMIAIKNQNLKA RFLIDLPLYLSKNKIKKGKDSAGYEIIKNDLEIEDINT INNKIINDSVKFTEVLMELEKYFILKDKCILSKNYIDN SEIPSLKQFSKVWIKENENEIINYRNAACHFHLPLLETF DNLLNVEQKFIKEELQNVSTINDLSKPQEYLILLFIK FKHNNFYLNLFNKNESKTIKNDKEVKKNRVLQKFIN QVILKKK</p>
<p>Porphyromonas gingivalis WP_053444417 (SEQ ID Nos. 87 and 88)</p>	<p>GTTGGATCT ACCCTCTAT TCGAAGGGT ACACACAA C</p>	<p>MTEQNEKPYNGTYTLEDKHFWA AFLNLARHNAYI TLAHIDRQLAYSKADITNDEDILFFKGQWKNLDNDL ERKARLRSLILKHF SFLEGAAYGKKLFE SQSSGNKSS KKKELSKKEKEELQANALSLDNLKSILFDLQKDKD FRNYYSHYRHPESSELPLFDGNMLQRLYNVFDVSVQ RVKRDHEHNDKVDPHRHFNHLVRKGKKDKYGNND NPFKHHFVDREGTVTEAGLLFFVSLFLEKRDAIWM QKKIRGFKGGTEAYQQMTNEVFCRSRISLPKLESL RTDDWMLLDMLNELVRCPSLYDRLREEDRARFRV PVDILSDEDDTDGTEEDPFKNTLVRHQDRFPYFALR YFDLKKVFTSLRFHIDLGTYHFAYKKNIGEQPEDRH LTRNLYGFGRIQDFAEEHRPEEWKRLVRDLDYFETG DKPYITQTPHYHIEKGKIGLRFVPEGQHLWPSPEVG ATRTGRSKYAQDKRLTAEAFLSVHELMPMMFYFL LREKYSEEVS AEKVQGRIKRVIEDVYAVYDAFARDE INTRDELDAACLADKGIRRGHLPRQMIAILSQEHKDM EEKVRKKLQEMIADTDHRLDMLDRQTDRKIRIGRK NAGLPKSGVVADWLVRDMMRFQPAKDTSGKPLN NSKANSTEYRMLQRALALFGGEKERLTPYFRQMNL TGGNNPHPFLETRWESHTNILSFYRSYLEARKAFL QSIGRSDRVENHRFLLLKEPKTDRQTLVAGWKGEFH LPRGIFTEAVRDCLIEMGYDEVGSYKEVGFMAKAVP LYFERASKDRVQPFYDYPFNVGNSLKPCKGRFLSKE KRAEWE SGKERFRLAKLKEILEAKEHPYHDFKS WQKFERELRLVKNQDIITWMMCRDLMEENKVEGL DTGTLYLKDIRTDVQEQGSNLVLRV KPMRLPVVV YRADSRGHVHKEQAPLATVYIEERDTKLLKQGNFKS FVKDRRLNGLFSFVDTGALAMEQYPISKLRVEYELA KYQTARVCAFEQTLELEESLLTRYPHLPDKNFRKML ESWSDPLLDKWPDLHGNVRLLI AVRNAF SHNQYPM YDETLFSSIRKYDPSSPDAIEERMGLNIAHRLSEEVK</p>

		QAKEMVERIIQA
Prevotella intermedia WP_050955369 (SEQ ID Nos. 89 and 90)	GTTGCATCT GCCTGCTGT TTGCAAGGT AAAAACAA C	MEDDKKTKESTNMLDNKHFVAAFLNLARHNVYIT VNHINKVLELKNKKDQDIIDNDQDILAIKTHWEKV NGDLNKTERLRELMTKHFPFLETAIYTKNKEDKEEV KQEKQAKAQSFDSLKHCLFLFLEKLQEARNYYSHY KYSESTKEPMLKELLKKNYNIFDDNIQLVIKD YQH NKDINPDEDFKHLDRTEEEFNYYFTTNKKG NITASG LLFFVSLFLEKKDAIWMQQKLRGFKDNRESKKKMT HEVFCRSRMLLPKLRLESTQTQDWILLDMLNELIRCP KSLYERLQGEYRKKFNVPFDSADEDYDAEQEPFKNT LVRHQDRFPYFALRYFDYNEIFTNLR FQIDLGT YHFS IYKKLIGGQKEDRHLTHKLYGFERIQEFAKQNR TDE WKAIVKDFDTYETSEEPYISETAPHYHLENQKIGIRF RNDNDEIWP SLKTNGENNEKRKYKLDKQ YQAE AFL SVHELLPMMFYLLLLKKEEPNNDKKNASIVEGFIKR EIRDIYKLYDAFANGEINNIDDLEKYCEDKGIPKRHL PKQM VAILYDEITKDMAEEAKRKQKEMVKD TKLL ATLEKQTQGEIEDGGRNIRLLKSGEIARWLVNDMMR FQPVQKDNEGNPLNNSKANSTEYQMLQRSLALYNK EEKPTRYFRQVNLINSNPHPFLKWKWECCNNILSF YRSYLTKKIEFLNKLKPEDWEKNQYFLKLKEPKTNR ETLVQGWKNGFNLPRGIFTEPIREWFKRHQNDSE EY EKVETLDRVGLVTKVIPLFFKKEDSKDKEEYLK KDA QKEINNCVQPFYGFYVNGNIHKPDEKDFLP SEERK KLWGDKKYKFKGYKAKVKSKKLT DKEKEEYRSYL EFQSWNKFERELRLVRNQDIVTWLLCTELIDK LKVE GLNVEELKKLRLKDIDTDTAKQEKNNILNRVMPMQ LPVTVYEIDDSHNIVKDRPLHTVYIEETKTKLLKQGN FKALVKDRRLNGLF SFVDTS SETELKSNPISKSLVEY ELGEYQNARIETIKDMLLLEETLIEKYKTLPTDNFSD MLNGWLEGKDEADKARFQNDVKLLVA VRNAF SUN QYPMRNRIAFANINPF SLS SADT SEEKKLDIANQLKD KTHKIIKRIIEIEKPIETKE
Fusobacterium necrophorum subsp. funduliforme ATCC 51357 contig00003 (SEQ ID Nos. 91 and 92)	GACTAAATC CAAGTAGAT TGGAATTTT AAC	MEKFRRQNRNSIIKIISNYDTKGIKELKVRYRKQAQ LDTFIIKTEIVNNDIFIKSIIEKAREKYRYSFLFDGEEK YHFKNKS SVEIVKKDIF SQT PDMIRNYKITLKISEKN PRVVEAEIEDLMNSTILKDGRRSARREKSMTERKLIE EKVAKNYSLLANCPMEEVD SIKIYKIKRFLT YRSNM LLYFASINSFLCEGIKGDNETEEIWHLKDNDVRKEK VRENFKNKLIQSTENYNSSLKNQIEEKEKLLRKEFKK GAFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSKL RHSLMHYDYQYFENLFENKKNDDLMKDLNLDL FKS LPLIRKMKLNNKVNYLEDGDTL FVLQKTKKAKTLY QIYDALCEQKNGFNKFINDFVSDGEENTVFKQIINE KFQSEMEFLEKRI SESEKKNEKLLKKKLD SMKAHFRN INSEDTKEAYFWDIHS SRNYKTKYNERKNLVNE YTE

		<p>LLGSSKEK KLLREEITKINRQLLKLKQEMEEITKKNS LFRLEYKMKIAFGFLFCEFDGNISFKFDEFDASNQEK IIQYHKNGEKYLTSFLKEEEKEKFNLEKMQKIIQKTE EEDWLLPETKNNLKFYLLTYLLLPYELKGDFLGFV KKHYIDIKNVDFIDENQNNIQVSQTVEKQEDYFYHK IRLFEKNTKKYEIVKYSIVPNEKLKQYFEDLGIDIKYL TVEQKSEVSEEKNKKVSLKNNGMFNKTILLFVFKYY QIAFKLFNDIELYSLFFLREKSGKPLEIFRKELESKMK DGYLNFQGQLLYVVYEVLVKNKDLKILSKKIDYRK DKSF SPEIAYLRNFLSHLNYSKFLDNFMKINTNKSDE NKEVLPSIKIQKMIQFFIKCNLQNQIDFDNFVNDFY MRKEKMFFIQLKQIFPDINSTEKQKMNEKEEILRNRY HLTDKKNEQIKDEHEAQSQLEYKILSLQKIYSSDKNN FYGRLKEEKLLFLEKQGGKLSMEEIKDKIAGDISDL LGILKKEITRDIKDKLTEKFRYCEEKLLNLSFYNHQD KKKEESIRVFLIRDKNSDNFKFESILDDGSNKIFISK GKEITIQCCDKVLETLLIEKNTLKISSNGKIISLIPHYSY SIDVKY</p>
<p>Fusobacterium necrophorum DJ- 2 contig0065, whole genome shotgun sequence (SEQ ID Nos. 93 and 94)</p>	<p>GACTAAATC CAAGTAGAT TGGAATTTT AAC</p>	<p>MEKFRRQNRSSIIKIIISNYDTKGIKELKVR YRKQ AQL DTFIIKTEIVNNDIFIKSIIKAREKYRYSFLFDGEEKY HFKNKSSVEIVKKDIFSQTPDNMIRNYKITLKISEKNP RVVEAEIEDLMNSTILKDGRRSARREKSMTERKLEE KVAENYSLLANCPMEEVD SIKIYKIKRFLTYRSNMLL YFASINSFLCEGIKGDNETEEIWHLKDNDVRKEKV KENFKNKLIQSTENYNS SLKNQIEEKEKLLRKE SKKG AFYRTIIKKLQQERIKELSEKSLTEDCEKIIKLYSELRH PLMHYDYQYFENLFENKENSELTKNLNLDIFKSLPL VRKMKLNNKVNYLEDNDFVLFVQKTKKAKTLYQIY DALCEQKNGFNKFNDFVSDGEENTVFKQIINEKFQ SEIEFLEKRISESEKKNEKLKKKLD SMKAHFRNINSE DTKEAYFWDIHS SRNYKTKYNERKNLVNEYTELLG SSKEK KLLREEITKINRQLLKLKQEMEEITKKNSLFR LEYKMKMAFGFLFCEFDGNISRFKDEFDASNQEKIIQ YHKNGEKYLTYFLKEEEKEKFNLLKQETIQKTGEE NWLLPQNKNNLKFYLLTYLLLPYELKGDFLGFVKK HYIDIKNVDFMDENQSSKIIESKEDDFYHKIRLFEKN TKKYEIVKYSIVPDKKLKQYFKDLGIDTKYLILDQKS EVSGEKNKKVSLKNNGMFNKTILLFVFKYYQIAFKL FNDIELYSLFFLREKSGKPFVFLKELKDKMIGKQLN FGQLLYWYEVLVKNKDLSEILSERIDYRKDMCFSA EIADLRNFLSHLNYSKFLDNFMKINTNKS DENKEVLI PSIKIQKMIKFIEECNLQSQIDFDNFVNDFYMRKEK MFFIQLKQIFPDINSTEKQKMNEKEEILRNRYHLTDK KNEQIKDEHEAQSQLEYKILSLQKIYSSDKNNFYGRL KEEKLLFLEKQEKKKLSMEEIKDKIAGDISDLLGILK KEITRDIKDKLTEKFRYCEEKLLNLSFYNHQDKKKE ESIRVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITI QCCDKVLETLLIEKNTLKISSNGKIISLIPHYSYSIDVK</p>

		Y
<p>Fusobacterium necrophorum BFTR-1 contig0068</p> <p>(SEQ ID Nos. 95 and 96)</p>	<p>GACTAAATC CAAGTAGAT TGGAATTT AAC</p>	<p>MKVR YRKQAQLDTFIIKTEIVNNDIFIKSIIEKAREKY RYSFLFDGEEKYHFKNKS SVEIVKNDIF SQTPDNMIR NYKITLKISEKNPRVVEAEIEDLMNSTILKDGRRSAR REKSMTERKLIIEKVAENYSLLANCPIEEVDSIKIYKI KRFLTYRSNMLLYFASINSFLCEGIK GKDNETEEIWH LKDNDVRKEKVKENFKNKLIQSTENYNSSLKNQIEE KEKLSSKEFKKGAFYRTIIKKLQQERIKEL SEKSLTED CEKIIKLYSELRHPLMHYDYQYFENLFENKENSELTK NLNLDIFKSLPLVRKMMLNKNVNYLEDNDTLFVLQ KTKKAKTLYQIYDALCEQKNGFNKFINDDFFVSDGEE NTVFKQIINEKF QSEMEFLEKRI SESEKKNEKLKKKL DSMKAHFRNINSEDTKEAYFWDIHSSRNYKTKYNER KNLVNEYTKLLGS SKEKKLLREEITKINRQLLKLKQE MEEITKKNSLFRLE YKMKIAFGFLFCFDGNI SKFKD EFDASNQEKIIQYHKNGEKYLT SFLKEEEKEKFNLEK MQKIIQKTEEEDWLLPETKNNLKFYLLTYLLLPYEL KGDFLGFVKKHYYDIKNVDFMDENQNNIQVSQTVE KQEDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLKQYF EDLGIDIKYLTGSVESGEKWLGENLGIDIKYLTVEQK SEVSEEKNKKVSLKNNGMFNKTILLFVFKYYQIAFK LFNDIELYSLFFLREKSEKPFEVFLLEELKDKMIGKQL NFGQLLYVVYE VL VKNKDLKIL SKKIDYRKDK SF S PEIAYLRNFL SULNYSKFLDNFMKINTNK SDENKEVL IPSIKIQKMIQFIEKCNLQNQIDFDNFVNDFFYMRKEK MFFIQLKQIFPDINSTEKQKKSEKEEILRKRYHLINKK NEQIKDEHE AQSQLYEKIL SLQKIFSCDKNNFYRRLK EEKLLFLEKQGKKKISMKEIKDKIASDISDLLGILKKE ITRDIKDKLTEKFRYCEEKLLNISFYNHQDKKKEEGI RVFLIRDKNSDNFKFESILDDGSNKIFISKNGKEITIQC CDKVLETLMIEKNTLKISSNGKIISLIPHYSYSIDVKY</p>
<p>Fusobacterium necrophorum subsp. funduliforme 1_1_36S contl.14</p> <p>(SEQ ID Nos. 97 and 98)</p>	<p>GT TAAAATT CCAATCTAC TTGGATTTA GTC</p>	<p>MTEKKSIIFFKNKS SVEIVKKDIF SQTPDNMIRNYKITL KISEKNPRVVEAEIEDLMNSTILKDGRRSARREKSM TERKLIIEK VAENYSLL ANCPMEEVD SIKIYKIKRFLT YRSNMLLYFASINSFLCEGIK GKDNETEEIWHLKD NDVRKEKVKENFKNKLIQSTENYNSSLKNQIEEKEKL LRKESKKGAFYRTIIKKLQQERIKELSEKSLTEDCEKI IKLYSELRHPLMHYDYQYFENLFENKENSELTKNLN LDIFKSLPLVRKMMLNKNVNYLEDNDTLFVLQKTK KAKTLYQIYDALCEQKNGFNKFINDDFFVSDGEENTV FKQIINEKF QSEMEFLEKRI SESEKKNEKLKKKFD SM KAHFHNINSEDTKEAYFWDIHSSRNYKTKYNERKNL VNEYTELLGS SKEKKLLREEITQINRKLKLLKQEMEE ITKKNSLFRLE YKMKIAFGFLFCFDGNI SKFKDEFD ASNQEKIIQYHKNGEKYLT YFLKEEEKEKFNLEKMQ KIIQKTEEEDWLLPETKNNLKFYLLTYLLLPYELKG DFLGFVKKHYYDIKNVDFMDENQNNIQVSQTVEKQ EDYFYHKIRLFEKNTKKYEIVKYSIVPNEKLKQYFED</p>

		LGIDIKYLTGSVESGEKWLGENLGIDIKYLTVEQKSE VSEEKIKKFL
Fusobacterium perfoetens ATCC 29250 T364DRAFT_sc affold00009.9_C (SEQ ID Nos. 99 and 100)	GACTAAAA CCAAGTAA ATTGGTATT TAAAC	MGKPNRS SIIKIIISNYDNKGIKEVKVRYNKQ AQLDT FLIKSELKDGKFILYSIVDKAREKYRYSFEIDKTNINK NEILIIKKDIYSNKEDKVIRKYILSFEVSEKNDRTIVTK IKDCLETQKKEKFERENTRRLISETERKLLSEETQKT YSKIACCSPEDIDSVKIYKIKRYL AYRSNMLLFF SLIN DIFVKGVVKDNGEEVGEIWRIDSKEIDEKKTYYDLLV ENFKKRMSQEFINYKQSIENKIEKNTNKIKEIEQKLL KEK YKKEINRLKKQLIELNRENDLLEKDKIEL SDEEI REDIEKILKIYSDLRHKLMMHYNQYFENLFENKKISK EKNEVDNLTELLDLNLFYRPLVRQLKLENKTNYLE KEDKITVLGVSDSAIKYYSYNFLCEQKNGFNFIN FFSNDGEENK SFKEKINL SLEKEIEIMEKETNEKIKEI NKNELQLMKEQKELGTA YVLDIHSNDYKISHNERN KNVKLQNDIMNGNRDKNALDKINKKL VELKIKMDK ITKRNSILRLKYKLQVAYGFLMEEYKGNIKKFKDEF DISKEKIKSYKSKGEKYLEVK SEKKYITKILNSIEDIH NITWLKNQEENLKFYVLTYYILLPFEFGRDGLGFVK KHYYDIKNVEFLDENNDRLTPEQLEKMKNDSFFNKI RLFENSKKYDILKESILT SERIGKYFSLNTGAKYFE YGGREENRGIFKNIIPIFKYYQIVLKL YNDVELAMLL TLSESEKDKINKIKELVTLKEKVSPKKID YEKKYKFS VLLDCFNRINLGGKDFLASEEVKEVAKTFTNLAYLR NKICHLNYSKFIDLLTIDTNKSTTDSEGKLLINDRIR KLIKFIENNQKMNISIDYNYINDYYMKKEKFIGQR KQAKTIIDSGKKANKRNKAEELLMYRVKKNENLI YELSKKLNELTKSELFLDKKLLKDIDFTDVKIKNKS FFELKNDVKEVANIKQALQKHSELIGIYKKEVIMAI KRSIVSKLIYDEEKVLSIIYDKTNKKYEDFLEIRRER DINKFQFLIDEKKEKLGYEKIIETKEKKKVVVKIQNN SELVSEPRIKNKDKKKAKTPEEISKLGILDLTNHYCF NLKITL
Fusobacterium ulcerans ATCC 49185 cont2.38 (SEQ ID Nos. 101 and 102)	GACTAAATC CATGTAAGT GGAATTTAA A	MENKGNKKIDFDEN YNILVAQIKE YFTKEIEN YNN RIDNIIDKKELLK YSEKKEE SEKNKKLEELNKLK SQK LKILTDEEIK ADVIKIIF SDRHSLMHYE KYFENL FENKNEELAELLNLNLFKNLTLLRQMKIENKTNYL EGREEFNIIGNIAKEVLGHYNLLAEQKNGFNFIN SFFVQDGTENLEFKKLIDEHFVNAKKRLERNIKKSK KLEKELEKMEQHYQRLNCA YVWDIHTSTTYKKLYN KRKSLIEEYNKQINEIKDKEVITAINVELLRIKKEMEE ITKSNLFLRLKYKMQIAYAFLEIEFGGNIKAFKDEFD CSKMEEVQKYLKKGVK YLYYKDKAEQKNYEFPPF EEIFENKDTHTNEEWLENT SENNLKFYILT YLLPME FKGDFLGVVKKH YYDIKNVDF TDESEKELSQVQLD KMIGD SFFUKIRLFKNTKRYEIK YSILT SDEIKRYFR LLELDVPYFYEKGTDEIGIFNKNIL TIFKYYQIFRL YNDLEIHGLFNI SSDLDKILRDLK SYGNKNINFREFL Y

		VIKQNNNS STEEEYRKIWENLEAKYLRLHLLTPEKEE IKTKTKEELEKLNNEISNLRNGICHLNYKEIEEILKTEI SEKNKEATLNEKIRKVINFIKENELDKVELGFNFINDF FMKKEQFMFGQIKQVKEGNSDSITTERERKEKNNKK LKET YELNCDNLSEF YET SNNLRERANS SSLLD SAF LKKIGLYKVKNNKVNKVKDEEKRIENIKRKLKDS SDIMGMYKAEVVKKLKEKLILIFKHDEEKRIYVTVY DTSKAVPENISKEILVKRNNSKEEYFFEDNNKKYVTE YYTLEITETNELKVIPAKKLEGKEFKTEKNKENKLM LNNHYCFNVKIIY
Anaerosalibacter sp. ND 1 genome assembly Anaerosalibacter massiliensis ND 1 (SEQ ID Nos. 103 and 104)	GGACTATAC TCACTAAGG TGAGAATA AAAC	MKSGRREKAKSNKSSIVRVII SNFDDKQVKEIKVLYT KQGGIDVIKFKSTEKDEKGRMKFNFDCAYNRLEEEEE FNSFGGKQKQSFVTTNEDLTELHVTKRHKTTGEIHK DYTIQGGKYTPIKQDRTKVTVSITDNKDHFDSDNDLGD KIRLSRSLTQ YTNRILLD ADVMKNYREI VCS DSEKVD ETINIDSQEIKINRFLSYRSNMIIYYQMINNPLLHYD GEEDKGGNDSINLINEIWKYENKKNDEKEKIIERSYK SIEK SINQYILNHNTE VE SGDKEKKIDI SEERKEDLK KTFILF SRLRHVMVHYNKYF YENLYSGKNFIIYNKD KSKSRRF SELLDL NIFKELSKIKL VKNRAVSNYLDKK TTIHVLNKNINAIKLLDIYRDICETKNGFNFINNMM TISGEEDKE YKEMVTK!ff NENMNKL SIYLENFKKH S DFKTNNKKKETYNLLKQELDEQKKLRLWFNAPYVY DIHS SKKYKEL YVERKK YVDIHSKLIE AGINNDNKK KLNEINVKLCELNTEMKEMTKLNSKYRLQYKLQLA FGFILEEFNLDIDKFVSAFDKDNNTISKFMKRETY LSKSLDRRDNRFKKLIKD YKFRDTE DIF CSDRENNLV KLYILMYILLPVEIRGDFLGFVKKNYYDLKHVDFIDK RNNNDNKDTFFHDLRLFEKNV KRLEVTSYSLSDGFLG KKSREKFGKELEKFIYKNVSIALPTNIDIKEFNKSLVL PMMKNYQIIFKLLNDIEISALFLIAKKEGNEGSITFKK VIDKVRKEDMNGNINF SQVMKMALNEKVNCQIRNS IAHINMKQL YIEPLNI YINNNQNKKT I SEQMEEIDICI TKGLTGKELNKNINDYYMKKEKLVFNLKLKRNN LVSIDAQQKNMKEKSILNKYDLNYKDENLNIKEIILK VNDLNNKQKLLKETTEGE SNYKNALSKDILLNGIIR KNINFKIKEMILGIIQQNEYRYVNINIYDKIRKEDHNI DLKINNK YIEI SCYENK SNE STDERINFKIK YMDLKV KNELL VP SCYEDIYIKKKIDLEIRYIENCK VVYIDIYY KKYNINLEFDGKTLFVKFNKDVKKNNQKVNLESNYI QNIKFIVS

Certain Cas13 effectors of the invention have the following PFS nucleotide preferences:

Species	PFS 5' pos. 1	PFS 5' pos. 2	3' PFS
Bergeyella zoohelcum		Not C	

WP_002664492			
Prevotella intermedia WP_036860899		Not C	
Prevotella buccae WP_004343973		Not C	NNA, NAN
Alistipes sp. ZOR0009 WP_047447901	Not C	R	NA
Prevotella sp. MA2016 WP_036929175	Not C	Not A	
Riemerella anatipestifer WP_004919755		Not C	NNA, NAN
Prevotella aurantiaca WP_025000926		R	
Prevotella saccharolytica WP_051522484	R	G	
Prevotella intermedia WP_061868553	R	R	
Capnocytophaga canimorsus WP_013997271		Not C	NNA, NAN
Porphyromonas gulae WP_039434803		Not C	NNA, NAN
Prevotella sp. P5-125 WP_044065294	R	Not C	
Flavobacterium branchiophilum WP_014084666	U or G	R	
Porphyromonas gingivalis WP_053444417		Not C	NNA, NAN
Prevotella intermedia WP_050955369		Not C	

[00139] To improve or otherwise alter the properties of the Cas13 enzyme, modifications of amino acids are implemented. The changeable residues are identified as a subset of the conserved charged residues. These residues have >80% conservation in the alignment of Figs. 6A-6L. These can be changed to an uncharged residue (typically an alanine). One or more of the indicated residue is mutated. Amino acid residue numbering corresponds to the consensus numbering as indicated in Figs. 7A-7O (top line), and reproduced below: (SEQ ID No. 105)

[00140] MWISIKTLIHHLGVLFFCDMGNLFGHMKIXKVVXHEKRXAKXXXPXKKVX
VKRKYSGGLLLNYNENPNKNK SXENILIKKKI SFXXLKSSSKLBKTINKPDXKKXX

XXLQWFLSEIVKKNRRNGLVLSVMLSVDKRXXEKIXEKXXXLK YFXXXXXXXXLXKL
HQEKPSKKLFNLKDLKEXEEXVFLFLKXFKNEJXYXXENDXXKDIEKILXEXLRXGF
XPADKKLKXKFLIEXXWGIFSXXXKLEPYXIQEDFXEXYIEDFKKLNKXXJXKSIEN
NKIVSQKSSDSQIYEXGKNIIMSXXGXIESIIEXXSKRXXXLDKYATXXLXEKLLLDE
XLXIEQXXXNXXEXXDKLASNLKXYXLXKLYFYVKXDKKKSXXEVAKAAVSAAK
DXNKDKYQNEVWXXHEXRKEDKRDFIXXXLEIXXIXKXIXKVXXIXKXAXXEAX
EXIKXXNIGKYRXXJDLFELEEDNXLNQFXXFVNIEXXKFFXHYXPNXIKRIXXXKN
DAXAXXLKXGELXKKVEKQLKNGALSIYXIXXGKAVYYXXFAMKXLADSDYWTX
KDLEXIKISEAFLRKFAGCSFAYXSLXAXNILQPECXXDILGKGDLLXKATVNIXQX
XSEHIMYLGKLRFINDDXLLXFKEDI AKSTXKXGXGXLXKNLIQFF GGESTWDNKIF
XAA YXXXLXGXENEDFLGWALRGAIXSIRNEXFHSFKIKKHXXXFLNIXNFIKXK
LXEFEKXXXXXKEXXHXXTSYXXXLIKLFXNEXXKXXLPXXIKELKLKSSGVX
MYYSXDDLKKLLENI YFKFSLLKIXEENXEXAXF VPSFKKVYXRADGVKGFY QQXX
XTRXHA YXLKLXPFDFXEEEXEAFNARYYLLKXIYYNXILEXXXEENEXXXXFLP
KFXNNXAFREXXNFXADXIEXYKRLQINKKKGAXKXKXKXQXKVXNXYNR
KXFA YAFENIRXMXFXETPREYMQYIQSEYXIENNGKEXKKSXXENKRNKDXFXHX
EKFLQVFIKGFXYJDXRXENFXFILXPEPQNGTKEYLYEEXXAILDEXXXNXLRX
XXITXNKXLKLEFJPEXKSDIKVXPXLVEEIYDYIKKIKINKIKKDXEJAFWQDAALY
LFCEKLLDARHLSXXLRXELIKYQFXKDIKXRAXXNGNXINHSXXXNXXXVXECT
DELEIIELEXLLNDRXSNDFKDYFDDEEAXIXXXXLCRIIFYAEYLXKYXKEEDDXXX
XAEXXXFXALEPFCQSDTAREAKNDIYXDGGXNPELRVPILNRGIXQXKKIYGTEXX
LEKLFDKNXLFBIDGXBIPXFKVSEEXAIXEXXEKKEIXEXSQYKXRGELHTEWXQ
KAREIEEYXXXXXKXKFKKPNXXFEKRFIEKHGQEYKKAXXXIXEYXWLKNKV
EXNXLNELHELLIXLLGRLIGYSALFERDLQYFXNGFHYXCLNNDXEKLAXYXNJSX
VXXKNRXIXKAXLYQIFAMYXXGLPFYSKDXDXXXAXXSGXKXSXXXXSXXTAGX
GKKJKKFKKYSXYXLIXXXLXXDXSKKLDXYLAGLELFENXEEHDNXTXIRNYIA
HFNYLXXAGXXADXSLELYNXLRDRLXSYDRKLNKNAVSKSLIDILDRHGMILKFKF
KXXXKXKXNDXXXAIXKXDXARITIXEPNGVTSEXFTYKLLXXVAALEIXSLEPK
KIRHLXXXARLLYYPKXATAQSQPDQKXXXKXKKNIXKGYIERXTNQVSSNQEEY
CELVKLLLETXXLXXLAVXGVAXBIGLHISRLRRIREDAIIVGRRYRFRVEIYVPPKSN
TSKLNAADLVRID

[00141] Mutated residues based on consensus sequence using MUSCLE alignment (www.ebi.ac.uk/Tools/isa/muscle/). **Corresponding positions in Lsh indicated**

consensus	Lsh	consensus	Lsh
K28	R9	E839	K 1134
K31	E12	R885	K 1187
R44	R29	E894	E 1196
E162	E154	R895	R 1197
E184	E179	D896	D 1198
K262	R362	K942	K1254
E288	K353	R960 (HEPN)	R1278
K357	K429	H965 (HEPN)	H1283
E360	Y432	D990	S1310
K338	K405	K992	R1312
R441 (HEPN)	D558	K994	N1314
H446 (HEPN)	N563		
E471	D616		
K482	K628		
K525	E679		
K558	K71 1		
D707	D943		
R790	11067		
K81 1	K 1103		
R833	K 1128		

[00142] Casl3 proteins having any one or more of the above amino acid residues, alone, or in combination, mutated display altered specificity and/or activity and/or alternative PAM recognition.

[00143] Casl3 alignments can be used to identify conserved residues. In a non-limiting alignment of Lw2 and FSL (Figs. 8A-8C), the following residues can be identified as conserved:

M35	K198	I478	A593	R717	F825
K36	N201	E479	L597	H722	Y829
T38	Y222	K494	I601	F740	K831
K39	D253	R495	L602	F742	D837

157	1266	N498	E61 1	K768	L852
E65	F267	S501	E613	1774	F858
G66	S280	E519	D630	K778	E867
L68	1303	N524	1631	1783	A871
N84	N306	Y529	G633	L787	L875
T86	R331	V530	K641	S789	K877
E88	Y338	G534	N646	V792	Y880
1103	K389	K535	V669	Y796	Y881
N105	Y390	Y539	F676	D799	F884
E123	K391	T549	S678	F812	F888
R128	1434	D551	N695	N818	F896
R129	K435	R577	E703	P820	N901
K139	L458	E580	A707	F821	V903
L152	D459	A581	1709	V822	N915
L194	E462	F582	1713	P823	K916
N196	L463	1587	1716	S824	R918

Q920	I1075	K1243	K1341	K1466	A1550	V1684
E951	K1076	Y1244	N1342	R1509	K1553	K1685
P956	F1092	G1245	K1343	N1510	S1554	E1689
Y959	K1097	D1255	N1350	I1512	D1557	
Q964	L1099	K1261	L1352	A1513	I1558	
I969	L1104	S1263	L1355	H1514	L1559	
N994	L1107	L1267	L1356	N1516	G1563	
F1000	K1113	E1269	I1359	Y1517	F1568	
I10001	Y1114	K1274	L1360	L1529	I1612	
Q1003	E1149	I1277	R1362	L1530	L1651	
F10005	E1151	E1278	V1363	E1534	E1652	
K1007	I1153	L1289	G1364	L1536	K1655	
G1008	L1155	H1290	Y1365	R1537	H1658	
F1009	L1158	A1294	I1369	Y1543	L1659	
N1019	D1166	N1320	R1371	D1544	K1663	
L1020	L1203	K1325	D1372	R1545	T1673	

K1021	D1222	E1327	F1385	K1546	SI677	
11023	G1224	Y1328	E1391	L1547	E1678	
N1028	11228	11334	D1459	K1548	E1679	
E1070	R1236	Y1337	K1463	N1549	C1681	

[00144] One or more of the indicated residue is mutated. Amino acid residue numbering corresponds to the numbering as indicated in Figs. 8A-8C (middle line between the two orthologous aligned sequences, indicating identical residues).

[00145] Any one or more of the residues indicated in Figs. 8A-8C, which are identical between Lew2 Casl3 and Lib Casl3 are mutated for modifying the Casl3 protein activity, specificity, or functionality. Casl3 proteins having any one or more of the above amino acid residues, alone, or in combination, mutated display altered specificity and/or activity and/or alternative PAM recognition.

[00146] Casl3 achieves RNA cleavage through conserved basic residues within its two HEPN domains. Mutation of the HEPN domain, such as (e.g. alanine) substitution of predicted HEPN domain catalytic residues can be used to convert Casl3 into an inactive programmable RNA-binding protein (dCasl3, analogous to dCas9).

[00147] According to the invention, a consensus sequence can be generated from multiple Casl3 orthologs, which can assist in locating conserved amino acid residues, and motifs, including but not limited to catalytic residues and HEPN motifs in Casl3 orthologs that mediate Casl3 function. One such consensus sequence, generated from the 33 orthologs mentioned above using Geneious alignment is:

MKI SKVXXXVXKXXXGKLXKXVNERNRX AKRL SNXLBK YIXXIDKIXKKKEXXK
FXAXEITLKLNQXXXBXLXKAXXDLRKDNXYSXJKKILHNEDINXEEEXELLINDXL
EKLXKIESXKYSYQKXXXNYXMSVQEHSKKSIXRIXESAKRNKEALDKFLKEYAXL
DPRMEXLAKLRKLELYFYFKNDXIXXEEEXNVXXHKXLKENHPDFVEXXXNKEN
AELNXYAIEKXKJLKYYFPXKXAKNSNDKIFEKQELKXWIHQJENAVRILLXXGK
VXYKLQXGYLAELWKIRINEIFIKYIXVGKAVAXFALRNXXKBENDILGGKIXKKLN
GITSFYEKIKAEIILQREXAVEVAFAANXLYAXDLXXIRXSILQFFGGASNWDXFLF
FHFATXSISDKKWNAELIXXKJGLVIREKLYSNNVAMFYKDDLEKLLNXLXXFXL
RASQVPSFKKVYVRXBFPQNLLKFNDEKDDDEAYSAXYYLLKEIYYNXFLPYFSAN
NXFFFVKNLVLKANKDKFXXAFXDIREMNXGSPIEYLXXTQXNXXNEGRKKEEKE
XDFIKFLLQIFXKGFDDYLKNNXXFILKFIPEPTXIEIXXELQAWYIVGKFLNARKXN

LLGXFXSYLKLDDIELRALRNENIKYQSSNXEKEVLEXCLELIGLLSLDLNDYFBDE
 XDFAXYJGKXLD FEKKXMKDLAELXPYDQNDGEOTIVNEINIXLAKKYGTLNLLEKJ
 XDKVSEKEIKEYYELKKEIEEYXXKGEELHEEWXQXKNRVEXRDILEYXEELXGQII
 NYNXLXNKVLLYFQLGLHYLLLDILGRLVGYTGIWERDAXLYQIAAMYXNGLPEYI
 XXXKNDKYKDGQIVGXKINXFKXDKKXLYNAGLELFENXNEHKNIXIRNYIAHFNY
 LSKAESLLXYSENLRXLF SYDRKLKN AVXKSLINILLRHGMVLKFKF GTDKK SVXIR
 SXKKIXHLKSIKKLYYPEVXVSKEYCKLVKXLLKYK (SEQ ID NO. 106)

[00148] HEPN sequence motifs identified from the above orthologs were identified. Non-limiting examples of amino acid residues that can be mutated to generate catalytically dead Cas3 mutants, based on the above consensus include, in or near HEPN1, D372, R377, Q/H382, and F383 or corresponding amino acids of an ortholog, and in or near HEPN2, K893, N894, R898, N899, H903, F904, Y906, Y927, D928, K930, K932 or corresponding amino acids of an ortholog.

[00149] In another non-limiting example, a sequence alignment tool to assist generation of a consensus sequence and identification of conserved residues is the MUSCLE alignment tool (www.ebi.ac.uk/Tools/fTisa/muscle/). For **example, using MUSCLE, the following amino acid locations conserved among Cas3 orthologs can be identified in *Leptotrichia wadei* Cas3:K2; K5; V6; E301; L331; 1335; N341; G351; K352; E375; L392; L396; D403; F446; 1466; 1470; R474 (HEPN); H475; H479 (HEPN), E508; P556; L561; 1595; Y596; F600; Y669; 1673; F681; L685; Y761; L676; L779; Y782; L836; D847; Y863; L869; 1872; K879; 1933; L954; 1958; R961; Y965; E970; R971; D972; R1046 (HEPN), H1051 (HEPN), Y1075; D1076; K1078; K1080; 11083; 11090.**

[00150] Figs. 6A-K show an alignment of Cas3 orthologs. Fig. 6L shows an exemplary sequence alignment of HEPN domains and highly conserved residues.

[00151] Cas3 HEPN may also target DNA, or potentially DNA and/or RNA. On the basis that that the HEPN domains of Cas3 are at least capable of binding to and, in their wild-type form, cutting RNA, then it is preferred that the Cas3 effector protein has RNase function. It may also, or alternatively, have DNase function.

[00152] Thus, in some embodiments, the effector protein may be a RNA-binding protein, such as a dead-Cas type effector protein, which may be optionally functionalized as described herein for instance with an transcriptional activator or repressor domain, NLS or other functional domain. In some embodiments, the effector protein may be a RNA-binding protein that cleaves a single strand of RNA. If the RNA bound is ssRNA, then the ssRNA is fully cleaved. In some embodiments, the effector protein may be a RNA-binding protein that

cleaves a double strand of RNA, for example if it comprises two RNase domains. If the RNA bound is dsRNA, then the dsRNA is fully cleaved.

[00153] RNase function in CRISPR systems is known, for example mRNA targeting has been reported for certain type III CRISPR-Cas systems (Hale et al., 2014, *Genes Dev*, vol. 28, 2432-2443; Hale et al., 2009, *Cell*, vol. 139, 945-956; Peng et al., 2015, *Nucleic acids research*, vol. 43, 406-417) and provides significant advantages. In the *Staphylococcus epidermis* type III-A system, transcription across targets results in cleavage of the target DNA and its transcripts, mediated by independent active sites within the Cas10-Csm ribonucleoprotein effector complex (see, Samai et al., 2015, *Cell*, vol. 151, 1164-1174). A CRISPR-Cas system, composition or method targeting RNA via the present effector proteins is thus provided.

[00154] The target RNA, i.e. the RNA of interest, is the RNA to be targeted by the present invention leading to the recruitment to, and the binding of the effector protein at, the target site of interest on the target RNA. The target RNA may be any suitable form of RNA. This may include, in some embodiments, mRNA. In other embodiments, the target RNA may include tRNA or rRNA. In other embodiments, the target RNA may include miRNA. In other embodiments, the target RNA may include siRNA.

Interfering RNA (RNAi) and microRNA (miRNA)

[00155] In other embodiments, the target RNA may include interfering RNA, i.e. RNA involved in an RNA interference pathway, such as shRNA, siRNA and so forth, both in eukaryotes and prokaryotes. In other embodiments, the target RNA may include microRNA (miRNA). Control over interfering RNA or miRNA may help reduce off-target effects (OTE) seen with those approaches by reducing the longevity of the interfering RNA or miRNA *in vivo* or *in vitro*.

[00156] In certain embodiments, the target is not the miRNA itself, but the miRNA binding site of a miRNA target.

[00157] In certain embodiments, miRNAs may be sequestered (such as including subcellularly relocated). In certain embodiments, miRNAs may be cut, such as without limitation at hairpins.

[00158] In certain embodiments, miRNA processing (such as including turnover) is increased or decreased.

[00159] If the effector protein and suitable guide are selectively expressed (for example spatially or temporally under the control of a suitable promoter, for example a tissue- or cell cycle-specific promoter and/or enhancer) then this could be used to 'protect' the cells or

systems *in vivo* or *in vitro*) from RNAi in those cells. This may be useful in neighbouring tissues or cells where RNAi is not required or for the purposes of comparison of the cells or tissues where the effector protein and suitable guide are and are not expressed (i.e. where the RNAi is not controlled and where it is, respectively). The effector protein may be used to control or bind to molecules comprising or consisting of RNA, such as ribozymes, ribosomes or riboswitches. In embodiments of the invention, the RNA guide can recruit the effector protein to these molecules so that the effector protein is able to bind to them.

[00160] The protein system of the invention can be applied in areas of RNAi technologies, without undue experimentation, from this disclosure, including therapeutic, assay and other applications (see, e.g., Guidi et al., PLoS Negl Trop Dis 9(5): e0003801. doi:10.1371/journal.pntd; Crotty et al., In vivo RNAi screens: concepts and applications. Shane Crotty ... 2015 Elsevier Ltd. Published by Elsevier Inc., Pesticide Biochemistry and Physiology (Impact Factor: 2.01). 01/2015; 120. DOI: 10.1016/j.pestbp.2015.01.002 and Makkonen et al., Viruses 2015, 7(4), 2099-2125; doi:10.3390/v7042099), because the present application provides the foundation for informed engineering of the system.

Ribosomal RNA (rRNA)

[00161] For example, azalide antibiotics such as azithromycin, are well known. They target and disrupt the 50S ribosomal subunit. The present effector protein, together with a suitable guide RNA to target the 50S ribosomal subunit, may be, in some embodiments, recruited to and bind to the 50S ribosomal subunit. Thus, the present effector protein in concert with a suitable guide directed at a ribosomal (especially the 50s ribosomal subunit) target is provided. Use of this use effector protein in concert with the suitable guide directed at the ribosomal (especially the 50s ribosomal subunit) target may include antibiotic use. In particular, the antibiotic use is analogous to the action of azalide antibiotics, such as azithromycin. In some embodiments, prokaryotic ribosomal subunits, such as the 70S subunit in prokaryotes, the 50S subunit mentioned above, the 30S subunit, as well as the 16S and 5S subunits may be targeted. In other embodiments, eukaryotic ribosomal subunits, such as the 80S subunit in eukaryotes, the 60S subunit, the 40S subunit, as well as the 28S, 18S, 5.8S and 5S subunits may be targeted.

[00162] In some embodiments, the effector protein may be a RNA-binding protein, optionally functionalized, as described herein. In some embodiments, the effector protein may be a RNA-binding protein that cleaves a single strand of RNA. In either case, but particularly where the RNA-binding protein cleaves a single strand of RNA, then ribosomal

function may be modulated and, in particular, reduced or destroyed. This may apply to any ribosomal RNA and any ribosomal subunit and the sequences of rRNA are well known.

[00163] Control of ribosomal activity is thus envisaged through use of the present effector protein in concert with a suitable guide to the ribosomal target. This may be through cleavage of, or binding to, the ribosome. In particular, reduction of ribosomal activity is envisaged. This may be useful in assaying ribosomal function *in vivo* or *in vitro*, but also as a means of controlling therapies based on ribosomal activity, *in vivo* or *in vitro*. Furthermore, control (i.e. reduction) of protein synthesis in an *in vivo* or *in vitro* system is envisaged, such control including antibiotic and research and diagnostic use.

Riboswitches

[00164] A riboswitch (also known as an aptozyme) is a regulatory segment of a messenger RNA molecule that binds a small molecule. This typically results in a change in production of the proteins encoded by the mRNA. Thus, control of riboswitch activity is thus envisaged through use of the present effector protein in concert with a suitable guide to the riboswitch target. This may be through cleavage of, or binding to, the riboswitch. In particular, reduction of riboswitch activity is envisaged. This may be useful in assaying riboswitch function *in vivo* or *in vitro*, but also as a means of controlling therapies based on riboswitch activity, *in vivo* or *in vitro*. Furthermore, control (i.e. reduction) of protein synthesis in an *in vivo* or *in vitro* system is envisaged. This control, as for rRNA may include antibiotic and research and diagnostic use.

Ribozymes

[00165] Ribozymes are RNA molecules having catalytic properties, analogous to enzymes (which are of course proteins). As ribozymes, both naturally occurring and engineered, comprise or consist of RNA, they may also be targeted by the present RNA-binding effector protein. In some embodiments, the effector protein may be a RNA-binding protein cleaves the ribozyme to thereby disable it. Control of ribozymal activity is thus envisaged through use of the present effector protein in concert with a suitable guide to the ribozymal target. This may be through cleavage of, or binding to, the ribozyme. In particular, reduction of ribozymal activity is envisaged. This may be useful in assaying ribozymal function *in vivo* or *in vitro*, but also as a means of controlling therapies based on ribozymal activity, *in vivo* or *in vitro*.

Gene expression, including RNA processing

[00166] The effector protein may also be used, together with a suitable guide, to target gene expression, including via control of RNA processing. The control of RNA processing

may include RNA processing reactions such as RNA splicing, including alternative splicing, via targeting of RNAPol; viral replication (in particular of satellite viruses, bacteriophages and retroviruses, such as HBV, HBC and HIV and others listed herein) including viroids in plants; and tRNA biosynthesis. The effector protein and suitable guide may also be used to control RNA activation (RNAa). RNAa leads to the promotion of gene expression, so control of gene expression may be achieved that way through disruption or reduction of RNAa and thus less promotion of gene expression. This is discussed more in detail below.

RNAi Screens

[00167] Identifying gene products whose knockdown is associated with phenotypic changes, biological pathways can be interrogated and the constituent parts identified, via RNAi screens. Control may also be exerted over or during these screens by use of the effector protein and suitable guide to remove or reduce the activity of the RNAi in the screen and thus reinstate the activity of the (previously interfered with) gene product (by removing or reducing the interference/repression).

[00168] Satellite RNAs (satRNAs) and satellite viruses may also be treated.

[00169] Control herein with reference to RNase activity generally means reduction, negative disruption or known-down or knock out.

***In vivo* RNA applications**

Inhibition of gene expression

[00170] The target-specific RNAses provided herein allow for very specific cutting of a target RNA. The interference at RNA level allows for modulation both spatially and temporally and in a non-invasive way, as the genome is not modified.

[00171] A number of diseases have been demonstrated to be treatable by mRNA targeting. While most of these studies relate to administration of siRNA, it is clear that the RNA targeting effector proteins provided herein can be applied in a similar way.

[00172] Examples of mRNA targets (and corresponding disease treatments) are VEGF, VEGF-R1 and RTP801 (in the treatment of AMD and/or DME), Caspase 2 (in the treatment of Naion)ADRB2 (in the treatment of intraocular pressure), TRPVI (in the treatment of Dry eye syndrome, Syk kinase (in the treatment of asthma), Apo B (in the treatment of hypercholesterolemia or hypobetalipoproteinemia), PLK1, KSP and VEGF (in the treatment of solid tumors), Ber-Abl (in the treatment of CML)(Burnett and Rossi Chem Biol. 2012, 19(1): 60-71)). Similarly, RNA targeting has been demonstrated to be effective in the treatment of RNA-virus mediated diseases such as HIV (targeting of HIV Tet and Rev), RSV

(targeting of RSV nucleocapsid) and HCV (targeting of miR-122) (Burnett and Rossi Chem Biol. 2012, 19(1): 60-71).

[00173] It is further envisaged that the RNA targeting effector protein of the invention can be used for mutation specific or allele specific knockdown. Guide RNA's can be designed that specifically target a sequence in the transcribed mRNA comprising a mutation or an allele-specific sequence. Such specific knockdown is particularly suitable for therapeutic applications relating to disorders associated with mutated or allele-specific gene products. For example, most cases of familial hypobetalipoproteinemia (FHBL) are caused by mutations in the ApoB gene. This gene encodes two versions of the apolipoprotein B protein: a short version (ApoB-48) and a longer version (ApoB-100). Several ApoB gene mutations that lead to FFIBL cause both versions of ApoB to be abnormally short. Specifically targeting and knockdown of mutated ApoB mRNA transcripts with an RNA targeting effector protein of the invention may be beneficial in treatment of FFIBL. As another example, Huntington's disease (HD) is caused by an expansion of CAG triplet repeats in the gene coding for the Huntingtin protein, which results in an abnormal protein. Specifically targeting and knockdown of mutated or allele-specific mRNA transcripts encoding the Huntingtin protein with an RNA targeting effector protein of the invention may be beneficial in treatment of HD.

[00174] It is noted that in this context, and more generally for the various applications as described herein, the use of a split version of the RNA targeting effector protein can be envisaged. Indeed, this may not only allow increased specificity but may also be advantageous for delivery. The Cas13 is split in the sense that the two parts of the Cas13 enzyme substantially comprise a functioning Cas13. Ideally, the split should always be so that the catalytic domain(s) are unaffected. That Cas13 may function as a nuclease or it may be a dead-Cas13 which is essentially an RNA-binding protein with very little or no catalytic activity, due to typically mutation(s) in its catalytic domains.

[00175] Each half of the split Cas13 may be fused to a dimerization partner. By means of example, and without limitation, employing rapamycin sensitive dimerization domains, allows to generate a chemically inducible split Cas13 for temporal control of Cas13 activity. Cas13 can thus be rendered chemically inducible by being split into two fragments and that rapamycin-sensitive dimerization domains may be used for controlled reassembly of the Cas13. The two parts of the split Cas13 can be thought of as the N' terminal part and the C terminal part of the split Cas13. The fusion is typically at the split point of the Cas13. In other

words, the C terminal of the N' terminal part of the split Cas13 is fused to one of the dimer halves, whilst the N' terminal of the C terminal part is fused to the other dimer half.

[00176] The Cas13 does not have to be split in the sense that the break is newly created. The split point is typically designed in silico and cloned into the constructs. Together, the two parts of the split Cas13, the N' terminal and C terminal parts, form a full Cas13, comprising preferably at least 70% or more of the wildtype amino acids (or nucleotides encoding them), preferably at least 80% or more, preferably at least 90% or more, preferably at least 95% or more, and most preferably at least 99% or more of the wildtype amino acids (or nucleotides encoding them). Some trimming may be possible, and mutants are envisaged. Non-functional domains may be removed entirely. What is important is that the two parts may be brought together and that the desired Cas13 function is restored or reconstituted. The dimer may be a homodimer or a heterodimer.

[00177] In certain embodiments, the Cas13 effector as described herein may be used for mutation-specific, or allele-specific targeting, such as . for mutation-specific, or allele-specific knockdown.

[00178] The RNA targeting effector protein can moreover be fused to another functional RNase domain, such as a non-specific RNase or Argonaute 2, which acts in synergy to increase the RNase activity or to ensure further degradation of the message.

[00179] *Modulation of gene expression through modulation of RNA function*

[00180] Apart from a direct effect on gene expression through cleavage of the mRNA, RNA targeting can also be used to impact specific aspects of the RNA processing within the cell, which may allow a more subtle modulation of gene expression. Generally, modulation can for instance be mediated by interfering with binding of proteins to the RNA, such as for instance blocking binding of proteins, or recruiting RNA binding proteins. Indeed, modulations can be ensured at different levels such as splicing, transport, localization, translation and turnover of the mRNA. Similarly in the context of therapy, it can be envisaged to address (pathogenic) malfunctioning at each of these levels by using RNA-specific targeting molecules. In these embodiments it is in many cases preferred that the RNA targeting protein is a "dead" Cas13 that has lost the ability to cut the RNA target but maintains its ability to bind thereto, such as the mutated forms of Cas13 described herein.

[00181] A) Alternative Splicing

[00182] Many of the human genes express multiple mRNAs as a result of alternative splicing. Different diseases have been shown to be linked to aberrant splicing leading to loss of function or gain of function of the expressed gene. While some of these diseases are

caused by mutations that cause splicing defects, a number of these are not. One therapeutic option is to target the splicing mechanism directly. The RNA targeting effector proteins described herein can for instance be used to block or promote splicing, include or exclude exons and influence the expression of specific isoforms and/or stimulate the expression of alternative protein products. Such applications are described in more detail below.

[00183] A RNA targeting effector protein binding to a target RNA can sterically block access of splicing factors to the RNA sequence. The RNA targeting effector protein targeted to a splice site may block splicing at the site, optionally redirecting splicing to an adjacent site. For instance a RNA targeting effector protein binding to the 5' splice site binding can block the recruitment of the U1 component of the spliceosome, favoring the skipping of that exon. Alternatively, a RNA targeting effector protein targeted to a splicing enhancer or silencer can prevent binding of transacting regulatory splicing factors at the target site and effectively block or promote splicing. Exon exclusion can further be achieved by recruitment of ILF2/3 to precursor mRNA near an exon by an RNA targeting effector protein as described herein. As yet another example, a glycine rich domain can be attached for recruitment of hnRNP A1 and exon exclusion (Del Gatto-Konczak et al. Mol Cell Biol. 1999 Jan; 19(1):25 1-60).

[00184] In certain embodiments, through appropriate selection of gRNA, specific splice variants may be targeted, while other splice variants will not be targeted.

[00185] In some cases the RNA targeting effector protein can be used to promote splicing (e.g. where splicing is defective). For instance a RNA targeting effector protein can be associated with an effector capable of stabilizing a splicing regulatory stem-loop in order to further splicing. The RNA targeting effector protein can be linked to a consensus binding site sequence for a specific splicing factor in order to recruit the protein to the target DNA.

[00186] Examples of diseases which have been associated with aberrant splicing include, but are not limited to Paraneoplastic Opsoclonus Myoclonus Ataxia (or POMA), resulting from a loss of Nova proteins which regulate splicing of proteins that function in the synapse, and Cystic Fibrosis, which is caused by defective splicing of a cystic fibrosis transmembrane conductance regulator, resulting in the production of nonfunctional chloride channels. In other diseases aberrant RNA splicing results in gain-of-function. This is the case for instance in myotonic dystrophy which is caused by a CUG triplet-repeat expansion (from 50 to >1500 repeats) in the 3'UTR of an mRNA, causing splicing defects.

[00187] The RNA targeting effector protein can be used to include an exon by recruiting a splicing factor (such as UI) to a 5' splicing site to promote excision of introns around a desired exon. Such recruitment could be mediated through a fusion with an arginine/serine rich domain, which functions as splicing activator (Gravely BR and Maniatis T, Mol Cell. 1998 (5):765-71).

[00188] It is envisaged that the RNA targeting effector protein can be used to block the splicing machinery at a desired locus, resulting in preventing exon recognition and the expression of a different protein product. An example of a disorder that may be treated is Duchenne muscular dystrophy (DMD), which is caused by mutations in the gene encoding for the dystrophin protein. Almost all DMD mutations lead to frameshifts, resulting in impaired dystrophin translation. The RNA targeting effector protein can be paired with splice junctions or exonic splicing enhancers (ESEs) thereby preventing exon recognition, resulting in the translation of a partially functional protein. This converts the lethal Duchenne phenotype into the less severe Becker phenotype.

[00189] B) RNA Modification

[00190] RNA editing is a natural process whereby the diversity of gene products of a given sequence is increased by minor modification in the RNA. Typically, the modification involves the conversion of adenosine (A) to inosine (I), resulting in an RNA sequence which is different from that encoded by the genome. RNA modification is generally ensured by the ADAR enzyme, whereby the pre-RNA target forms an imperfect duplex RNA by base-pairing between the exon that contains the adenosine to be edited and an intronic non-coding element. A classic example of A-I editing is the glutamate receptor GluR-B mRNA, whereby the change results in modified conductance properties of the channel (Higuchi M, et al. Cell. 1993;75:1361-70).

[00191] According to the invention, enzymatic approaches are used to induce transitions (A \leftrightarrow G or C \leftrightarrow U changes) or transversions (any purine to any pyrimidine or vice versa) in the RNA bases of a given transcript. Transitions can be directly induced by using adening (ADAR1/2), APOBEC) or cytosine deaminases (AID) which convert A to I or C to U, respectively. Transversions can be indirectly induced by localizing reactive oxygen species damage to the bases of interest, which causes chemical modifications to be added to the affected bases, such as the conversion of guanine to oxo-guanine. An oxo-guanine is recognized as a T and will thus base pair with an adenine causing translation to be affected. Proteins that can be recruited for ROS-mediated base damage include APEX and mini-SOG.

With both approaches, these effectors can be fused to a catalytically inactive Cas13 and be recruited to sites on transcripts where these types of mutations are desired.

[00192] In humans, a heterozygous functional-null mutation in the *ADARI* gene leads to a skin disease, human pigmentary genodermatosis (Miyamura Y, et al. *Am J Hum Genet.* 2003;73:693-9). It is envisaged that the RNA targeting effector proteins of the present invention can be used to correct malfunctioning RNA modification.

[00193] It is further envisaged that RNA adenosine methylase (N(6)-methyladenosine) can be fused to the RNA targeting effector proteins of the invention and targeted to a transcript of interest. This methylase causes reversible methylation, has regulatory roles and may affect gene expression and cell fate decisions by modulating multiple RNA-related cellular pathways (Fu et al *Nat Rev Genet.* 2014;15(5):293-306).

[00194] Q_Polyadenylation

[00195] Polyadenylation of an mRNA is important for nuclear transport, translation efficiency and stability of the mRNA, and all of these, as well as the process of polyadenylation, depend on specific RBPs. Most eukaryotic mRNAs receive a 3' poly(A) tail of about 200 nucleotides after transcription. Polyadenylation involves different RNA-binding protein complexes which stimulate the activity of a poly(A)polymerase (Minvielle-Sebastia L et al. *Curr Opin Cell Biol.* 1999;11:352-7). It is envisaged that the RNA-targeting effector proteins provided herein can be used to interfere with or promote the interaction between the RNA-binding proteins and RNA.

[00196] Examples of diseases which have been linked to defective proteins involved in polyadenylation are oculopharyngeal muscular dystrophy (OPMD) (Brais B, et al. *Nat Genet.* 1998;18:164-7).

D) RNA export

[00197] After pre-mRNA processing, the mRNA is exported from the nucleus to the cytoplasm. This is ensured by a cellular mechanism which involves the generation of a carrier complex, which is then translocated through the nuclear pore and releases the mRNA in the cytoplasm, with subsequent recycling of the carrier.

[00198] Overexpression of proteins (such as TAP) which play a role in the export of RNA has been found to increase export of transcripts that are otherwise inefficiently exported in *Xenopus* (Katahira J, et al. *EMBO J.* 1999;18:2593-609).

[00199] E) mRNA Localization

[00200] mRNA localization ensures spatially regulated protein production. Localization of transcripts to a specific region of the cell can be ensured by localization elements. In

particular embodiments, it is envisaged that the effector proteins described herein can be used to target localization elements to the RNA of interest. The effector proteins can be designed to bind the target transcript and shuttle them to a location in the cell determined by its peptide signal tag. More particularly for instance, a RNA targeting effector protein fused to one or more nuclear localization signal (NLS) and/or one or more nuclear export signal (NES) can be used to alter RNA localization.

[00201] Further examples of localization signals include the zipcode binding protein (ZBP1) which ensures localization of β -actin to the cytoplasm in several asymmetric cell types, KDEL retention sequence (localization to endoplasmic reticulum), nuclear export signal (localization to cytoplasm), mitochondrial targeting signal (localization to mitochondria), peroxisomal targeting signal (localization to peroxisome) and m6A marking/YTFIDF2 (localization to p-bodies). Other approaches that are envisaged are fusion of the RNA targeting effector protein with proteins of known localization (for instance membrane, synapse).

[00202] Alternatively, the effector protein according to the invention may for instance be used in localization-dependent knockdown. By fusing the effector protein to a appropriate localization signal, the effector is targeted to a particular cellular compartment. Only target RNAs residing in this compartment will effectively be targeted, whereas otherwise identical targets, but residing in a different cellular compartment will not be targeted, such that a localization dependent knockdown can be established.

[00203] F) Translation

[00204] The RNA targeting effector proteins described herein can be used to enhance or repress translation. It is envisaged that upregulating translation is a very robust way to control cellular circuits. Further, for functional studies a protein translation screen can be favorable over transcriptional upregulation screens, which have the shortcoming that upregulation of transcript does not translate into increased protein production.

[00205] It is envisaged that the RNA targeting effector proteins described herein can be used to bring translation initiation factors, such as EIF4G in the vicinity of the 5' untranslated repeat (5'UTR) of a messenger RNA of interest to drive translation (as described in De Gregorio et al. EMBO J. 1999;18(17):4865-74 for a non-reprogrammable RNA binding protein). As another example GLD2, a cytoplasmic poly(A) polymerase, can be recruited to the target mRNA by an RNA targeting effector protein. This would allow for directed polyadenylation of the target mRNA thereby stimulating translation.

[00206] Similarly, the RNA targeting effector proteins envisaged herein can be used to block translational repressors of mRNA, such as ZBP1 (Huttelmaier S, et al. Nature. 2005;438:512-5). By binding to translation initiation site of a target RNA, translation can be directly affected.

[00207] In addition, fusing the RNA targeting effector proteins to a protein that stabilizes mRNAs, e.g. by preventing degradation thereof such as RNase inhibitors, it is possible to increase protein production from the transcripts of interest.

[00208] It is envisaged that the RNA targeting effector proteins described herein can be used to repress translation by binding in the 5'UTR regions of a RNA transcript and preventing the ribosome from forming and beginning translation.

[00209] Further, the RNA targeting effector protein can be used to recruit Caf1, a component of the CCR4-NOT deadenylase complex, to the target mRNA, resulting in deadenylation of the target transcript and inhibition of protein translation.

[00210] For instance, the RNA targeting effector protein of the invention can be used to increase or decrease translation of therapeutically relevant proteins. Examples of therapeutic applications wherein the RNA targeting effector protein can be used to downregulate or upregulate translation are in amyotrophic lateral sclerosis (ALS) and cardiovascular disorders. Reduced levels of the glial glutamate transporter EAAT2 have been reported in ALS motor cortex and spinal cord, as well as multiple abnormal EAAT2 mRNA transcripts in ALS brain tissue. Loss of the EAAT2 protein and function thought to be the main cause of excitotoxicity in ALS. Restoration of EAAT2 protein levels and function may provide therapeutic benefit. Hence, the RNA targeting effector protein can be beneficially used to upregulate the expression of EAAT2 protein, e.g. by blocking translational repressors or stabilizing mRNA as described above. Apolipoprotein A1 is the major protein component of high density lipoprotein (HDL) and ApoA1 and HDL are generally considered as atheroprotective. It is envisaged that the RNA targeting effector protein can be beneficially used to upregulate the expression of ApoA1, e.g. by blocking translational repressors or stabilizing mRNA as described above.

[00211] G) mRNA Turnover

[00212] Translation is tightly coupled to mRNA turnover and regulated mRNA stability. Specific proteins have been described to be involved in the stability of transcripts (such as the ELAV/Hu proteins in neurons, Keene JD, 1999, Proc Natl Acad Sci U S A. 96:5-7) and tristetraprolin (TTP). These proteins stabilize target mRNAs by protecting the messages from degradation in the cytoplasm (Peng SS et al., 1988, EMBO J. 17:3461-70).

[00213] It can be envisaged that the RNA-targeting effector proteins of the present invention can be used to interfere with or to promote the activity of proteins acting to stabilize mRNA transcripts, such that mRNA turnover is affected. For instance, recruitment of human TTP to the target RNA using the RNA targeting effector protein would allow for adenylate-uridylylate-rich element (AU-rich element) mediated translational repression and target degradation. AU-rich elements are found in the 3' UTR of many mRNAs that code for proto-oncogenes, nuclear transcription factors, and cytokines and promote RNA stability. As another example, the RNA targeting effector protein can be fused to HuR, another mRNA stabilization protein (Hinman MN and Lou H, Cell Mol Life Sci 2008;65:3168-81), and recruit it to a target transcript to prolong its lifetime or stabilize short-lived mRNA.

[00214] It is further envisaged that the RNA-targeting effector proteins described herein can be used to promote degradation of target transcripts. For instance, m6A methyltransferase can be recruited to the target transcript to localize the transcript to P-bodies leading to degradation of the target.

[00215] As yet another example, an RNA targeting effector protein as described herein can be fused to the non-specific endonuclease domain PiLT N-terminus (PIN), to recruit it to a target transcript and allow degradation thereof.

[00216] Patients with paraneoplastic neurological disorder (PND)- associated encephalomyelitis and neuropathy are patients who develop autoantibodies against Hu-proteins in tumors outside of the central nervous system (Szabo A et al. 1991, Cell.;67:325-33 which then cross the blood-brain barrier. It can be envisaged that the RNA-targeting effector proteins of the present invention can be used to interfere with the binding of auto-antibodies to mRNA transcripts.

[00217] Patients with dystrophy type 1 (DM1), caused by the expansion of (CUG)_n in the 3' UTR of dystrophin myotonia-protein kinase (DMPK) gene, are characterized by the accumulation of such transcripts in the nucleus. It is envisaged that the RNA targeting effector proteins of the invention fused with an endonuclease targeted to the (CUG)_n repeats could inhibit such accumulation of aberrant transcripts.

[00218] H) Interaction with Multi-Functional Proteins

[00219] Some RNA-binding proteins bind to multiple sites on numerous RNAs to function in diverse processes. For instance, the hnRNP A1 protein has been found to bind exonic splicing silencer sequences, antagonizing the splicing factors, associate with telomere ends (thereby stimulating telomere activity) and bind miRNA to facilitate Drosha-mediated processing thereby affecting maturation. It is envisaged that the RNA-binding effector

proteins of the present invention can interfere with the binding of RNA-binding proteins at one or more locations.

[00220] I) RNA folding

[00221] RNA adopts a defined structure in order to perform its biological activities. Transitions in conformation among alternative tertiary structures are critical to most RNA-mediated processes. However, RNA folding can be associated with several problems. For instance, RNA may have a tendency to fold into, and be upheld in, improper alternative conformations and/or the correct tertiary structure may not be sufficiently thermodynamically favored over alternative structures. The RNA targeting effector protein, in particular a cleavage-deficient or dead RNA targeting protein, of the invention may be used to direct folding of (m)RNA and/or capture the correct tertiary structure thereof.

Use of RNA-targeting effector protein in modulating cellular status

[00222] In certain embodiments Cas13 in a complex with crRNA is activated upon binding to target RNA and subsequently cleaves any nearby ssRNA targets (i.e. "collateral" or "bystander" effects). Cas13, once primed by the cognate target, can cleave other (non-complementary) RNA molecules. Such promiscuous RNA cleavage could potentially cause cellular toxicity, or otherwise affect cellular physiology or cell status.

[00223] Accordingly, in certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell dormancy. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell cycle arrest. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in reduction of cell growth and/or cell proliferation. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell anergy. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell apoptosis. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell necrosis. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell death. In certain embodiments, the non-naturally occurring

or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of programmed cell death.

[00224] In certain embodiments, the invention relates to a method for induction of cell dormancy comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell cycle arrest comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for reduction of cell growth and/or cell proliferation comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell anergy comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell apoptosis comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell necrosis comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell death comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of programmed cell death comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein.

[00225] The methods and uses as described herein may be therapeutic or prophylactic and may target particular cells, cell (sub)populations, or cell/tissue types. In particular, the methods and uses as described herein may be therapeutic or prophylactic and may target particular cells, cell (sub)populations, or cell/tissue types expressing one or more target sequences, such as one or more particular target RNA (e.g. ss RNA). Without limitation, target cells may for instance be cancer cells expressing a particular transcript, e.g. neurons of a given class, (immune) cells causing e.g. autoimmunity, or cells infected by a specific (e.g. viral) pathogen, etc.

[00226] Accordingly, in certain embodiments, the invention relates to a method for treating a pathological condition characterized by the presence of undesirable cells (host

cells), comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating a pathological condition characterized by the presence of undesirable cells (host cells). In certain embodiments, the invention relates the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating a pathological condition characterized by the presence of undesirable cells (host cells). It is to be understood that preferably the CRISPR-Cas system targets a target specific for the undesirable cells. In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating cancer. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating cancer. In certain embodiments, the invention relates to a method for treating, preventing, or alleviating cancer comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cancer cells. In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating infection of cells by a pathogen. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating infection of cells by a pathogen. In certain embodiments, the invention relates to a method for treating, preventing, or alleviating infection of cells by a pathogen comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cells infected by the pathogen (e.g. a pathogen derived target). In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating an autoimmune disorder. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating an autoimmune disorder. In certain embodiments, the invention relates to a method for treating, preventing,

or alleviating an autoimmune disorder comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cells responsible for the autoimmune disorder (e.g. specific immune cells).

Use of RNA-targeting effector protein in RNA detection or protein detection

[00227] It is further envisaged that the RNA targeting effector protein can be used for detection of nucleic acids or proteins in a biological sample. The samples can be cellular or cell-free.

[00228] in Northern blot assays. Northern blotting involves the use of electrophoresis to separate RNA samples by size. The RNA targeting effector protein can be used to specifically bind and detect the target RNA sequence.

[00229] A RNA targeting effector protein can also be fused to a fluorescent protein (such as GFP) and used to track RNA localization in living cells. More particularly, the RNA targeting effector protein can be inactivated in that it no longer cleaves RNA. In particular embodiments, it is envisaged that a split RNA targeting effector protein can be used, whereby the signal is dependent on the binding of both subproteins, in order to ensure a more precise visualization. Alternatively, a split fluorescent protein can be used that is reconstituted when multiple RNA targeting effector protein complexes bind to the target transcript. It is further envisaged that a transcript is targeted at multiple binding sites along the mRNA so the fluorescent signal can amplify the true signal and allow for focal identification. As yet another alternative, the fluorescent protein can be reconstituted from a split intein.

[00230] RNA targeting effector proteins are for instance suitably used to determine the localization of the RNA or specific splice variants, the level of mRNA transcript, up- or down regulation of transcripts and disease-specific diagnosis. The RNA targeting effector proteins can be used for visualization of RNA in (living) cells using e.g. fluorescent microscopy or flow cytometry, such as fluorescence-activated cell sorting (FACS) which allows for high-throughput screening of cells and recovery of living cells following cell sorting. Further, expression levels of different transcripts can be assessed simultaneously under stress, e.g. inhibition of cancer growth using molecular inhibitors or hypoxic conditions on cells. Another application would be to track localization of transcripts to synaptic connections during a neural stimulus using two photon microscopy.

[00231] In certain embodiments, the components or complexes according to the invention as described herein can be used in multiplexed error-robust fluorescence in situ hybridization

(MERFISH; Chen et al. Science; 2015; 348(6233)), such as for instance with (fluorescently) labeled Cas13 effectors.

In vitro apex labeling

[00232] Cellular processes depend on a network of molecular interactions among protein, RNA, and DNA. Accurate detection of protein-DNA and protein-RNA interactions is key to understanding such processes. In vitro proximity labeling technology employs an affinity tag combined with e.g. a photoactivatable probe to label polypeptides and RNAs in the vicinity of a protein or RNA of interest in vitro. After UV irradiation the photoactivatable group reacts with proteins and other molecules that are in close proximity to the tagged molecule, thereby labelling them. Labelled interacting molecules can subsequently be recovered and identified. The RNA targeting effector protein of the invention can for instance be used to target a probe to a selected RNA sequence.

[00233] These applications could also be applied in animal models for in vivo imaging of disease relevant applications or difficult-to culture cell types.

[00234] The invention provides agents and methods for diagnosing and monitoring health states through non-invasive sampling of cell free RNA, including testing for risk and guiding RNA-targeted therapies, and is useful in setting where rapid administration of therapy is important to treatment outcomes. In one embodiment, the invention provides cancer detection methods and agents for circulating tumor RNA, including for monitoring recurrence and/or development of common drug resistance mutations. In another embodiment, the invention provides detection methods and agents for detection and/or identification of bacterial species directly from blood or serum to monitor, e.g., disease progression and sepsis. In an embodiment of the invention, the Cas13 proteins and derivative are used to distinguish and diagnose common diseases such as rhinovirus or upper respiratory tract infections from more serious infections such as bronchitis.

[00235] The invention provides methods and agents for rapid genotyping for emergency pharmacogenomics, including guidance for administration of anticoagulants during myocardial infarction or stroke treatment based on, e.g., VKORC1, CYP2C9, and CYP2C19 genotyping.

[00236] The invention provides agents and methods for monitoring food contamination by bacteria at all points along a food production and delivery chain. In another embodiment, the invention provides for quality control and monitoring, e.g. by identification of food sources and determination of purity. In one non-limiting example, the invention may be used to identify or confirm a food sources, such as a species of animal meat and seafood.

[00237] In another embodiment, the invention is used in forensic determinations. For example, crime scene samples containing blood or other bodily fluids. In an embodiment of the invention, the invention is used to identify nucleic acid samples from fingerprints.

Use of RNA-targeting effector protein in RNA origami/in vitro assembly lines - combinatorics

[00238] RNA origami refers to nanoscale folded structures for creating two-dimensional or three-dimensional structures using RNA as integrated template. The folded structure is encoded in the RNA and the shape of the resulting RNA is thus determined by the synthesized RNA sequence (Geary, et al. 2014. Science, 345 (6198). pp. 799-804). The RNA origami may act as scaffold for arranging other components, such as proteins, into complexes. The RNA targeting effector protein of the invention can for instance be used to target proteins of interest to the RNA origami using a suitable guide RNA.

Use of RNA-targeting effector protein in RNA isolation or purification, enrichment or depletion

[00239] It is further envisaged that the RNA targeting effector protein when complexed to RNA can be used to isolate and/or purify the RNA. The RNA targeting effector protein can for instance be fused to an affinity tag that can be used to isolate and/or purify the RNA-RNA targeting effector protein complex. Such applications are for instance useful in the analysis of gene expression profiles in cells. In particular embodiments, it can be envisaged that the RNA targeting effector proteins can be used to target a specific noncoding RNA (ncRNA) thereby blocking its activity, providing a useful functional probe. In certain embodiments, the effector protein as described herein may be used to specifically enrich for a particular RNA (including but not limited to increasing stability, etc.), or alternatively to specifically deplete a particular RNA (such as without limitation for instance particular splice variants, isoforms, etc.).

Interrogation of lincRNA function and other nuclear RNAs

[00240] Current RNA knockdown strategies such as siRNA have the disadvantage that they are mostly limited to targeting cytosolic transcripts since the protein machinery is cytosolic. The advantage of a RNA targeting effector protein of the present invention, an exogenous system that is not essential to cell function, is that it can be used in any compartment in the cell. By fusing a NLS signal to the RNA targeting effector protein, it can be guided to the nucleus, allowing nuclear RNAs to be targeted. It is for instance envisaged to probe the function of lincRNAs. Long intergenic non-coding RNAs (lincRNAs) are a

vastly underexplored area of research. Most lincRNAs have as of yet unknown functions which could be studied using the RNA targeting effector protein of the invention.

Identification of RNA binding proteins

[00241] Identifying proteins bound to specific RNAs can be useful for understanding the roles of many RNAs. For instance, many lincRNAs associate with transcriptional and epigenetic regulators to control transcription. Understanding what proteins bind to a given lincRNA can help elucidate the components in a given regulatory pathway. A RNA targeting effector protein of the invention can be designed to recruit a biotin ligase to a specific transcript in order to label locally bound proteins with biotin. The proteins can then be pulled down and analyzed by mass spectrometry to identify them.

Assembly of complexes on RNA and substrate shuttling

[00242] RNA targeting effector proteins of the invention can further be used to assemble complexes on RNA. This can be achieved by functionalizing the RNA targeting effector protein with multiple related proteins (e.g. components of a particular synthesis pathway). Alternatively, multiple RNA targeting effector proteins can be functionalized with such different related proteins and targeted to the same or adjacent target RNA. Useful applications of assembling complexes on RNA are for instance facilitating substrate shuttling between proteins.

Synthetic biology

[00243] The development of biological systems have a wide utility, including in clinical applications. It is envisaged that the programmable RNA targeting effector proteins of the invention can be used fused to split proteins of toxic domains for targeted cell death, for instance using cancer-linked RNA as target transcript. Further, pathways involving protein-protein interaction can be influenced in synthetic biological systems with e.g. fusion complexes with the appropriate effectors such as kinases or other enzymes.

Protein splicing: inteins

[00244] Protein splicing is a post-translational process in which an intervening polypeptide, referred to as an intein, catalyzes its own excision from the polypeptides flanking it, referred to as exteins, as well as subsequent ligation of the exteins. The assembly of two or more RNA targeting effector proteins as described herein on a target transcript could be used to direct the release of a split intein (Topilina and Mills *Mob DNA*. 2014 Feb 4;5(1):5), thereby allowing for direct computation of the existence of a mRNA transcript and subsequent release of a protein product, such as a metabolic enzyme or a transcription factor (for downstream actuation of transcription pathways). This application may have significant

relevance in synthetic biology (see above) or large-scale bioproduction (only produce product under certain conditions).

Inducible, dosed and self-inactivating systems

[00245] In one embodiment, fusion complexes comprising an RNA targeting effector protein of the invention and an effector component are designed to be inducible, for instance light inducible or chemically inducible. Such inducibility allows for activation of the effector component at a desired moment in time.

[00246] Light inducibility is for instance achieved by designing a fusion complex wherein CRY2PHR/CIBN pairing is used for fusion. This system is particularly useful for light induction of protein interactions in living cells (Konermann S, et al. *Nature*. 2013;500:472-476).

[00247] Chemical inducibility is for instance provided for by designing a fusion complex wherein FKBP/FRB (FK506 binding protein / FKBP rapamycin binding) pairing is used for fusion. Using this system rapamycin is required for binding of proteins (Zetsche et al. *Nat Biotechnol*. 2015;33(2): 139-42 describes the use of this system for Cas9) .

[00248] Further, when introduced in the cell as DNA, the RNA targeting effector protein of the inventions can be modulated by inducible promoters, such as tetracycline or doxycycline controlled transcriptional activation (Tet-On and Tet-Off expression system), hormone inducible gene expression system such as for instance an ecdysone inducible gene expression system and an arabinose-inducible gene expression system. When delivered as RNA, expression of the RNA targeting effector protein can be modulated via a riboswitch, which can sense a small molecule like tetracycline (as described in Goldfless et al. *Nucleic Acids Res*. 2012;40(9):e64).

[00249] In one embodiment, the delivery of the RNA targeting effector protein of the invention can be modulated to change the amount of protein or crRNA in the cell, thereby changing the magnitude of the desired effect or any undesired off-target effects.

[00250] In one embodiment, the RNA targeting effector proteins described herein can be designed to be self-inactivating. When delivered to a cell as RNA, either mRNA or as a replication RNA therapeutic (Wroblecka et al *Nat Biotechnol*. 2015 Aug; 33(8): 839-841), they can self-inactivate expression and subsequent effects by destroying the own RNA, thereby reducing residency and potential undesirable effects.

[00251] For further in vivo applications of RNA targeting effector proteins as described herein, reference is made to Mackay JP et al (*Nat Struct Mol Biol*. 2011 Mar;18(3):256-61), Nelles et al (*Bioessays*. 2015 Jul;37(7):732-9) and Abil Z and Zhao H (*Mol Biosyst*. 2015

Oct;1 l(10):2658-65), which are incorporated herein by reference. In particular, the following applications are envisaged in certain embodiments of the invention, preferably in certain embodiments by using catalytically inactive Casl3: enhancing translation (e.g. Casl3 - translation promotion factor fusions (e.g. eIF4 fusions)); repressing translation (e.g. gRNA targeting ribosome binding sites); exon skipping (e.g. gRNAs targeting splice donor and/or acceptor sites); exon inclusion (e.g. gRNA targeting a particular exon splice donor and/or acceptor site to be included or Casl3 fused to or recruiting spliceosome components (e.g. U1 snRNA)); accessing RNA localization (e.g. Casl3 - marker fusions (e.g. EGFP fusions)); altering RNA localization (e.g. Casl3 - localization signal fusions (e.g. NLS or NES fusions)); RNA degradation (in this case no catalytically inactive Casl3 is to be used if relied on the activity of Casl3, alternatively and for increased specificity, a split Casl3 may be used); inhibition of non-coding RNA function (e.g. miRNA), such as by degradation or binding of gRNA to functional sites (possibly titrating out at specific sites by relocalization by Casl3-signal sequence fusions).

[00252] As described herein before and demonstrated in the Examples, Casl3 function is robust to 5' or 3' extensions of the crRNA and to extension of the crRNA loop. It is therefore envisaged that MS2 loops and other recruitment domains can be added to the crRNA without affecting complex formation and binding to target transcripts. Such modifications to the crRNA for recruitment of various effector domains are applicable in the uses of a RNA targeted effector proteins described above.

[00253] As demonstrated in the Examples, Casl3, in particular LshCasl3, is capable of mediating resistance to RNA phages. It is therefore envisaged that Casl3 can be used to immunize, e.g. animals, humans and plants, against RNA-only pathogens, including but not limited to retroviruses (e.g. lentiviruses, such as HIV), HCV, Ebola virus and Zika virus.

[00254] The present inventors have shown that Casl3 can process (cleaves) its own array. This applies to both the wildtype Casl3 protein and the mutated Casl3 protein containing one or more mutated amino acid residues R597, H602, R1278 and H1283, such as one or more of the modifications selected from R597A, H602A, R1278A and H1283A. It is therefore envisaged that multiple crRNAs designed for different target transcripts and/or applications can be delivered as a single pre-crRNA or as a single transcript driven by one promoter. Such method of delivery has the advantages that it is substantially more compact, easier to synthesize and easier to delivery in viral systems. Preferably, amino acid numbering as described herein refers to Lsh Casl3 protein. It will be understood that exact amino acid

positions may vary for orthologues of Lsh Casl3, which can be adequately determined by protein alignment, as is known in the art, and as described herein elsewhere.

[00255] Aspects of the invention also encompass methods and uses of the compositions and systems described herein in genome or transcriptome engineering, e.g. for altering or manipulating the (protein) expression of one or more genes or the one or more gene products, in prokaryotic or eukaryotic cells, *in vitro*, *in vivo* or *ex vivo*.

[00256] In an aspect, the invention provides methods and compositions for modulating, e.g., reducing, (protein) expression of a target RNA in cells. In the subject methods, a Casl3 system of the invention is provided that interferes with transcription, stability, and / or translation of an RNA.

[00257] In certain embodiments, an effective amount of Casl3 system is used to cleave RNA or otherwise inhibit RNA expression. In this regard, the system has uses similar to siRNA and shRNA, thus can also be substituted for such methods. The method includes, without limitation, use of a Casl3 system as a substitute for e.g., an interfering ribonucleic acid (such as an siRNA or shRNA) or a transcription template thereof, e.g., a DNA encoding an shRNA. The Casl3 system is introduced into a target cell, e.g., by being administered to a mammal that includes the target cell,

[00258] Advantageously, a Casl3 system of the invention is specific. For example, whereas interfering ribonucleic acid (such as an siRNA or shRNA) polynucleotide systems are plagued by design and stability issues and off-target binding, a Casl3 system of the invention can be designed with high specificity.

Destabilized Casl3

[00259] In certain embodiments, the effector protein (CRISPR enzyme; Casl3) according to the invention as described herein is associated with or fused to a destabilization domain (DD). In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, 4HT. As such, in some embodiments, one of the at least one DDs is ER50 and a stabilizing ligand therefor is 4HT or CMP8. In some embodiments, the DD is DHFR50. A corresponding stabilizing ligand for this DD is, in some embodiments, TMP. As such, in some embodiments, one of the at least one DDs is DHFR50 and a stabilizing ligand therefor is TMP. In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, CMP8. CMP8 may therefore be an alternative stabilizing ligand to 4HT in the ER50 system. While it may be possible that CMP8 and 4HT can/should be used in a competitive matter, some cell types may be more

susceptible to one or the other of these two ligands, and from this disclosure and the knowledge in the art the skilled person can use CMP8 and/or 4HT.

[00260] In some embodiments, one or two DDs may be fused to the N-terminal end of the CRISPR enzyme with one or two DDs fused to the C-terminal of the CRISPR enzyme. In some embodiments, the at least two DDs are associated with the CRISPR enzyme and the DDs are the same DD, i.e. the DDs are homologous. Thus, both (or two or more) of the DDs could be ER50 DDs. This is preferred in some embodiments. Alternatively, both (or two or more) of the DDs could be DHFR50 DDs. This is also preferred in some embodiments. In some embodiments, the at least two DDs are associated with the CRISPR enzyme and the DDs are different DDs, i.e. the DDs are heterologous. Thus, one of the DDs could be ER50 while one or more of the DDs or any other DDs could be DHFR50. Having two or more DDs which are heterologous may be advantageous as it would provide a greater level of degradation control. A tandem fusion of more than one DD at the N or C-term may enhance degradation; and such a tandem fusion can be, for example ER50-ER50-Cas13 or DHFR-DHFR-Cas13. It is envisaged that high levels of degradation would occur in the absence of either stabilizing ligand, intermediate levels of degradation would occur in the absence of one stabilizing ligand and the presence of the other (or another) stabilizing ligand, while low levels of degradation would occur in the presence of both (or two of more) of the stabilizing ligands. Control may also be imparted by having an N-terminal ER50 DD and a C-terminal DHFR50 DD.

[00261] In some embodiments, the fusion of the CRISPR enzyme with the DD comprises a linker between the DD and the CRISPR enzyme. In some embodiments, the linker is a GlySer linker. In some embodiments, the DD-CRISPR enzyme further comprises at least one Nuclear Export Signal (NES). In some embodiments, the DD-CRISPR enzyme comprises two or more NESs. In some embodiments, the DD-CRISPR enzyme comprises at least one Nuclear Localization Signal (NLS). This may be in addition to an NES. In some embodiments, the CRISPR enzyme comprises or consists essentially of or consists of a localization (nuclear import or export) signal as, or as part of, the linker between the CRISPR enzyme and the DD. HA or Flag tags are also within the ambit of the invention as linkers. Applicants use NLS and/or NES as linker and also use Glycine Serine linkers as short as GS up to (GGGGS)₃.

[00262] [0019] Destabilizing domains have general utility to confer instability to a wide range of proteins; see, e.g., Miyazaki, J Am Chem Soc. Mar 7, 2012; 134(9): 3942-3945, incorporated herein by reference. CMP8 or 4-hydroxytamoxifen can be destabilizing

domains. More generally, A temperature-sensitive mutant of mammalian DHFR (DHFRts), a destabilizing residue by the N-end rule, was found to be stable at a permissive temperature but unstable at 37 °C. The addition of methotrexate, a high-affinity ligand for mammalian DHFR, to cells expressing DHFRts inhibited degradation of the protein partially. This was an important demonstration that a small molecule ligand can stabilize a protein otherwise targeted for degradation in cells. A rapamycin derivative was used to stabilize an unstable mutant of the FRB domain of mTOR (FRB*) and restore the function of the fused kinase, GSK-3p.6,7 This system demonstrated that ligand-dependent stability represented an attractive strategy to regulate the function of a specific protein in a complex biological environment. A system to control protein activity can involve the DD becoming functional when the ubiquitin complementation occurs by rapamycin induced dimerization of FK506-binding protein and FKBP12. Mutants of human FKBP12 or ecDHFR protein can be engineered to be metabolically unstable in the absence of their high-affinity ligands, Shield-1 or trimethoprim (TMP), respectively. These mutants are some of the possible destabilizing domains (DDs) useful in the practice of the invention and instability of a DD as a fusion with a CRISPR enzyme confers to the CRISPR protein degradation of the entire fusion protein by the proteasome. Shield-1 and TMP bind to and stabilize the DD in a dose-dependent manner. The estrogen receptor ligand binding domain (ERLBD, residues 305-549 of ERS1) can also be engineered as a destabilizing domain. Since the estrogen receptor signaling pathway is involved in a variety of diseases such as breast cancer, the pathway has been widely studied and numerous agonist and antagonists of estrogen receptor have been developed. Thus, compatible pairs of ERLBD and drugs are known. There are ligands that bind to mutant but not wild-type forms of the ERLBD. By using one of these mutant domains encoding three mutations (L384M, M421G, G521R)¹², it is possible to regulate the stability of an ERLBD-derived DD using a ligand that does not perturb endogenous estrogen-sensitive networks. An additional mutation (Y537S) can be introduced to further destabilize the ERLBD and to configure it as a potential DD candidate. This tetra-mutant is an advantageous DD development. The mutant ERLBD can be fused to a CRISPR enzyme and its stability can be regulated or perturbed using a ligand, whereby the CRISPR enzyme has a DD. Another DD can be a 12-kDa (107-amino-acid) tag based on a mutated FKBP protein, stabilized by Shield1 ligand; see, e.g., Nature Methods 5, (2008). For instance a DD can be a modified FK506 binding protein 12 (FKBP12) that binds to and is reversibly stabilized by a synthetic, biologically inert small molecule, Shield-1; see, e.g., Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein

function in living cells using synthetic small molecules. Cell. 2006;126:995-1004; Banaszynski LA, Sellmyer MA, Contag CH, Wandless TJ, Thorne SH. Chemical control of protein stability and function in living mice. Nat Med. 2008;14:1 123-1 127; Maynard-Smith LA, Chen LC, Banaszynski LA, Ooi AG, Wandless TJ. A directed approach for engineering conditional protein stability using biologically silent small molecules. The Journal of biological chemistry. 2007;282:24866-24872; and Rodriguez, Chem Biol. Mar 23, 2012; 19(3): 391-398—all of which are incorporated herein by reference and may be employed in the practice of the invention in selected a DD to associate with a CRISPR enzyme in the practice of this invention. As can be seen, the knowledge in the art includes a number of DDs, and the DD can be associated with, e.g., fused to, advantageously with a linker, to a CRISPR enzyme, whereby the DD can be stabilized in the presence of a ligand and when there is the absence thereof the DD can become destabilized, whereby the CRISPR enzyme is entirely destabilized, or the DD can be stabilized in the absence of a ligand and when the ligand is present the DD can become destabilized; the DD allows the CRISPR enzyme and hence the CRISPR-Cas complex or system to be regulated or controlled—turned on or off so to speak, to thereby provide means for regulation or control of the system, e.g., in an in vivo or in vitro environment. For instance, when a protein of interest is expressed as a fusion with the DD tag, it is destabilized and rapidly degraded in the cell, e.g., by proteasomes. Thus, absence of stabilizing ligand leads to a D associated Cas being degraded. When a new DD is fused to a protein of interest, its instability is conferred to the protein of interest, resulting in the rapid degradation of the entire fusion protein. Peak activity for Cas is sometimes beneficial to reduce off-target effects. Thus, short bursts of high activity are preferred. The present invention is able to provide such peaks. In some senses the system is inducible. In some other senses, the system repressed in the absence of stabilizing ligand and de-repressed in the presence of stabilizing ligand.

Cas13 mutations

[00263] In certain embodiments, the effector protein (CRISPR enzyme; Cas13; effector protein) according to the invention as described herein is a catalytically inactive or dead Cas13 effector protein (dCas13). In some embodiments, the dCas13 effector comprises mutations in the nuclease domain. In some embodiments, the dCas13 effector protein has been truncated. To reduce the size of a fusion protein of the Cas13 effector and the one or more functional domains, the C-terminus of the Cas13 effector can be truncated while still maintaining its RNA binding function. For example, at least 20 amino acids, at least 50 amino acids, at least 80 amino acids, or at least 100 amino acids, or at least 150 amino acids,

or at least 200 amino acids, or at least 250 amino acids, or at least 300 amino acids, or at least 350 amino acids, or up to 120 amino acids, or up to 140 amino acids, or up to 160 amino acids, or up to 180 amino acids, or up to 200 amino acids, or up to 250 amino acids, or up to 300 amino acids, or up to 350 amino acids, or up to 400 amino acids, may be truncated at the C-terminus of the Cas13b effector. Specific examples of Cas13 truncations include C-terminal Δ 984-1090, C-terminal Δ 1026-1090, and C-terminal Δ 1053-1090, C-terminal Δ 934-1090, C-terminal Δ 884-1090, C-terminal Δ 834-1090, C-terminal Δ 784-1090, and C-terminal Δ 734-1090, wherein amino acid positions correspond to amino acid positions of *Prevotella* sp. P5-125 Cas13b protein. See Figs. 9A-9B.

Application of RNA targeting RNA targeting -CRISPR system to plants and yeast

Definitions:

[00264] In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants. Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, Clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini. The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants.

[00265] The methods for modulating gene expression using the RNA targeting system as described herein can be used to confer desired traits on essentially any plant. A wide variety

of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., *Arabidopsis*). Thus, the methods and CRISPR-Cas systems can be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders Magnoliales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papaverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Resales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, Santales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales; the methods and CRISPR-Cas systems can be used with monocotyledonous plants such as those belonging to the orders Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchidales, or with plants belonging to Gymnospermae, e.g. those belonging to the orders Finales, Ginkgoales, Cycadales, Araucariales, Cupressales and Gnetales.

[00266] The RNA targeting CRISPR systems and methods of use described herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa*, *Alseodaphne*, *Anacardium*, *Arachis*, *Beilschmiedia*, *Brassica*, *Carthamiis*, ***Cocculus***, *Croton*, *Cucumis*, *dimis*, *Citrullus*, *Capsicum*, *Catharanthus*, *Cocos*, *Coffea*, *Cucurbita*, *Daucus*, *Duguetia*, *Eschscholzia*, *Ficus*, *Fragaria*, *Glaucium*, *Glycine*, *Gossypium*, *Helianthus*, *Hevea*, *Hyoscyamus*, *Lactuca*, *Landolphia*, *Linum*, *Litsea*, *Lycopersicon*, *Lupinus*, *Mecumha*, *Majonica*, *Matricaria*, *Medicago*,

Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrtis, Prunus, Raphanus, Ricinus, Senecio, Sinomemum, Stephania, Sinapis, Solarium, Theobroma, TrifoUum, Trigonella, Vicia, Vinca, It/s, and Vigna; and the genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festnca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, l ea, Abies, Cunninghamia, Ephedra, Picea, Piniis, and Pseudotsuga.

[00267] The RNA targeting CRISPR systems and methods of use can also be used over a broad range of "algae" or "algae cells", including for example algae selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae), Bacillariophyta (diatoms), Eustigmatophyta and dinoflagellates as well as the prokaryotic phylum Cyanobacteria (blue-green algae). The term "algae" includes for example algae selected from : Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cyllindrotheca, Dunaliella, Emiiiana, Eugiena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraseimis, Thalassiosira, and Trichodesmium.

[00268] A part of a plant, i.e., a "plant tissue" may be treated according to the methods of the present invention to produce an improved plant. Plant tissue also encompasses plant cells. The term "plant cell" as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in in vitro tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[00269] A "protoplast" refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

[00270] The term "transformation" broadly refers to the process by which a plant host is genetically modified by the introduction of DNA by means of *Agrobacteria* or one of a variety of chemical or physical methods. As used herein, the term "plant host" refers to plants, including any cells, tissues, organs, or progeny of the plants. Many suitable plant tissues or plant cells can be transformed and include, but are not limited to, protoplasts, somatic embryos, pollen, leaves, seedlings, stems, calli, stolons, microtubers, and shoots. A

plant tissue also refers to any clone of such a plant, seed, progeny, propagule whether generated sexually or asexually, and descendants of any of these, such as cuttings or seed.

[00271] The term "transformed" as used herein, refers to a cell, tissue, organ, or organism into which a foreign DNA molecule, such as a construct, has been introduced. The introduced DNA molecule may be integrated into the genomic DNA of the recipient cell, tissue, organ, or organism such that the introduced DNA molecule is transmitted to the subsequent progeny. In these embodiments, the "transformed" or "transgenic" cell or plant may also include progeny of the cell or plant and progeny produced from a breeding program employing such a transformed plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of the introduced DNA molecule. Preferably, the transgenic plant is fertile and capable of transmitting the introduced DNA to progeny through sexual reproduction.

[00272] The term "progeny", such as the progeny of a transgenic plant, is one that is born of, begotten by, or derived from a plant or the transgenic plant. The introduced DNA molecule may also be transiently introduced into the recipient cell such that the introduced DNA molecule is not inherited by subsequent progeny and thus not considered "transgenic". Accordingly, as used herein, a "non-transgenic" plant or plant cell is a plant which does not contain a foreign DNA stably integrated into its genome.

[00273] The term "plant promoter" as used herein is a promoter capable of initiating transcription in plant cells, whether or not its origin is a plant cell. Exemplary suitable plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria such as *Agrobacterium* or *Rhizobium* which comprise genes expressed in plant cells.

[00274] As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[00275] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an *S. cerevisiae*, *Kluyveromyces marxianus*, or *Issatchenkia orientalis* cell. Other yeast cells may include without limitation *Candida* spp. (e.g., *Candida albicans*),

Yarrowia spp. (e.g., *Yarrowia lipolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora* spp. (e.g., *Neurospora crassa*), *Fusarium* spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., *Issatchenkia orientalis*, a.k.a. *Pichia kudriavzevii* and *Candida acidothermophilum*). In some embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation *Aspergillus* spp. (e.g., *Aspergillus niger*), *Trichoderma* spp. (e.g., *Trichoderma reesei*), *Rhizopus* spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

[00276] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains may include, without limitation, JAY270 and ATCC4124.

[00277] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest. Without wishing to be bound to theory, it is thought that the abundance of guideRNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the Cas13 CRISPRs system described herein may take advantage of using a certain fungal cell type.

[00278] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S.*

cerevisiae strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S. cerevisiae* strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[00279] As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in *Yeast Protocols*, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) *Biotechnology (NY)* 9(1 1): 1067-72. Yeast vectors may contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2 μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

Stable integration of RNA targeting CRISP system components in the genome of plants and plant cells

[00280] In particular embodiments, it is envisaged that the polynucleotides encoding the components of the RNA targeting CRISPR system are introduced for stable integration into the genome of a plant cell. In these embodiments, the design of the transformation vector or the expression system can be adjusted depending on when, where and under what conditions the guide RNA and/or the RNA targeting gene(s) are expressed.

[00281] In particular embodiments, it is envisaged to introduce the components of the RNA targeting CRISPR system stably into the genomic DNA of a plant cell. Additionally or alternatively, it is envisaged to introduce the components of the RNA targeting CRISPR

system for stable integration into the DNA of a plant organelle such as, but not limited to a plastid, a mitochondrion or a chloroplast.

[00282] The expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the guide RNA and/or RNA targeting enzyme in a plant cell; a 5' untranslated region to enhance expression; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the one or more guide RNAs and/or the RNA targeting gene sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

[00283] The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

In a particular embodiment, a RNA targeting CRISPR expression system comprises at least:

- (a) a nucleotide sequence encoding a guide RNA (gRNA) that hybridizes with a target sequence in a plant, and wherein the guide RNA comprises a guide sequence and a direct repeat sequence, and
 - (b) a nucleotide sequence encoding a RNA targeting protein,
- wherein components (a) or (b) are located on the same or on different constructs, and whereby the different nucleotide sequences can be under control of the same or a different regulatory element operable in a plant cell.

[00284] DNA construct(s) containing the components of the RNA targeting CRISPR system, and, where applicable, template sequence may be introduced into the genome of a plant, plant part, or plant cell by a variety of conventional techniques. The process generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom.

In particular embodiments, the DNA construct may be introduced into the plant cell using techniques such as but not limited to electroporation, microinjection, aerosol beam injection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see also Fu et al., *Transgenic Res.* 2000 Feb;9(1):1-9). The basis of particle bombardment is the acceleration of particles coated with gene/s of interest toward cells, resulting in the penetration of the protoplasm by the particles and typically stable integration into the genome. (see e.g. Klein et al, *Nature* (1987), Klein et al, *Bio/Technology* (1992), Casas et al, *Proc. Natl. Acad. Sci. USA* (1993).).

[00285] In particular embodiments, the DNA constructs containing components of the RNA targeting CRISPR system may be introduced into the plant by *Agrobacterium-mediated* transformation. The DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The foreign DNA can be incorporated into the genome of plants by infecting the plants or by incubating plant protoplasts with *Agrobacterium* bacteria, containing one or more Ti (tumor-inducing) plasmids. (see e.g. Fraey et al., (1985), Rogers et al., (1987) and U.S. Pat. No. 5,563,055).

Plant promoters

[00286] In order to ensure appropriate expression in a plant cell, the components of the Cas13 CRISPR system described herein are typically placed under control of a plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged.

[00287] A constitutive plant promoter is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One non-limiting example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. The present invention envisages methods for modifying RNA sequences and as such also envisages regulating expression of plant biomolecules. In particular embodiments of the present invention it is thus advantageous to place one or more elements of the RNA targeting CRISPR system under the control of a promoter that can be regulated. "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the RNA targeting CRISPR components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter. Tissue-preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the RNA targeting CRISPR system are found in Kawamata et al., (1997) *Plant Cell Physiol* 38:792-803; Yamamoto et al., (1997) *Plant J* 12:255-65; Hire et al., (1992) *Plant Mol Biol* 20:207-18, Kuster et al., (1995) *Plant Mol Biol* 29:759-72, and Capana et al., (1994) *Plant Mol Biol* 25:681-91.

Examples of promoters that are inducible and that allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include a RNA targeting CRISPR enzyme, a light-responsive cytochrome heterodimer (e.g. from *Arabidopsis thaliana*), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283, which is hereby incorporated by reference in its entirety.

[00288] In particular embodiments, transient or inducible expression can be achieved by using, for example, chemical-regulated promoters, i.e. whereby the application of an exogenous chemical induces gene expression. Modulating of gene expression can also be obtained by a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize In2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) *Plant Cell Physiol* 38:568-77), the maize GST promoter (GST-11-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1 a promoter (Ono et al., (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) *Mol Gen Genet* 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

Translocation to and/or expression in specific plant organelles

[00289] The expression system may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[00290] In particular embodiments, it is envisaged that the RNA targeting CRISPR system is used to specifically modify expression and/or translation of chloroplast genes or to ensure expression in the chloroplast. For this purpose use is made of chloroplast transformation methods or compartmentalization of the RNA targeting CRISPR components to the chloroplast. For instance, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[00291] Methods of chloroplast transformation are known in the art and include Particle bombardment, PEG treatment, and microinjection. Additionally, methods involving the translocation of transformation cassettes from the nuclear genome to the plastid can be used as described in WO20 10061 186.

[00292] Alternatively, it is envisaged to target one or more of the RNA targeting CRISPR components to the plant chloroplast. This is achieved by incorporating in the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the RNA targeting protein. The CTP is removed in a processing step during translocation into the chloroplast. Chloroplast targeting of expressed proteins is well known to the skilled artisan (see for instance Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180) . In such embodiments it is also desired to target the one or more guide RNAs to the plant chloroplast. Methods and constructs which can be used for translocating guide RNA into the chloroplast by means of a chloroplast localization sequence are described, for instance, in US 20040142476, incorporated herein by reference. Such variations of constructs can be incorporated into the expression systems of the invention to efficiently translocate the RNA targeting -guide RNA(s).

Introduction of polynucleotides encoding the CRISPR-RNA targeting system in Algal cells.

[00293] Transgenic algae (or other plants such as rape) may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol) or other products. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00294] US 8945839 describes a method for engineering Micro-Algae (*Chlamydomonas reinhardtii* cells) species) using Cas9. Using similar tools, the methods of the RNA targeting CRISPR system described herein can be applied on *Chlamydomonas* species and other algae. In particular embodiments, RNA targeting protein and guide RNA(s) are introduced in algae expressed using a vector that expresses RNA targeting protein under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA is optionally delivered using a vector containing T7 promoter. Alternatively, RNA targeting mRNA and in vitro transcribed guide RNA can be delivered to algal cells. Electroporation protocols are available to the skilled person such as the standard recommended protocol from the GeneArt *Chlamydomonas* Engineering kit.

Introduction of polynucleotides encoding RNA targeting components in yeast cells

[00295] In particular embodiments, the invention relates to the use of the RNA targeting CRISPR system for RNA editing in yeast cells. Methods for transforming yeast cells which can be used to introduce polynucleotides encoding the RNA targeting CRISPR system components are well known to the artisan and are reviewed by Kawai et al., 2010, *Bioeng Bugs*. 2010 Nov-Dec; 1(6): 395-403). Non-limiting examples include transformation of yeast cells by lithium acetate treatment (which may further include carrier DNA and PEG treatment), bombardment or by electroporation.

Transient expression of RNA targeting CRISPR system components in plants and plant cell

[00296] In particular embodiments, it is envisaged that the guide RNA and/or RNA targeting gene are transiently expressed in the plant cell. In these embodiments, the RNA targeting CRISPR system can ensure modification of RNA target molecules only when both the guide RNA and the RNA targeting protein is present in a cell, such that gene expression can further be controlled. As the expression of the RNA targeting enzyme is transient, plants regenerated from such plant cells typically contain no foreign DNA. In particular embodiments the RNA targeting enzyme is stably expressed by the plant cell and the guide sequence is transiently expressed.

[00297] In particularly preferred embodiments, the RNA targeting CRISPR system components can be introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, *Annu Rev Phytopathol.* 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, tobnavirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses are non-integrative vectors, which is of interest in the context of avoiding the production of GMO plants.

[00298] In particular embodiments, the vector used for transient expression of RNA targeting CRISPR constructs is for instance a pEAQ vector, which is tailored for *Agrobacterium*-mediated transient expression (Sainsbury F. et al., *Plant Biotechnol J.* 2009 Sep;7(7):682-93) in the protoplast. Precise targeting of genomic locations was demonstrated using a modified Cabbage Leaf Curl virus (CaLCuV) vector to express gRNAs in stable transgenic plants expressing a CRISPR enzyme (Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[00299] In particular embodiments, double-stranded DNA fragments encoding the guide RNA and/or the RNA targeting gene can be transiently introduced into the plant cell. In such embodiments, the introduced double-stranded DNA fragments are provided in sufficient quantity to modify RNA molecule(s) in the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for direct DNA transfer in plants are known by the skilled artisan (see for instance Davey et al. *Plant Mo! Biol.* 1989 Sep;13(3):273-85.)

[00300] In other embodiments, an RNA polynucleotide encoding the RNA targeting protein is introduced into the plant cell, which is then translated and processed by the host cell generating the protein in sufficient quantity to modify the RNA molecule(s) cell (in the presence of at least one guide RNA) but which does not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for introducing mRNA to plant protoplasts for transient expression are known by the skilled artisan (see for instance in Gallie, *Plant Cell Reports* (1993), 13;119-122). Combinations of the different methods described above are also envisaged.

Delivery of RNA targeting CRISPR components to the plant cell

[00301] In particular embodiments, it is of interest to deliver one or more components of the RNA targeting CRISPR system directly to the plant cell. This is of interest, inter alia, for the generation of non-transgenic plants (see below). In particular embodiments, one or more of the RNA targeting components is prepared outside the plant or plant cell and delivered to the cell. For instance in particular embodiments, the RNA targeting protein is prepared in vitro prior to introduction to the plant cell. RNA targeting protein can be prepared by various methods known by one of skill in the art and include recombinant production. After expression, the RNA targeting protein is isolated, refolded if needed, purified and optionally treated to remove any purification tags, such as a His-tag. Once crude, partially purified, or more completely purified RNA targeting protein is obtained, the protein may be introduced to the plant cell.

[00302] In particular embodiments, the RNA targeting protein is mixed with guide RNA targeting the RNA of interest to form a pre-assembled ribonucleoprotein.

[00303] The individual components or pre-assembled ribonucleoprotein can be introduced into the plant cell via electroporation, by bombardment with RNA targeting -associated gene product coated particles, by chemical transfection or by some other means of transport across a cell membrane. For instance, transfection of a plant protoplast with a pre-assembled CRISPR ribonucleoprotein has been demonstrated to ensure targeted modification of the

plant genome (as described by Woo et al. *Nature Biotechnology*, 2015; DOI: 10.1038/nbt.3389). These methods can be modified to achieve targeted modification of RNA molecules in the plants.

[00304] In particular embodiments, the RNA targeting CRISPR system components are introduced into the plant cells using nanoparticles. The components, either as protein or nucleic acid or in a combination thereof, can be uploaded onto or packaged in nanoparticles and applied to the plants (such as for instance described in WO 2008042156 and US 20130185823). In particular, embodiments of the invention comprise nanoparticles uploaded with or packed with DNA molecule(s) encoding the RNA targeting protein, DNA molecules encoding the guide RNA and/or isolated guide RNA as described in WO2015089419.

[00305] Further means of introducing one or more components of the RNA targeting CRISPR system to the plant cell is by using cell penetrating peptides (CPP). Accordingly, in particular, embodiments the invention comprises compositions comprising a cell penetrating peptide linked to an RNA targeting protein. In particular embodiments of the present invention, an RNA targeting protein and/or guide RNA(s) is coupled to one or more CPPs to effectively transport them inside plant protoplasts (Ramakrishna (2014, *Genome Res.* 2014 Jun;24(6): 1020-7 for Cas9 in human cells). In other embodiments, the RNA targeting gene and/or guide RNA(s) are encoded by one or more circular or non-circular DNA molecule(s) which are coupled to one or more CPPs for plant protoplast delivery. The plant protoplasts are then regenerated to plant cells and further to plants. CPPs are generally described as short peptides of fewer than 35 amino acids either derived from proteins or from chimeric sequences which are capable of transporting biomolecules across cell membrane in a receptor independent manner. CPP can be cationic peptides, peptides having hydrophobic sequences, amphipatic peptides, peptides having proline-rich and anti-microbial sequence, and chimeric or bipartite peptides (Pooga and Langel 2005). CPPs are able to penetrate biological membranes and as such trigger the movement of various biomolecules across cell membranes into the cytoplasm and to improve their intracellular routing, and hence facilitate interaction of the biomolecule with the target. Examples of CPP include amongst others: Tat, a nuclear transcriptional activator protein required for viral replication by HIV type1, penetratin, Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin $\beta 3$ signal peptide sequence; polyarginine peptide Arg5 sequence, Guanine rich-molecular transporters, sweet arrow peptide, etc...

Target RNA envisaged for plant, algae or fungal applications

[00306] The target RNA, i.e. the RNA of interest, is the RNA to be targeted by the present invention leading to the recruitment to, and the binding of the RNA targeting protein at, the target site of interest on the target RNA. The target RNA may be any suitable form of RNA. This may include, in some embodiments, mRNA. In other embodiments, the target RNA may include transfer RNA (tRNA) or ribosomal RNA (rRNA). In other embodiments the target RNA may include interfering RNA (RNAi), microRNA (miRNA), microswitches, microzymes, satellite RNAs and RNA viruses. The target RNA may be located in the cytoplasm of the plant cell, or in the cell nucleus or in a plant cell organelle such as a mitochondrion, chloroplast or plastid.

[00307] In particular embodiments, the RNA targeting CRISPR system is used to cleave RNA or otherwise inhibit RNA expression.

Use of RNA targeting CRISPR system for modulating plant gene expression via RNA modulation

[00308] The RNA targeting protein may also be used, together with a suitable guide RNA, to target gene expression, via control of RNA processing. The control of RNA processing may include RNA processing reactions such as RNA splicing, including alternative splicing or specifically targeting certain splice variants or isoforms; viral replication (in particular of plant viruses, including virioids in plants and tRNA biosynthesis. The RNA targeting protein in combination with a suitable guide RNA may also be used to control RNA activation (RNAa). RNAa leads to the promotion of gene expression, so control of gene expression may be achieved that way through disruption or reduction of RNAa and thus less promotion of gene expression.

[00309] The RNA targeting effector protein of the invention can further be used for antiviral activity in plants, in particular against RNA viruses. The effector protein can be targeted to the viral RNA using a suitable guide RNA selective for a selected viral RNA sequence. In particular, the effector protein may be an active nuclease that cleaves RNA, such as single stranded RNA. provided is therefore the use of an RNA targeting effector protein of the invention as an antiviral agent. Examples of viruses that can be counteracted in this way include, but are not limited to, Tobacco mosaic virus (TMV), Tomato spotted wilt virus (TSWV), Cucumber mosaic virus (CMV), Potato virus Y (PVY), Cauliflower mosaic virus (CaMV) (RT virus), Plum pox virus (PPV), Brome mosaic virus (BMV) and Potato virus X (PVX).

[00310] Examples of modulating RNA expression in plants, algae or fungi, as an alternative of targeted gene modification are described herein further.

[00311] Of particular interest is the regulated control of gene expression through regulated cleavage of mRNA. This can be achieved by placing elements of the RNA targeting under the control of regulated promoters as described herein.

Use of the RNA targeting CRISPR system to restore the functionality of tRNA molecules.

[00312] Pring et al describe RNA editing in plant mitochondria and chloroplasts that alters mRNA sequences to code for different proteins than the DNA. (Plant Mol. Biol. (1993) 21 (6): 1163- 1170. doi: 10.1007/BF00023611). In particular embodiments of the invention, the elements of the RNA targeting CRISPR system specifically targetting mitochondrial and chloroplast mRNA can be introduced in a plant or plant cell to express different proteins in such plant cell organelles mimicking the processes occurring *in vivo*.

Use of the RNA targeting CRISPR system as an alternative to RNA interference to inhibit RNA expression.

[00313] The RNA targeting CRISPR system has uses similar to RNA inhibition or RNA interference, thus can also be substituted for such methods. In particular embodiment, the methods of the present invention include the use of the RNA targeting CRISPR as a substitute for e.g. an interfering ribonucleic acid (such as an siRNA or shRNA or a dsRNA). Examples of inhibition of RNA expression in plants, algae or fungi as an alternative of targeted gene modification are described herein further.

Use of the RNA targeting CRISPR system to control RNA interference.

[00314] Control over interfering RNA or miRNA may help reduce off-target effects (OTE) seen with those approaches by reducing the longevity of the interfering RNA or miRNA *in vivo* or *in vitro*. In particular embodiments, the target RNA may include interfering RNA, i.e. RNA involved in an RNA interference pathway, such as shRNA, siRNA and so forth. In other embodiments, the target RNA may include microRNA (miRNA) or double stranded RNA (dsRNA).

[00315] In other particular embodiments, if the RNA targeting protein and suitable guide RNA(s) are selectively expressed (for example spatially or temporally under the control of a regulated promoter, for example a tissue- or cell cycle-specific promoter and/or enhancer) this can be used to 'protect' the cells or systems (*in vivo* or *in vitro*) from RNAi in those cells. This may be useful in neighbouring tissues or cells where RNAi is not required or for the purposes of comparison of the cells or tissues where the effector protein and suitable guide are and are not expressed (i.e. where the RNAi is not controlled and where it is, respectively). The RNA targeting protein may be used to control or bind to molecules comprising or consisting of RNA, such as ribozymes, ribosomes or riboswitches. In embodiments of the

invention, the guide RNA can recruit the RNA targeting protein to these molecules so that the RNA targeting protein is able to bind to them.

[00316] The RNA targeting CRISPR system of the invention can be applied in areas of *in-planta* RNAi technologies, without undue experimentation, from this disclosure, including insect pest management, plant disease management and management of herbicide resistance, as well as in plant assay and for other applications (see, for instance Kim et al., in *Pesticide Biochemistry and Physiology* (Impact Factor: 2.01). 01/2015; 120. DOI: 10.1016/j.pestbp.2015.01.002; Sharma et al. in *Academic Journals* (2015), Vol.12(18) pp2303-2312); Green J.M, in *Pest Management Science*, Vol 70(9), pp 1351-1357), because the present application provides the foundation for informed engineering of the system.

Use of RNA targeting CRISPR system to modify riboswitches and control metabolic regulation in Plants, Algae and Fungi

[00317] Riboswitches (also known as aptozymes) are regulatory segments of messenger RNA that bind small molecules and in turn regulate gene expression. This mechanism allows the cell to sense the intracellular concentration of these small molecules. A particular riboswitch typically regulates its adjacent gene by altering the transcription, the translation or the splicing of this gene. Thus, in particular embodiments of the present invention, control of riboswitch activity is envisaged through the use of the RNA targeting protein in combination with a suitable guide RNA to target the riboswitch. This may be through cleavage of, or binding to, the riboswitch. In particular embodiments, reduction of riboswitch activity is envisaged. Recently, a riboswitch that binds thiamin pyrophosphate (TPP) was characterized and found to regulate thiamin biosynthesis in plants and algae. Furthermore it appears that this element is an essential regulator of primary metabolism in plants (Bocobza and Aharoni, *Plant J.* 2014 Aug; 79(4):693-703. doi: 10.1111/tbj.12540. Epub 2014 Jun 17). TPP riboswitches are also found in certain fungi, such as in *Neurospora crassa*, where it controls alternative splicing to conditionally produce an Upstream Open Reading Frame (uORF), thereby affecting the expression of downstream genes (Cheah MT et al., (2007)*Nature* 447 (7143): 497-500. doi:10.1038/nature05769) The RNA targeting CRISPR system described herein may be used to manipulate the endogenous riboswitch activity in plants, algae or fungi and as such alter the expression of downstream genes controlled by it. In particular embodiments, the RNA targeting CRISPR system may be used in assaying riboswitch function in vivo or in vitro and in studying its relevance for the metabolic network. In particular embodiments the RNA targeting CRISPR system may potentially be used for engineering of riboswitches as metabolite sensors in plants and platforms for gene control.

Use of RNA targeting CRISPR system in RNAi Screens for plants, algae or fungi

[00318] Identifying gene products whose knockdown is associated with phenotypic changes, biological pathways can be interrogated and the constituent parts identified, via RNAi screens. In particular embodiments of the invention, control may also be exerted over or during these screens by use of the Guide 29 or Guide 30 protein and suitable guide RNA described herein to remove or reduce the activity of the RNAi in the screen and thus reinstate the activity of the (previously interfered with) gene product (by removing or reducing the interference/repression).

Use of RNA targeting proteins for visualization of RNA molecules *in vivo* and *in vitro*

[00319] In particular embodiments, the invention provides a nucleic acid binding system. In situ hybridization of RNA with complementary probes is a powerful technique. Typically fluorescent DNA oligonucleotides are used to detect nucleic acids by hybridization. Increased efficiency has been attained by certain modifications, such as locked nucleic acids (LNAs), but there remains a need for efficient and versatile alternatives. As such, labelled elements of the RNA targeting system can be used as an alternative for efficient and adaptable system for in situ hybridization

Further applications of the RNA targeting CRISPR system in plants and yeasts*Use of RNA targeting CRISPR system in biofuel production*

[00320] The term "biofuel" as used herein is an alternative fuel made from plant and plant-derived resources. Renewable biofuels can be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. There are two types of biofuels: bioethanol and biodiesel. Bioethanol is mainly produced by the sugar fermentation process of cellulose (starch), which is mostly derived from maize and sugar cane. Biodiesel on the other hand is mainly produced from oil crops such as rapeseed, palm, and soybean. Biofuels are used mainly for transportation.

Enhancing plant properties for biofuel production

[00321] In particular embodiments, the methods using the RNA targeting CRISPR system as described herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolysing agents for a more efficient release of sugars for fermentation. In particular embodiments, the biosynthesis of cellulose and/or lignin are modified. Cellulose is the major component of the cell wall. The biosynthesis of cellulose and lignin are co-

regulated. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, the methods described herein are used to downregulate lignin biosynthesis in the plant so as to increase fermentable carbohydrates. More particularly, the methods described herein are used to downregulate at least a first lignin biosynthesis gene selected from the group consisting of 4-coumarate 3-hydroxylase (C3H), phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), caffeic acid O-methyltransferase (COMT), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5-hydroxylase (F5H), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl CoA-reductase (CCR), 4-coumarate-CoA ligase (4CL), monolignol-lignin-specific glycosyltransferase, and aldehyde dehydrogenase (ALDH) as disclosed in WO 2008064289 A2.

[00322] In particular embodiments, the methods described herein are used to produce plant mass that produces lower levels of acetic acid during fermentation (see also WO 2010096488).

Modifying yeast for Biofuel production

[00323] In particular embodiments, the RNA targeting enzyme provided herein is used for bioethanol production by recombinant micro-organisms. For instance, RNA targeting enzymes can be used to engineer micro-organisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. More particularly, the invention provides methods whereby the RNA targeting CRISPR complex is used to modify the expression of endogenous genes required for biofuel production and/or to modify endogenous genes which may interfere with the biofuel synthesis. More particularly the methods involve stimulating the expression in a micro-organism such as a yeast of one or more nucleotide sequence encoding enzymes involved in the conversion of pyruvate to ethanol or another product of interest. In particular embodiments the methods ensure the stimulation of expression of one or more enzymes which allows the micro-organism to degrade cellulose, such as a cellulase. In yet further embodiments, the RNA targeting CRISPR complex is used to suppress endogenous metabolic pathways which compete with the biofuel production pathway.

Modifying Algae and plants for production of vegetable oils or biofuels

[00324] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol),

for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00325] US 8945839 describes a method for engineering Micro-Algae (*Chlamydomonas reinhardtii* cells) species) using Cas9. Using similar tools, the methods of the RNA targeting CRISPR system described herein can be applied on *Chlamydomonas* species and other algae. In particular embodiments, the RNA targeting effector protein and guide RNA are introduced in algae expressed using a vector that expresses the RNA targeting effector protein under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA will be delivered using a vector containing T7 promoter. Alternatively, in vitro transcribed guide RNA can be delivered to algae cells. Electroporation protocol follows standard recommended protocol from the GeneArt *Chlamydomonas* Engineering kit.

Particular applications of the RNA targeting enzymes in plants

[00326] In particular embodiments, present invention can be used as a therapy for virus removal in plant systems as it is able to cleave viral RNA. Previous studies in human systems have demonstrated the success of utilizing CRISPR in targeting the single strand RNA virus, hepatitis C (A. Price, et al., Proc. Natl. Acad. Sci, 2015). These methods may also be adapted for using the RNA targeting CRISPR system in plants.

Improved plants

[00327] The present invention also provides plants and yeast cells obtainable and obtained by the methods provided herein. The improved plants obtained by the methods described herein may be useful in food or feed production through the modified expression of genes which, for instance ensure tolerance to plant pests, herbicides, drought, low or high temperatures, excessive water, etc.

[00328] The improved plants obtained by the methods described herein, especially crops and algae may be useful in food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamin levels than would normally be seen in the wildtype. In this regard, improved plants, especially pulses and tubers are preferred.

[00329] Improved algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00330] The invention also provides for improved parts of a plant. Plant parts include, but are not limited to, leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. Plant parts as envisaged herein may be viable, nonviable, regeneratable, and/or non-regeneratable.

[00331] It is also encompassed herein to provide plant cells and plants generated according to the methods of the invention. Gametes, seeds, embryos, either zygotic or somatic, progeny or hybrids of plants comprising the genetic modification, which are produced by traditional breeding methods, are also included within the scope of the present invention. Such plants may contain a heterologous or foreign DNA sequence inserted at or instead of a target sequence. Alternatively, such plants may contain only an alteration (mutation, deletion, insertion, substitution) in one or more nucleotides. As such, such plants will only be different from their progenitor plants by the presence of the particular modification.

[00332] In an embodiment of the invention, a Cas13 system is used to engineer pathogen resistant plants, for example by creating resistance against diseases caused by bacteria, fungi or viruses. In certain embodiments, pathogen resistance can be accomplished by engineering crops to produce a Cas13 system that will be ingested by an insect pest, leading to mortality. In an embodiment of the invention, a Cas13 system is used to engineer abiotic stress tolerance. In another embodiment, a Cas13 system is used to engineer drought stress tolerance or salt stress tolerance, or cold or heat stress tolerance. Younis et al. 2014, Int. J. Biol. Sci. 10; 1150 reviewed potential targets of plant breeding methods, all of which are amenable to correction or improvement through use of a Cas13 system described herein. Some non-limiting target crops include Arabidops Zea mays is thaliana, Oryza sativa L, Prunus domestica L., Gossypium hirsutum, Nicotiana rustica, Zea mays, Medicago sativa, Nicotiana benthamiana and Arabidopsis thaliana

[00333] In an embodiment of the invention, a Cas13 system is used for management of crop pests. For example, a Cas13 system operable in a crop pest can be expressed from a plant host or transferred directly to the target, for example using a viral vector.

[00334] In an embodiment, the invention provides a method of efficiently producing homozygous organisms from a heterozygous non-human starting organism. In an embodiment, the invention is used in plant breeding. In another embodiment, the invention is used in animal breeding. In such embodiments, a homozygous organism such as a plant or animal is made by preventing or suppressing recombination by interfering with at least one target gene involved in double strand breaks, chromosome pairing and/or strand exchange.

Application of the CAS13 proteins in optimized functional RNA targeting systems

[00335] In an aspect the invention provides a system for specific delivery of functional components to the RNA environment. This can be ensured using the CRISPR systems comprising the RNA targeting effector proteins of the present invention which allow specific targeting of different components to RNA. More particularly such components include

activators or repressors, such as activators or repressors of RNA translation, degradation, etc. Applications of this system are described elsewhere herein.

[00336] According to one aspect the invention provides non-naturally occurring or engineered composition comprising a guide RNA comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the guide RNA is modified by the insertion of one or more distinct RNA sequence(s) that bind an adaptor protein. In particular embodiments, the RNA sequences may bind to two or more adaptor proteins (e.g. aptamers), and wherein each adaptor protein is associated with one or more functional domains. The guide RNAs of the Cas13 enzymes described herein are shown to be amenable to modification of the guide sequence. In particular embodiments, the guide RNA is modified by the insertion of distinct RNA sequence(s) 5' of the direct repeat, within the direct repeat, or 3' of the guide sequence. When there is more than one functional domain, the functional domains can be same or different, e.g., two of the same or two different activators or repressors. In an aspect the invention provides a herein-discussed composition, wherein the one or more functional domains are attached to the RNA targeting enzyme so that upon binding to the target RNA the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function; In an aspect the invention provides a herein-discussed composition, wherein the composition comprises a CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the RNA targeting enzyme and at least two of which are associated with the gRNA.

[00337] Accordingly, In an aspect the invention provides non-naturally occurring or engineered CRISPR-Cas complex composition comprising the guide RNA as herein-discussed and a CRISPR enzyme which is an RNA targeting enzyme, wherein optionally the RNA targeting enzyme comprises at least one mutation, such that the RNA targeting enzyme has no more than 5% of the nuclease activity of the enzyme not having the at least one mutation, and optionally one or more comprising at least one or more nuclear localization sequences. In particular embodiments, the guide RNA is additionally or alternatively modified so as to still ensure binding of the RNA targeting enzyme but to prevent cleavage by the RNA targeting enzyme (as detailed elsewhere herein).

[00338] In particular embodiments, the RNA targeting enzyme is a Cas13 enzyme which has a diminished nuclease activity of at least 97%, or 100% as compared with the Cas13 enzyme not having the at least one mutation. In an aspect the invention provides a herein-discussed composition, wherein the Cas13 enzyme comprises two or more mutations. The

mutations may be selected from mutations of one or more of the following amino acid residues: R597, H602, R1278, and H1283, such as for instance one or more of the following mutations: R597A, H602A, R1278A, and H1283A, according to *Leptotrichia shahii* Cas13 protein or a corresponding position in an ortholog.

[00339] In particular embodiments, an RNA targeting system is provided as described herein above comprising two or more functional domains. In particular embodiments, the two or more functional domains are heterologous functional domain. In particular embodiments, the system comprises an adaptor protein which is a fusion protein comprising a functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain. In particular embodiments, the linker includes a GlySer linker. Additionally or alternatively, one or more functional domains are attached to the RNA effector protein by way of a linker, optionally a GlySer linker. In particular embodiments, the one or more functional domains are attached to the RNA targeting enzyme through one or both of the HEPN domains.

[00340] In an aspect the invention provides a herein-discussed composition, wherein the one or more functional domains associated with the adaptor protein or the RNA targeting enzyme is a domain capable of activating or repressing RNA translation. In an aspect the invention provides a herein-discussed composition, wherein at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity, or molecular switch activity or chemical inducibility or light inducibility.

[00341] In an aspect the invention provides a herein-discussed composition comprising an aptamer sequence. In particular embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In an aspect the invention provides a herein-discussed composition, wherein the aptamer sequence is two or more aptamer sequences specific to different adaptor protein. In an aspect the invention provides a herein-discussed composition, wherein the adaptor protein comprises MS2, PP7, QP, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, Mi1, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕ Ct β -5, ϕ Ct β -8, (λ)Cbl2r, (λ)Cbl23r, 7s, PRR1. Accordingly, in particular embodiments, the aptamer is selected from a binding protein specifically binding any one of the adaptor proteins listed above. In an aspect the invention provides a herein-discussed composition,

wherein the cell is a eukaryotic cell. In an aspect the invention provides a herein-discussed composition, wherein the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell, whereby the mammalian cell is optionally a mouse cell. In an aspect the invention provides a herein-discussed composition, wherein the mammalian cell is a human cell.

[00342] In an aspect the invention provides a herein above-discussed composition wherein there is more than one gRNA, and the gRNAs target different sequences whereby when the composition is employed, there is multiplexing. In an aspect the invention provides a composition wherein there is more than one gRNA modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins.

[00343] In an aspect the invention provides a herein-discussed composition wherein one or more adaptor proteins associated with one or more functional domains is present and bound to the distinct RNA sequence(s) inserted into the guide RNA(s).

[00344] In an aspect the invention provides a herein-discussed composition wherein the guide RNA is modified to have at least one non-coding functional loop; e.g., wherein the at least one non-coding functional loop is repressive; for instance, wherein at least one non-coding functional loop comprises Alu.

[00345] In an aspect the invention provides a method for modifying gene expression comprising the administration to a host or expression in a host in vivo of one or more of the compositions as herein-discussed.

[00346] In an aspect the invention provides a herein-discussed method comprising the delivery of the composition or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed in vivo. In an aspect the invention provides a herein-discussed method wherein the expression in vivo is via a lentivirus, an adenovirus, or an AAV.

[00347] In an aspect the invention provides a mammalian cell line of cells as herein-discussed, wherein the cell line is, optionally, a human cell line or a mouse cell line. In an aspect the invention provides a transgenic mammalian model, optionally a mouse, wherein the model has been transformed with a herein-discussed composition or is a progeny of said transformant.

[00348] In an aspect the invention provides a nucleic acid molecule(s) encoding guide RNA or the RNA targeting CRISPR-Cas complex or the composition as herein-discussed. In an aspect the invention provides a vector comprising: a nucleic acid molecule encoding a guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the direct repeat of the gRNA is modified by

the insertion of distinct RNA sequence(s) that bind(s) to two or more adaptor proteins, and wherein each adaptor protein is associated with one or more functional domains; or, wherein the gRNA is modified to have at least one non-coding functional loop. In an aspect the invention provides vector(s) comprising nucleic acid molecule(s) encoding: non-naturally occurring or engineered CRISPR-Cas complex composition comprising the gRNA herein-discussed, and an RNA targeting enzyme, wherein optionally the RNA targeting enzyme comprises at least one mutation, such that the RNA targeting enzyme has no more than 5% of the nuclease activity of the RNA targeting enzyme not having the at least one mutation, and optionally one or more comprising at least one or more nuclear localization sequences. In an aspect a vector can further comprise regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide RNA (gRNA) and/or the nucleic acid molecule encoding the RNA targeting enzyme and/or the optional nuclear localization sequence(s).

[00349] In one aspect, the invention provides a kit comprising one or more of the components described hereinabove. In some embodiments, the kit comprises a vector system as described above and instructions for using the kit.

[00350] In an aspect the invention provides a method of screening for gain of function (GOF) or loss of function (LOF) or for screening non-coding RNAs or potential regulatory regions (e.g. enhancers, repressors) comprising the cell line of as herein-discussed or cells of the model herein-discussed containing or expressing the RNA targeting enzyme and introducing a composition as herein-discussed into cells of the cell line or model, whereby the gRNA includes either an activator or a repressor, and monitoring for GOF or LOF respectively as to those cells as to which the introduced gRNA includes an activator or as to those cells as to which the introduced gRNA includes a repressor.

[00351] In an aspect the invention provides a library of non-naturally occurring or engineered compositions, each comprising a RNA targeting CRISPR guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target RNA sequence of interest in a cell, an RNA targeting enzyme, wherein the RNA targeting enzyme comprises at least one mutation, such that the RNA targeting enzyme has no more than 5% of the nuclease activity of the RNA targeting enzyme not having the at least one mutation, wherein the gRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains, wherein the composition comprises one or more or two or more adaptor proteins, wherein the each protein is associated with one or more functional domains, and wherein the gRNAs

comprise a genome wide library comprising a plurality of RNA targeting guide RNAs (gRNAs). In an aspect the invention provides a library as herein-discussed, wherein the RNA targeting RNA targeting enzyme has a diminished nuclease activity of at least 97%, or 100% as compare with the RNA targeting enzyme not having the at least one mutation. In an aspect the invention provides a library as herein-discussed, wherein the adaptor protein is a fusion protein comprising the functional domain. In an aspect the invention provides a library as herein discussed, wherein the gRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the one or two or more adaptor proteins. In an aspect the invention provides a library as herein discussed, wherein the one or two or more functional domains are associated with the RNA targeting enzyme. In an aspect the invention provides a library as herein discussed, wherein the cell population of cells is a population of eukaryotic cells. In an aspect the invention provides a library as herein discussed, wherein the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell. In an aspect the invention provides a library as herein discussed, wherein the mammalian cell is a human cell. In an aspect the invention provides a library as herein discussed, wherein the population of cells is a population of embryonic stem (ES) cells.

[00352] In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 100 or more RNA sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 1000 or more RNA sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 20,000 or more sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of the entire transcriptome. In an aspect the invention provides a library as herein discussed, wherein the targeting is of a panel of target sequences focused on a relevant or desirable pathway. In an aspect the invention provides a library as herein discussed, wherein the pathway is an immune pathway. In an aspect the invention provides a library as herein discussed, wherein the pathway is a cell division pathway.

[00353] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a gene with modified expression. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors encoding the components of the system described herein above into a eukaryotic cell, and (b) allowing a CRISPR complex to bind to a target polynucleotide so as to modify expression of a gene, thereby generating a model eukaryotic cell comprising modified gene expression.

[00354] The structural information provided herein allows for interrogation of guide RNA interaction with the target RNA and the RNA targeting enzyme permitting engineering or alteration of guide RNA structure to optimize functionality of the entire RNA targeting CRISPR-Cas system. For example, the guide RNA may be extended, without colliding with the RNA targeting protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[00355] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00356] The skilled person will understand that modifications to the guide RNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide RNA may be modified, by introduction of a distinct RNA sequence(s) 5' of the direct repeat, within the direct repeat, or 3' of the guide sequence.

[00357] The modified guide RNA, the inactivated RNA targeting enzyme (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[00358] Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00359] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR RNA targeting events.

(See, e.g., Piatt et al., Cell (2014), <http://dx.doi.org/10.1016/j.cell.2014.09.014>, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667), which are not believed prior to the present invention or application). For example, the target cell comprises RNA targeting CRISPR enzyme conditionally or inducibly (e.g. in the form of Cre dependent constructs) and/or the adapter protein conditionally or inducibly and, on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of s RNA targeting enzyme expression and/or adaptor expression in the target cell. By applying the teachings and compositions of the current invention with the known method of creating a CRISPR complex, inducible gene expression affected by functional domains are also an aspect of the current invention. Alternatively, the adaptor protein may be provided as a conditional or inducible element with a conditional or inducible s RNA targeting enzyme to provide an effective model for screening purposes, which advantageously only requires minimal design and administration of specific gRNAs for a broad number of applications.

Guide RNA according to the invention comprising a dead guide sequence

[00360] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity (i.e. without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Indeed, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the assay involves synthesizing a CRISPR target RNA and guide RNAs comprising mismatches with the target RNA, combining these with the RNA targeting enzyme and analyzing cleavage based on gels based on the presence of bands generated by cleavage products, and quantifying cleavage based upon relative band intensities.

[00361] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition RNA targeting CRISPR-Cas system comprising a functional RNA targeting as described herein, and guide RNA (gRNA) wherein the gRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the RNA targeting CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable RNA cleavage activity of a non-mutant RNA targeting enzyme of the

system.. It is to be understood that any of the gRNAs according to the invention as described herein elsewhere may be used as dead gRNAs / gRNAs comprising a dead guide sequence as described herein below. Any of the methods, products, compositions and uses as described herein elsewhere is equally applicable with the dead gRNAs / gRNAs comprising a dead guide sequence as further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[00362] The ability of a dead guide sequence to direct sequence-specific binding of a CRISPR complex to an RNA target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the dead guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence. For instance, cleavage of a target RNA polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the dead guide sequence to be tested and a control guide sequence different from the test dead guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A dead guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell.

[00363] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences are typically shorter than respective guide sequences which result in active RNA cleavage. In particular embodiments, dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same.

[00364] As explained below and known in the art, one aspect of gRNA - RNA targeting specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the RNA targeting enzyme. Thus, structural data available for validated dead guide sequences may be used for designing Cas13 specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains HEPN of two or more Cas13 effector proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such Cas13 specific

equivalents, allowing for formation of the CRISPR complex and successful binding to the target RNA, while at the same time, not allowing for successful nuclease activity.

[00365] The use of dead guides in the context herein as well as the state of the art provides a surprising and unexpected platform for network biology and/or systems biology in both in vitro, ex vivo, and in vivo applications, allowing for multiplex gene targeting, and in particular bidirectional multiplex gene targeting. Prior to the use of dead guides, addressing multiple targets has been challenging and in some cases not possible. With the use of dead guides, multiple targets, and thus multiple activities, may be addressed, for example, in the same cell, in the same animal, or in the same patient. Such multiplexing may occur at the same time or staggered for a desired timeframe.

[00366] For example, the dead guides allow to use gRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA comprising a dead guide may be modified to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA comprising a dead guide to incorporate protein-interacting aptamers (Koner mann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble multiple distinct effector domains. Such may be modeled after natural processes.

[00367] Thus, one aspect is a gRNA of the invention which comprises a dead guide, wherein the gRNA further comprises modifications which provide for gene activation or repression, as described herein. The dead gRNA may comprise one or more aptamers. The aptamers may be specific to gene effectors, gene activators or gene repressors. Alternatively, the aptamers may be specific to a protein which in turn is specific to and recruits / binds a specific gene effector, gene activator or gene repressor. If there are multiple sites for activator or repressor recruitment, it is preferred that the sites are specific to either activators or repressors. If there are multiple sites for activator or repressor binding, the sites may be specific to the same activators or same repressors. The sites may also be specific to different activators or different repressors. The effectors, activators, repressors may be present in the form of fusion proteins.

[00368] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized CRISPR system to a gene locus in an

organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the 20 nt sequence downstream of each CRISPR motif by: i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the first 15 nt of the sequence in the genome of the organism; c) selecting the sequence for use in a guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected if the GC content is 50% or less. In an embodiment, the sequence is selected if the GC content is 40% or less. In an embodiment, the sequence is selected if the GC content is 30% or less. In an embodiment, two or more sequences are analyzed and the sequence having the lowest GC content is selected. In an embodiment, off-target matches are determined in regulatory sequences of the organism. In an embodiment, the gene locus is a regulatory region. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[00369] In an aspect, the invention provides a dead guide RNA for targeting a functionalized CRISPR system to a gene locus in an organism. In an embodiment of the invention, the dead guide RNA comprises a targeting sequence wherein the CG content of the target sequence is 70% or less, and the first 15 nt of the targeting sequence does not match an off-target sequence downstream from a CRISPR motif in the regulatory sequence of another gene locus in the organism. In certain embodiments, the GC content of the targeting sequence 60% or less, 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In certain embodiments, the GC content of the targeting sequence is from 70% to 60% or from 60% to 50% or from 50% to 40% or from 40% to 30%. In an embodiment, the targeting sequence has the lowest CG content among potential targeting sequences of the locus.

[00370] In an embodiment of the invention, the first 15 nt of the dead guide match the target sequence. In another embodiment, first 14 nt of the dead guide match the target sequence. In another embodiment, the first 13 nt of the dead guide match the target sequence. In another embodiment first 12 nt of the dead guide match the target sequence. In another embodiment, first 11 nt of the dead guide match the target sequence. In another embodiment, the first 10 nt of the dead guide match the target sequence. In an embodiment of the invention the first 15 nt of the dead guide does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 14 nt, or the first 13 nt of the dead guide, or the first 12 nt of the guide, or the first 11 nt of the dead guide, or the first 10 nt of the dead guide, does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus.

In other embodiments, the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt of the dead guide do not match an off-target sequence downstream from a CRISPR motif in the genome.

[00371] In certain embodiments, the dead guide RNA includes additional nucleotides at the 3'-end that do not match the target sequence. Thus, a dead guide RNA that includes the first 20-28 nt, downstream of a CRISPR motif can be extended in length at the 3' end.

General provisions

[00372] In an aspect, the invention provides a nucleic acid binding system. In situ hybridization of RNA with complementary probes is a powerful technique. Typically fluorescent DNA oligonucleotides are used to detect nucleic acids by hybridization. Increased efficiency has been attained by certain modifications, such as locked nucleic acids (LNAs), but there remains a need for efficient and versatile alternatives. The invention provides an efficient and adaptable system for in situ hybridization.

[00373] In embodiments of the invention the terms guide sequence and guide RNA are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. Preferably the guide sequence is 10 - 30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target

sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome.

[00374] In general, and throughout this specification, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00375] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be

selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[00376] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and HI promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced

into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.).

[00377] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00378] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence.

[00379] In certain embodiments, the CRISPR system as provided herein can make use of a crRNA or analogous polynucleotide comprising a guide sequence, wherein the polynucleotide is an RNA, a DNA or a mixture of RNA and DNA, and/or wherein the polynucleotide comprises one or more nucleotide analogs. The sequence can comprise any structure, including but not limited to a structure of a native crRNA, such as a bulge, a hairpin or a stem loop structure. In certain embodiments, the polynucleotide comprising the guide sequence forms a duplex with a second polynucleotide sequence which can be an RNA or a DNA sequence.

[00380] In certain embodiments, guides of the invention comprise non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, boranophosphate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2' and 4' carbons of the ribose ring, or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited

to, 2-aminopurine, 5-bromo-uridine, pseudouridine (Ψ), N⁶-methylpseudouridine (rae Ψ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2'-**O**-methyl (M), 2'-**O**-methyl 3'-phosphorothioate (MS), S-constrained ethyl (cEt), or 2'-**O**-methyl 3'-thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guide RNAs can comprise increased stability and increased activity as compared to unmodified guide RNAs, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, *Nat Biotechnol.* 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Allerson et al., *J. Med. Chem.* 2005, 48:901-904; Bramsen et al., *Front. Genet.*, 2012, 3:154; Deng et al., *PNAS*, 2015, 112:1 1870-1 1875; Sharma et al., *MedChemComm.*, 2014, 5:1454-1471; Li et al., *Nature Biomedical Engineering*, 2017, 1, 0066 DOI:10.1038/s4 1551-017-0066).

[00381] In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, *J. Biotech.* 233:74-83). In certain embodiments, a guide comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to Cas9, Cpf1, or C2c1. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, 5' and/or 3' end, stem-loop regions, and the seed region. In certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., *Nature Biomedical Engineering*, 2017, 1:0066). In certain embodiments, 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, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-**O**-methyl (M), 2'-**O**-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), or 2'-**O**-methyl-3'-thioPACE (MSP). Such modification can enhance genome editing efficiency (see Hendel et al., *Nat. Biotechnol.* (2015) 33(9): 985-989). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-**O**-Me, 2'-F or S-constrained

ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E71 10-E71 11). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine. In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., eLife, 2017, 6:e25312, DOI: 10.7554)

[00382] In one aspect of the invention, the guide comprises a modified crRNA for Cpf1, having a 5'-handle and a guide segment further comprising a seed region and a 3'-terminus. In some embodiments, the modified guide can be used with a Cpf1 of any one of *Acidaminococcus* sp. BV3L6 Cpf1 (AsCpf1); *Francisella tularensis* subsp. Novicida U 112 Cpf1 (FnCpf1); *L. bacterium* MC2017 Cpf1 (Lb3Cpf1); *Butyrivibrio proteoclasticus* Cpf1 (BpCpf1); *Parcubacteria bacterium* GWC201 1_GWC2_44_17 Cpf1 (PbCpf1); *Peregrinibacteria bacterium* GW201 1_GWA_33_10 Cpf1 (PeCpf1); *Leptospira inadai* Cpf1 (LiCpf1); *Smithella* sp. SC_K08D17 Cpf1 (SsCpf1); *L. bacterium* MA2020 Cpf1 (Lb2Cpf1); *Porphyromonas crevioricanis* Cpf1 (PeCpf1); *Porphyromonas macacae* Cpf1 (PmCpf1); *Candidatus Methanoplasma termitum* Cpf1 (CMtCpf1); *Eubacterium eligens* Cpf1 (EeCpf1); *Moraxella bovoculi* 237 Cpf1 (MbCpf1); *Prevotella disiens* Cpf1 (PdCpf1); or *L. bacterium* ND2006 Cpf1 (LbCpf1).

[00383] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (Ψ), N1-methylpseudouridine (meI Ψ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), or 2'-O-methyl-3'-thioPACE (MSP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the

chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 or 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpf1 CrRNA improve gene cutting efficiency (see Li, et al., *Nature Biomedical Engineering*, 2017, 1:0066). In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'- O-methyl (M) analogs.

[00384] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU.

[00385] In one aspect, the guide comprises portions that are chemically linked or conjugated via a non-phosphodiester bond. In one aspect, the guide comprises, in non-limiting examples, a tracr sequence and a tracr mate sequence portion or a direct repeat and a targeting sequence portion that are chemically linked or conjugated via a non-nucleotide loop. In some embodiments, the portions are joined via a non-phosphodiester covalent linker. Examples of the covalent linker include but are not limited to a chemical moiety selected from the group consisting of carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[00386] In some embodiments, portions of the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., *Methods in Molecular Biology* Col 288, *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, New Jersey (2012)). In some embodiments, the non-targeting guide portions can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., *Bioconjugate Techniques*, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide,

haloalkyl, sulfonyl, ally, propargyl, diene, alkyne, and azide. Once a non-targeting portions of a guide is functionalized, a covalent chemical bond or linkage can be formed between the two oligonucleotides. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[00387] In some embodiments, one or more portions of a guide can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., *J. Am. Chem. Soc.* (1998) 120: 11820-1 1821; Scaringe, *Methods Enzymol.* (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., *J. Am. Chem. Soc.* (2011) 133: 11540-1 1546; Hendel et al., *Nat. Biotechnol.* (2015) 33:985-989).

[00388] In some embodiments, the guide portions can be covalently linked using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., *Angew. Chem. Int. Ed.* (2009) 48:6974-6998; Manoharan, M. *Curr. Opin. Chem. Biol.* (2004) 8: 570-9; Behlke et al., *Oligonucleotides* (2008) 18: 305-19; Watts, et al., *Drug. Discov. Today* (2008) 13: 842-55; Shukla, et al., *ChemMedChem* (2010) 5: 328-49.

[00389] In some embodiments, the guide portions can be covalently linked using click chemistry. In some embodiments, guide portions can be covalently linked using a triazole linker. In some embodiments, guide portions can be covalently linked using Huisgen 1,3-dipolar cycloaddition reaction involving an alkyne and azide to yield a highly stable triazole linker (He et al., *ChemBioChem* (2015) 17: 1809-1812; WO 2016/186745). In some embodiments, guide portions are covalently linked by ligating a 5'-hexyne portion and a 3'-azide portion. In some embodiments, either or both of the 5'-hexyne guide portion and a 3'-azide guide portion can be protected with 2'-acetoxyethyl orthoester (2'-ACE) group, which can be subsequently removed using Dharmacon protocol (Scaringe et al., *J. Am. Chem. Soc.* (1998) 120: 11820-1 1821; Scaringe, *Methods Enzymol.* (2000) 317: 3-18).

[00390] In some embodiments, guide portions can be covalently linked via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are

not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of ethylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiesteres, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[00391] The linker (e.g., a non-nucleotide loop) can be of any length. In some embodiments, the linker has a length equivalent to about 0-16 nucleotides. In some embodiments, the linker has a length equivalent to about 0-8 nucleotides. In some embodiments, the linker has a length equivalent to about 0-4 nucleotides. In some embodiments, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in WO201 1/008730.

[00392] In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid -targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid -targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay

as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid - targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[00393] In some embodiments, a nucleic acid-targeting guide RNA is selected to reduce the degree secondary structure within the RNA-targeting guide RNA. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[00394] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In

certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[00395] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[00396] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides, preferably at least 18 nt, such as at least 19, 20, 21, 22, or more nt. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[00397] Applicants also perform a challenge experiment to verify the RNA targeting and cleaving capability of a Cas13. This experiment closely parallels similar work in *E. coli* for the heterologous expression of StCas9 (Sapranauskas, R. et al. *Nucleic Acids Res* 39, 9275-9282 (2011)). Applicants introduce a plasmid containing both a PAM and a resistance gene into the heterologous *E. coli*, and then plate on the corresponding antibiotic. If there is RNA cleavage of the plasmid transcribed resistance gene, Applicants observe no viable colonies.

[00398] In further detail, the assay is as follows for a DNA target, but may be adapted accordingly for an RNA target. Two *E. coli* strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g. pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the DNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned. One has a 8 random bp 5' of the proto-spacer (e.g. total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g. total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains were harvested and plasmid DNA was

isolated. Plasmid DNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransformed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM.

[00399] For minimization of toxicity and off-target effect, it will be important to control the concentration of nucleic acid-targeting guide RNA delivered. Optimal concentrations of nucleic acid -targeting guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci. The concentration that gives the highest level of on-target modification while minimizing the level of off-target modification should be chosen for *in vivo* delivery. The nucleic acid-targeting system is derived advantageously from a Type VI CRISPR system. In some embodiments, one or more elements of a nucleic acid-targeting system is derived from a particular organism comprising an endogenous RNA-targeting system. In particular embodiments, the Type VI RNA-targeting Cas enzyme is Cas13. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Cse1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In embodiments, the Type VI protein such as Cas13 as referred to herein also encompasses a homologue or an orthologue of a Type VI protein such as Cas13. The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. In particular embodiments, the homologue or orthologue of a Type VI protein such as Cas13 as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with a Type VI protein such as Cas13. In further

embodiments, the homologue or orthologue of a Type VI protein such as Cas13 as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type Type VI protein such as Cas13.

[00400] In an embodiment, the Type VI RNA-targeting Cas protein may be a Cas13 ortholog of an organism of a genus which includes but is not limited to *Leptotrichia*, *Listeria*, *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* and *Campylobacter*. Species of organism of such a genus can be as otherwise herein discussed.

[00401] Some methods of identifying orthologs of CRISPR-Cas system enzymes may involve identifying tracr sequences in genomes of interest. Identification of tracr sequences may relate to the following steps: Search for the direct repeats or tracr mate sequences in a database to identify a CRISPR region comprising a CRISPR enzyme. Search for homologous sequences in the CRISPR region flanking the CRISPR enzyme in both the sense and antisense directions. Look for transcriptional terminators and secondary structures. Identify any sequence that is not a direct repeat or a tracr mate sequence but has more than 50% identity to the direct repeat or tracr mate sequence as a potential tracr sequence. Take the potential tracr sequence and analyze for transcriptional terminator sequences associated therewith.

[00402] It will be appreciated that any of the functionalities described herein may be engineered into CRISPR enzymes from other orthologs, including chimeric enzymes comprising fragments from multiple orthologs. Examples of such orthologs are described elsewhere herein. Thus, chimeric enzymes may comprise fragments of CRISPR enzyme orthologs of an organism which includes but is not limited to *Leptotrichia*, *Listeria*, *Corynebacter*, *Sutterella*, *Legionella*, *Treponema*, *Filifactor*, *Eubacterium*, *Streptococcus*, *Lactobacillus*, *Mycoplasma*, *Bacteroides*, *Flaviivola*, *Flavobacterium*, *Sphaerochaeta*, *Azospirillum*, *Gluconacetobacter*, *Neisseria*, *Roseburia*, *Parvibaculum*, *Staphylococcus*, *Nitratifactor*, *Mycoplasma* and *Campylobacter*. A chimeric enzyme can comprise a first fragment and a second fragment, and the fragments can be of CRISPR enzyme orthologs of organisms of genera herein mentioned or of species herein mentioned; advantageously the fragments are from CRISPR enzyme orthologs of different species.

[00403] In embodiments, the Type VI RNA-targeting effector protein, in particular the Cas13 protein as referred to herein also encompasses a functional variant of Cas13 or a

homologue or an orthologue thereof. A "functional variant" of a protein as used herein refers to a variant of such protein which retains at least partially the activity of that protein. Functional variants may include mutants (which may be insertion, deletion, or replacement mutants), including polymorphs, etc. Also included within functional variants are fusion products of such protein with another, usually unrelated, nucleic acid, protein, polypeptide or peptide. Functional variants may be naturally occurring or may be man-made. Advantageous embodiments can involve engineered or non-naturally occurring Type VI RNA-targeting effector protein.

[00404] In an embodiment of the invention, there is provided a neffector protein which comprises an amino acid sequence having at least 80% sequence homology to the wild-type sequence of any of *Leptotrichia shahii* CasI3, *Lachnospiraceae* bacterium MA2020 CasI3, *Lachnospiraceae* bacterium NK4A179 CasI3, *Clostridium aminophilum* (DSM 10710) CasI3, *Carnobacterium gallinarum* (DSM 4847) CasI3, *Paludibacter propionigenes* (WB4) CasI3, *Listeria weihenstephanensis* (FSL R9-0317) CasI3, *Listeriaceae* bacterium (FSL M6-0635) CasI3, *Listeria newyorkensis* (FSL M6-0635) CasI3, *Leptotrichia wadei* (F0279) CasI3, *Rhodobacter capsulatus* (SB 1003) CasI3, *Rhodobacter capsulatus* (R121) CasI3, *Rhodobacter capsulatus* (DE442) CasI3, *Leptotrichia wadei* (Lw2) CasI3, or *Listeria seeligeri* CasI3.

[00405] In an embodiment of the invention, the effector protein comprises an amino acid sequence having at least 80% sequence homology to a Type VI effector protein consensus sequence including but not limited to a consensus sequence described herein.

[00406] In an embodiment of the invention, the effector protein comprises at least one HEPN domain, including but not limited to HEPN domains described herein, HEPN domains known in the art, and domains recognized to be HEPN domains by comparison to consensus sequences and motifs. Several such domains are provided herein. In one non-limiting example, a consensus sequence can be derived from the sequences of CasI3 orthologs provided herein.

[00407] In an embodiment of the invention, the effector protein comprises one or more HEPN domains comprising a RxxxxH motif sequence. The RxxxxH motif sequence can be, without limitation, from an HEPN domain described herein or an HEPN domain known in the art. RxxxxH motifs sequences further include motif sequences created by combining portions of two or more HEPN domains. As noted, consensus sequences can be derived from the sequences of the 15 orthologs disclosed in US 62/432,240 (BI-10035). For example, from the above sequence alignment, the first HEPN domain comprises a R{N/H}xxxH motif whereas the second HEPN domain comprises a R{N/K}xxxH motif.

[00408] In an embodiment of the invention, a HEPN domain comprises at least one RxxxxH motif comprising the sequence of R{N/H/K}XiX₂X₃H. In an embodiment of the invention, a HEPN domain comprises a RxxxxH motif comprising the sequence of R{N/H}XiX₂X₃H. In an embodiment of the invention, a HEPN domain comprises the sequence of R{N/K}XiX₂X₃H. In certain embodiments, Xi is R, S, D, E, Q, N, G, Y, or H. In certain embodiments, X₂ is I, S, T, V, or L. In certain embodiments, X₃ is L, F, N, Y, V, I, S, D, E, or A.

[00409] Additional effectors for use according to the invention can be identified by their proximity to casl genes, for example, though not limited to, within the region 20 kb from the start of the casl gene and 20 kb from the end of the casl gene. In certain embodimenst, the effector protein comprises at least one HEPN domain and at least 500 amino acids, and wherein the Casl3 effector protein is naturally present in a prokaryotic genome within 20 kb upstream or downstream of a Casl gene or a CRISPR array.

[00410] In an embodiment, nucleic acid molecule(s) encoding the Type VI RNA-targeting effector protein, in particular Cas 13 or an ortholog or homolog thereof, may be codon-optimized for expression in an eukaryotic cell. A eukaryote can be as herein discussed. Nucleic acid molecule(s) can be engineered or non-naturally occurring.

[00411] In an embodiment, the Type VI RNA-targeting effector protein, in particular Casl3 or an ortholog or homolog thereof, may comprise one or more mutations (and hence nucleic acid molecule(s) coding for same may have mutation(s)). The mutations may be artificially introduced mutations and may include but are not limited to one or more mutations in a catalytic domain. Examples of catalytic domains with reference to a Cas9 enzyme may include but are not limited to RuvC I, RuvC II, RuvC III and HNH domains.

[00412] In an embodiment, the Type VI protein such as Casl3 or an ortholog or homolog thereof, may comprise one or more mutations. The mutations may be artificially introduced mutations and may include but are not limited to one or more mutations in a catalytic domain. Examples of catalytic domains with reference to a Cas enzyme may include but are not limited to HEPN domains.

[00413] In an embodiment, the Type VI protein such as Casl3 or an ortholog or homolog thereof, may be used as a generic nucleic acid binding protein with fusion to or being operably linked to a functional domain. Exemplary functional domains may include but are not limited to translational initiator, translational activator, translational repressor, nucleases, in particular ribonucleases, a spliceosome, beads, a light inducible/controllable domain or a chemically inducible/controllable domain.

[00414] In some embodiments, the unmodified nucleic acid-targeting effector protein may have cleavage activity. In some embodiments, the RNA-targeting effector protein may direct cleavage of one or both nucleic acid (DNA or RNA) strands at the location of or near a target sequence, such as within the target sequence and/or within the complement of the target sequence or at sequences associated with the target sequence. In some embodiments, the nucleic acid -targeting Cas protein may direct cleavage of one or both DNA or RNA strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a nucleic acid -targeting Cas protein that may be mutated with respect to a corresponding wild-type enzyme such that the mutated nucleic acid -targeting Cas protein lacks the ability to cleave RNA strands of a target polynucleotide containing a target sequence. As a further example, two or more catalytic domains of Cas (e.g. HEPN domain) may be mutated to produce a mutated Cas substantially lacking all RNA cleavage activity. In some embodiments, a nucleic acid -targeting effector protein may be considered to substantially lack all RNA cleavage activity when the RNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. An effector protein may be identified with reference to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the Type VI CRISPR system. Most preferably, the effector protein is a Type VI protein such as Cas13. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[00415] Again, it will be appreciated that the terms Cas and CRISPR enzyme and CRISPR protein and Cas protein are generally used interchangeably and at all points of reference herein refer by analogy to novel CRISPR effector proteins further described in this application, unless otherwise apparent, such as by specific reference to Cas9. As mentioned above, many of the residue numberings used herein refer to the effector protein from the Type VI CRISPR locus. However, it will be appreciated that this invention includes many more effector proteins from other species of microbes. In certain embodiments, Cas may be constitutively present or inducibly present or conditionally present or administered or delivered. Cas optimization may be used to enhance function or to develop new functions,

one can generate chimeric Cas proteins. And Cas may be used as a generic nucleic acid binding protein.

[00416] Typically, in the context of an endogenous nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in cleavage of one or both DNA or RNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, or more base pairs from) the target sequence. As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 200, 500, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest).

[00417] An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US20 13/074667) as an example of a codon optimized sequence (from knowledge in the art and this disclosure, codon optimizing coding nucleic acid molecule(s), especially as to effector protein (e.g., Cas13) is within the ambit of the skilled artisan).. Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a DNA/RNA-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in

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

[00418] In some embodiments, a vector encodes a nucleic acid-targeting effector protein such as the Cas13, or an ortholog or homolog thereof comprising one or more nuclear localization sequences (NLSs) or nuclear export sequences (NESs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs. In some embodiments, the RNA-targeting effector protein comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs or NESs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS or NES at the amino-terminus and zero or at one or more NLS or NES at the carboxy terminus). When more than one NLS or NES is present, each may be selected independently of the others, such that a single NLS or NES may be present in more than one copy and/or in combination with one or more other NLSs or NESs present in one or more copies. In some embodiments, an NLS or NES is considered near the N- or C-terminus when the nearest amino acid of the NLS or NES is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV; the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK); the c-myc NLS having the amino acid sequence PAAKRVKLD or RQRRNELKRSP; the hRNPA1 M9 NLS having the sequence

NQS SNFGPMKGGNFGGRSSGPYGGGGQYF AKPRNQGGY; the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV of the IBB domain from importin-alpha; the sequences VSRKRPRP and PPKKARED of the myoma T protein; the sequence POPKKKPL of human p53; the sequence SALIKKKKKMAP of mouse c-abl IV; the sequences DRLRR and PKQKKRK of the influenza virus NS1; the sequence RKLKKKIKKL of the Hepatitis virus delta antigen; the sequence REKKKFLKRR of the mouse Mx1 protein; the sequence KRKGDEV DGVDEVAKKKSKK of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs or NESs are of sufficient strength to drive accumulation of the DNA/RNA-targeting Cas protein in a detectable amount in respectively the nucleus or cytoplasm of a eukaryotic cell. In general, strength of nuclear/cytoplasmic localization activity may derive from the number of NLSs or NESs in the nucleic acid-targeting effector protein, the particular NLS(s) or NES(s) used, or a combination of these factors. Detection of accumulation in the nucleus/cytoplasm may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI) or cytoplasm. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus/cytoplasm may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for RNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by RNA-targeting complex formation and/or RNA-targeting Cas protein activity), as compared to a control not exposed to the nucleic acid-targeting Cas protein or nucleic acid-targeting complex, or exposed to a nucleic acid-targeting Cas protein lacking the one or more NLSs or NESs. In preferred embodiments of the herein described Cas13 effector protein complexes and systems the codon optimized Cas13 effector proteins comprise an NLS or NES attached to the C-terminal of the protein.

[00419] In some embodiments, one or more vectors driving expression of one or more elements of a nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct formation of a nucleic acid-targeting complex at one or more target sites. For example, a nucleic acid-targeting effector enzyme and a nucleic acid-targeting guide RNA could each be operably linked to

separate regulatory elements on separate vectors. RNA(s) of the nucleic acid-targeting system can be delivered to a transgenic nucleic acid-targeting effector protein animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses nucleic acid-targeting effector protein; or an animal or mammal that is otherwise expressing nucleic acid-targeting effector protein or has cells containing nucleic acid-targeting effector protein, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo nucleic acid-targeting effector protein. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the nucleic acid-targeting system not included in the first vector. nucleic acid-targeting system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA may be operably linked to and expressed from the same promoter. Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a nucleic acid-targeting system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US20 13/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises two or more insertion sites, so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or

more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a nucleic acid-targeting effector protein. nucleic acid-targeting effector protein or nucleic acid-targeting guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a particle or nanoparticle complex. nucleic acid-targeting effector protein mRNA can be delivered prior to the nucleic acid-targeting guide RNA to give time for nucleic acid-targeting effector protein to be expressed. nucleic acid-targeting effector protein mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of nucleic acid-targeting guide RNA. Alternatively, nucleic acid-targeting effector protein mRNA and nucleic acid-targeting guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of nucleic acid-targeting effector protein mRNA + guide RNA. Additional administrations of nucleic acid-targeting effector protein mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome and/or transcriptome modification.

[00420] In one aspect, the invention provides methods for using one or more elements of a nucleic acid-targeting system. The nucleic acid-targeting complex of the invention provides an effective means for modifying a target RNA. The nucleic acid-targeting complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target RNA in a multiplicity of cell types. As such the nucleic acid-targeting complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary nucleic acid-targeting complex comprises a RNA-targeting effector protein complexed with a guide RNA hybridized to a target sequence within the target locus of interest.

[00421] In one embodiment, this invention provides a method of cleaving a target RNA. The method may comprise modifying a target RNA using a nucleic acid-targeting complex that binds to the target RNA and effect cleavage of said target RNA. In an embodiment, the nucleic acid-targeting complex of the invention, when introduced into a cell, may create a break (e.g., a single or a double strand break) in the RNA sequence. For example, the method can be used to cleave a disease RNA in a cell. For example, an exogenous RNA template comprising a sequence to be integrated flanked by an upstream sequence and a downstream sequence may be introduced into a cell. The upstream and downstream sequences share sequence similarity with either side of the site of integration in the RNA.

Where desired, a donor RNA can be mRNA. The exogenous RNA template comprises a sequence to be integrated (e.g., a mutated RNA). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include RNA encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function. The upstream and downstream sequences in the exogenous RNA template are selected to promote recombination between the RNA sequence of interest and the donor RNA. The upstream sequence is a RNA sequence that shares sequence similarity with the RNA sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a RNA sequence that shares sequence similarity with the RNA sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous RNA template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted RNA sequence. Preferably, the upstream and downstream sequences in the exogenous RNA template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted RNA sequence. In some methods, the upstream and downstream sequences in the exogenous RNA template have about 99% or 100% sequence identity with the targeted RNA sequence. An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp. In some methods, the exogenous RNA template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous RNA template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996). In a method for modifying a target RNA by integrating an exogenous RNA template, a break (e.g., double or single stranded break in double or single stranded DNA or RNA) is introduced into the DNA or RNA sequence by the nucleic acid-targeting complex, the break is repaired via homologous recombination with an exogenous RNA template such that the template is integrated into the RNA target. The presence of a double-stranded break facilitates integration of the template. In other embodiments, this invention provides a method of modifying expression of a RNA in a eukaryotic cell. The method comprises increasing or decreasing expression of a target

polynucleotide by using a nucleic acid-targeting complex that binds to the RNA (e.g., mRNA or pre-mRNA). In some methods, a target RNA can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a RNA-targeting complex to a target sequence in a cell, the target RNA is inactivated such that the sequence is not translated, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein or microRNA or pre-microRNA transcript is not produced. The target RNA of a RNA-targeting complex can be any RNA endogenous or exogenous to the eukaryotic cell. For example, the target RNA can be a RNA residing in the nucleus of the eukaryotic cell. The target RNA can be a sequence (e.g., mRNA or pre-mRNA) coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA). Examples of target RNA include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated RNA. Examples of target RNA include a disease associated RNA. A "disease-associated" RNA refers to any RNA which is yielding translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a RNA transcribed from a gene that becomes expressed at an abnormally high level; it may be a RNA transcribed from a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated RNA also refers to a RNA transcribed from a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The translated products may be known or unknown, and may be at a normal or abnormal level. The target RNA of a RNA-targeting complex can be any RNA endogenous or exogenous to the eukaryotic cell. For example, the target RNA can be a RNA residing in the nucleus of the eukaryotic cell. The target RNA can be a sequence (e.g., mRNA or pre-mRNA) coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA).

[00422] In some embodiments, the method may comprise allowing a nucleic acid-targeting complex to bind to the target RNA to effect cleavage of said target RNA or RNA thereby modifying the target RNA, wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said target RNA. In one aspect, the invention provides a method of modifying expression of RNA in a eukaryotic cell. In some embodiments, the method comprises allowing a nucleic acid-targeting complex to bind to the RNA such that said binding results

in increased or decreased expression of said RNA; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA. Similar considerations and conditions apply as above for methods of modifying a target RNA. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In one aspect, the invention provides for methods of modifying a target RNA in a eukaryotic cell, which may be *in vivo*, *ex vivo* or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[00423] Indeed, in any aspect of the invention, the nucleic acid-targeting complex may comprise a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence.

[00424] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving RNA sequence targeting, that relate to the nucleic acid-targeting system and components thereof. In advantageous embodiments, the effector protein/enzyme is a Type VI protein such as Cas13. An advantage of the present methods is that the CRISPR system minimizes or avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target RNA.

[00425] In relation to a nucleic acid-targeting complex or system preferably, the tracr sequence has one or more hairpins and is 30 or more nucleotides in length, 40 or more nucleotides in length, or 50 or more nucleotides in length; the crRNA sequence is between 10 to 30 nucleotides in length, the nucleic acid-targeting effector protein is a Type VI effector protein.

[00426] In certain embodiments, the effector protein may be a *Listeria* sp. Cas13p, preferably *Listeria seeligeria* Cas13p, more preferably *Listeria seeligeria* serovar 1/2b str. SLCC3954 Cas13p and the crRNA sequence may be 44 to 47 nucleotides in length, with a 5' 29-nt direct repeat (DR) and a 15-nt to 18-nt spacer.

[00427] In certain embodiments, the effector protein may be a *Leptotrichia* sp. Cas13p, preferably *Leptotrichia shahii* Cas13p, more preferably *Leptotrichia shahii* DSM 19757 Cas13p and the crRNA sequence may be 42 to 58 nucleotides in length, with a 5' direct repeat of at least 24 nt, such as a 5' 24-28-nt direct repeat (DR) and a spacer of at least 14 nt, such

as a 14-nt to 28-nt spacer, or a spacer of at least 18 nt, such as 19, 20, 21, 22, or more nt, such as 18-28, 19-28, 20-28, 21-28, or 22-28 nt.

[00428] More preferably, the effector protein may be a *Leptotrichia* sp., preferably *Leptotrichia wadei* F0279, or a *Listeria* sp., preferably *Listeria newyorkensis* FSL M6-0635.

[00429] In certain embodiments, the effector protein may be a Type VI loci effector protein, more particularly a CasI3p, and the crRNA sequence may be 36 to 63 nucleotides in length, preferably 37-nt to 62-nt in length, or 38-nt to 61-nt in length, or 39-nt to 60-nt in length, more preferably 40-nt to 59-nt in length, or 41-nt to 58-nt in length, most preferably 42-nt to 57-nt in length. For example, the crRNA may comprise, consist essentially of or consist of a direct repeat (DR), preferably a 5' DR, 26-nt to 31-nt in length, preferably 27-nt to 30-nt in length, even more preferably 28-nt or 29-nt in length or at least 28 or 29 nt in length, and a spacer 10-nt to 32-nt in length, preferably 11-nt to 31-nt in length, more preferably 12-nt to 30-nt in length, even more preferably 13-nt to 29-nt in length, and most preferably 14-nt to 28-nt in length, such as 18-28 nt, 19-28 nt, 20-28 nt, 21-28 nt, or 22-28 nt.

[00430] In certain embodiments, the effector protein may be a Type VI loci effector protein, more particularly a CasI3p, and the tracrRNA sequence (if present) may be at least 60-nt long, such as at least 65-nt in length, or at least 70-nt in length, such as from 60-nt to 70-nt in length, or from 60-nt to 70-nt in length, or from 70-nt to 80-nt in length, or from 80-nt to 90-nt in length, or from 90-nt to 100-nt in length, or from 100-nt to 110-nt in length, or from 110-nt to 120-nt in length, or from 120-nt to 130-nt in length, or from 130-nt to 140-nt in length, or from 140-nt to 150-nt in length, or more than 150-nt in length.

[00431] In certain embodiments, the effector protein may be a Type VI loci effector protein, more particularly a CasI3p, and no tracrRNA may be required for cleavage.

[00432] The use of two different aptamers (each associated with a distinct nucleic acid-targeting guide RNAs) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different nucleic acid-targeting guide RNAs, to activate expression of one DNA or RNA, whilst repressing another. They, along with their different guide RNAs can be administered together, or substantially together, in a multiplexed approach. A large number of such modified nucleic acid-targeting guide RNAs can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of effector protein molecules need to be delivered, as a comparatively small number of effector protein molecules can be used with a large number modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first

activator and a second activator. The first and second activators may be the same, but they are preferably different activators. Three or more or even four or more activators (or repressors) may be used, but package size may limit the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

[00433] It is also envisaged that the nucleic acid-targeting effector protein-guide RNA complex as a whole may be associated with two or more functional domains. For example, there may be two or more functional domains associated with the nucleic acid-targeting effector protein, or there may be two or more functional domains associated with the guide RNA (via one or more adaptor proteins), or there may be one or more functional domains associated with the nucleic acid-targeting effector protein and one or more functional domains associated with the guide RNA (via one or more adaptor proteins).

[00434] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 ((GGGS)₃) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the guide RNAs and the functional domain (activator or repressor), or between the nucleic acid-targeting effector protein and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

[00435] The invention comprehends a nucleic acid-targeting complex comprising a nucleic acid-targeting effector protein and a guide RNA, wherein the nucleic acid-targeting effector protein comprises at least one mutation, such that the nucleic acid-targeting Cas protein has no more than 5% of the activity of the nucleic acid-targeting Cas protein not having the at least one mutation and, optionally, at least one or more nuclear localization sequences; the guide RNA comprises a guide sequence capable of hybridizing to a target sequence in a RNA of interest in a cell; and wherein: the nucleic acid-targeting effector protein is associated with two or more functional domains; or at least one loop of the guide RNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with two or more functional domains; or the nucleic acid-targeting effector protein is associated with one or more functional domains and at least one loop of the guide RNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains.

Delivery generally**Cas13 Effector Protein Complexes Can Deliver Functional Effectors**

[00436] Unlike CRISPR-Cas-mediated gene knockout, which permanently eliminates expression by mutating the gene at the DNA level, CRISPR-Cas knockdown allows for temporary reduction of gene expression through the use of artificial transcription or translation factors. Mutating key residues in both DNA or RNA cleavage domains of the Cas13 protein results in the generation of a catalytically inactive Cas13. A catalytically inactive Cas13 complexes with a guide RNA and localizes to the or RNA sequence specified by that guide RNA's targeting domain, however, it does not cleave the target RNA. Fusion of the inactive Cas13 protein to an effector domain, e.g., a transcription or translation repression domain, enables recruitment of the effector to any or RNA site specified by the guide RNA. In certain embodiments, Cas13 may be fused to a transcriptional repression domain and recruited to the promoter region of a gene. Especially for gene repression, it is contemplated herein that blocking the binding site of an endogenous transcription factor would aid in downregulating gene expression. In another embodiment, an inactive Cas13 can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene. In further embodiments, Cas13 may be fused to a translation repression domain.

[00437] In an embodiment, a guide RNA molecule can be targeted to a known transcription response element(s) (e.g., promoters, enhancers, etc.), a known upstream activating sequences, and/or sequences of unknown or known function that are suspected of being able to control (protein) expression of the target RNA.

[00438] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced. In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to an RNA target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not translated, affecting the expression level of the protein in the cell.

[00439] In particular embodiments, the CRISPR enzyme comprises one or more mutations selected from the group consisting of R597A, H602A, R1278A and H1283A and/or the one

or more mutations are in the HEPN domain of the CRISPR enzyme or is a mutation as otherwise discussed herein. In some embodiments, the CRISPR enzyme has one or more mutations in a catalytic domain, wherein when transcribed, the direct repeat sequence forms a single stem loop and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the enzyme further comprises a functional domain. In some embodiments, the functional domain is a . In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (eg SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

Delivery of the Cas13 Effector Protein Complex or Components Thereof

[00440] Through this disclosure and the knowledge in the art, TALEs, CRISPR-Cas systems, or components thereof or nucleic acid molecules thereof or nucleic acid molecules encoding or providing components thereof may be delivered by a delivery system herein described both generally and in detail.

[00441] Vector delivery, e.g., plasmid, viral delivery: The CRISPR enzyme, for instance a Type V protein such as Cas13, and/or any of the present RNAs, for instance a guide RNA, can be delivered using any suitable vector, e.g., plasmid or viral vectors, such as adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Effector proteins and one or more guide RNAs can be packaged into one or more vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[00442] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered

saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[00443] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1×10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1×10^6 particles (for example, about 1×10^6 - 1×10^{12} particles), more preferably at least about 1×10^7 particles, more preferably at least about 1×10^8 particles (e.g., about 1×10^8 - 1×10^{11} particles or about 1×10^8 - 1×10^{12} particles), and most preferably at least about 1×10^9 particles (e.g., about 1×10^9 - 1×10^{10} particles or about 1×10^9 - 1×10^{12} particles), or even at least about 1×10^{10} particles (e.g., about 1×10^{10} - 1×10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1×10^{14} particles, preferably no more than about 1×10^{13} particles, even more preferably no more than about 1×10^{12} particles, even more preferably no more than about 1×10^{11} particles, and most preferably no more than about 1×10^{10} particles (e.g., no more than about 1×10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^6 particle units (pu), about 2×10^6 pu, about 4×10^6 pu, about 1×10^7 pu, about 2×10^7 pu, about 4×10^7 pu, about 1×10^8 pu, about 2×10^8 pu, about 4×10^8 pu, about 1×10^9 pu, about 2×10^9 pu, about 4×10^9 pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector.

See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[00444] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1×10^{10} to about 1×10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1×10^5 to 1×10^{50} genomes AAV, from about 1×10^8 to 1×10^{20} genomes AAV, from about 1×10^{10} to about 1×10^{16} genomes, or about 1×10^{11} to about 1×10^{16} genomes AAV. A human dosage may be about 1×10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

[00445] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μ g to about 10 μ g per 70 kg individual. Plasmids of the invention will generally comprise (i) a promoter; (ii) a sequence encoding an nucleic acid-targeting CRISPR enzyme, operably linked to said promoter; (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii). The plasmid can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on a different vector.

[00446] The doses herein are based on an average 70 kg individual. The frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), or scientist skilled in the art. It is also noted that mice used in experiments are typically about 20g and from mice experiments one can scale up to a 70 kg individual.

[00447] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference. Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian

cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[00448] Indeed, RNA delivery is a useful method of *in vivo* delivery. It is possible to deliver nucleic acid-targeting Cas protein/Cas9 and guide RNA/gRNA (and, for instance, HR repair template) into cells using liposomes or particles. Thus delivery of the nucleic acid-targeting Cas protein/CRISPR enzyme, such as a Cas/Cas9 and/or delivery of the guide RNAs of the invention may be in RNA form and via microvesicles, liposomes or particles. For example, Cas mRNA and guide RNA can be packaged into liposomal particles for delivery *in vivo*. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[00449] Means of delivery of RNA also preferred include delivery of RNA via nanoparticles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, *Advanced Functional Materials*, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C, Cortez, C, Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, *Journal of Internal Medicine*, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the RNA-targeting system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA *in vitro* and *in vivo*." *Nat Protoc.* 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA *in vitro* and *in vivo*. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purified and characterized from transfected cell supernatant, then RNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain. Vitamin E (α-tocopherol) may be conjugated with nucleic acid-targeting Cas protein and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (*HUMAN GENE THERAPY* 22:711-719 (June 2011)) for

delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, CA) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of nucleic acid-targeting effector protein conjugated to α -tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μ mol of nucleic acid-targeting effector protein targeted to the brain may be contemplated. Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKC γ for *in vivo* gene silencing in the spinal cord of rats. Zou et al. administered about 10 μ l of a recombinant lentivirus having a titer of 1×10^9 transducing units (TU)/ml by an intrathecal catheter. A similar dosage of nucleic acid-targeting effector protein expressed in a lentiviral vector targeted to the brain may be contemplated for humans in the present invention, for example, about 10-50 ml of nucleic acid-targeting effector protein targeted to the brain in a lentivirus having a titer of 1×10^9 transducing units (TU)/ml may be contemplated.

[00450] In terms of local delivery to the brain, this can be achieved in various ways. For instance, material can be delivered intrastrially e.g., by injection. Injection can be performed stereotactically via a craniotomy.

Packaging and Promoters generally

[00451] Ways to package nucleic acid-targeting effector coding nucleic acid molecules, e.g., DNA, into vectors, e.g., viral vectors, to mediate genome modification *in vivo* include:

To achieve NHEJ-mediated gene knockout:

Single virus vector:

Vector containing two or more expression cassettes:

Promoter-nucleic acid-targeting effector protein coding nucleic acid molecule - terminator

Promoter- guide RNA 1-terminator

Promoter- guide RNA (N)-terminator (up to size limit of vector)

Double virus vector:

Vector 1 containing one expression cassette for driving the expression of nucleic acid-targeting effector protein

Promoter- nucleic acid-targeting effector protein coding nucleic acid molecule-terminator

Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

Promoter- guide RNA1 -terminator

Promoter- guide RNA1 (N)-terminator (up to size limit of vector)

To mediate homology-directed repair.

In addition to the single and double virus vector approaches described above, an additional vector is used to deliver a homology-direct repair template.

[00452] The promoter used to drive nucleic acid-targeting effector protein coding nucleic acid molecule expression can include:

AAV ITR can serve as a promoter: this is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce potential toxicity due to over expression of nucleic acid-targeting effector protein.

For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

For brain or other CNS expression, can use promoters: Synapsin1 for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

For liver expression, can use Albumin promoter.

For lung expression, can use SP-B.

For endothelial cells, can use ICAM.

For hematopoietic cells can use IFNbeta or CD45.

For Osteoblasts can use OG-2.

[00453] The promoter used to drive guide RNA can include:

Pol III promoters such as U6 or HI

Use of Pol II promoter and intronic cassettes to express guide RNA

Adeno associated virus (AAV)

[00454] nucleic acid-targeting effector protein and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or

viral vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. Doses may be based on or extrapolated to an average 70 kg individual (e.g., a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific genome / transcriptome modification, the expression of nucleic acid-targeting effector protein can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression (e.g., for targeting CNS disorders) might use the Synapsin I promoter.

[00455] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons:

- Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response) and
- Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[00456] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that nucleic acid-targeting effector protein (such as a Type V protein such as Cas13) as well as a promoter and transcription terminator may be all fit into the same viral vector. Therefore embodiments of the invention include utilizing homologs of nucleic acid-targeting effector protein that are shorter.

[00457] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination

thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The herein promoters and vectors are preferred individually. A tabulation of certain AAV serotypes as to these cells (see Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)) is as follows:

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

Lentivirus

[00458] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[00459] Lentiviruses may be prepared as follows. After cloning pCasESIO (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media was changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells were transfected with 10 µg of lentiviral transfer plasmid (pCasESIO) and the following packaging plasmids: 5 µg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat). Transfection was done in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100uL Plus reagent). After 6 hours, the media was changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

[00460] Lentivirus may be purified as follows. Viral supernatants were harvested after 48 hours. Supernatants were first cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They were then spun in a ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets were resuspended in 50ul of DMEM overnight at 4C. They were then aliquotted and immediately frozen at -80°C.

[00461] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 - 285). In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980-991 (September 2012)) and this vector may be modified for the nucleic acid-targeting system of the present invention.

[00462] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti-CCR5 -specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the nucleic acid-targeting system of the present invention. A minimum of 2.5×10^6 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 µmol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2×10^6 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm² tissue culture flasks coated with fibronectin (25 mg/cm²) (RetroNectin, Takara Bio Inc.).

[00463] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281 180, 20090007284, US201 101 17189; US200900 17543; US20070054961, US201003 17109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US201 10293571; US201 10293571, US20040013648, US20070025970, US200901 11106 and US Patent No. US7259015.

RNA delivery

[00464] RNA delivery: The nucleic acid-targeting Cas protein, for instance a Type V protein such as Cas13, and/or guide RNA, can also be delivered in the form of RNA. nucleic

acid-targeting Cas protein (such as a Type VI protein such as Cas13) mRNA can be generated using *in vitro* transcription. For example, nucleic acid-targeting effector protein (such as a Type V protein such as Cas13) mRNA can be synthesized using a PCR cassette containing the following elements: T7_promoter-kozak sequence (GCCACC)-effector protein-3' UTR from beta globin-polyA tail (a string of 120 or more adenines). The cassette can be used for transcription by T7 polymerase. Guide RNAs can also be transcribed using *in vitro* transcription from a cassette containing T7_promoter-GG- guide RNA sequence.

[00465] To enhance expression and reduce possible toxicity, the nucleic acid-targeting effector protein-coding sequence and/or the guide RNA can be modified to include one or more modified nucleoside e.g., using pseudo-U or 5-Methyl-C.

[00466] mRNA delivery methods are especially promising for liver delivery currently.

[00467] Much clinical work on RNA delivery has focused on RNAi or antisense, but these systems can be adapted for delivery of RNA for implementing the present invention. References below to RNAi *etc.* should be read accordingly.

Particle delivery systems and/or formulations:

[00468] Several types of particle delivery systems and/or formulations are known to be useful in a diverse spectrum of biomedical applications. In general, a particle is defined as a small object that behaves as a whole unit with respect to its transport and properties. Particles are further classified according to diameter. Coarse particles cover a range between 2,500 and 10,000 nanometers. Fine particles are sized between 100 and 2,500 nanometers. Ultrafine particles, or nanoparticles, are generally between 1 and 100 nanometers in size. The basis of the 100-nm limit is the fact that novel properties that differentiate particles from the bulk material typically develop at a critical length scale of under 100 nm.

[00469] As used herein, a particle delivery system/formulation is defined as any biological delivery system/formulation which includes a particle in accordance with the present invention. A particle in accordance with the present invention is any entity having a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, inventive particles have a greatest dimension of less than 10 μm . In some embodiments, inventive particles have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, or 100 nm. Typically, inventive particles have a greatest dimension (e.g., diameter) of 500 nm or less. In some

embodiments, inventive particles have a greatest dimension (e.g., diameter) of 250 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 200 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 150 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 100 nm or less. Smaller particles, e.g., having a greatest dimension of 50 nm or less are used in some embodiments of the invention. In some embodiments, inventive particles have a greatest dimension ranging between 25 nm and 200 nm.

[00470] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), ultraviolet-visible spectroscopy, dual polarisation interferometry and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo (herein cargo refers to e.g., one or more components of CRISPR-Cas system e.g., CRISPR enzyme or mRNA or guide RNA, or any combination thereof, and may include additional carriers and/or excipients) to provide particles of an optimal size for delivery for any *in vitro*, *ex vivo* and/or *in vivo* application of the present invention. In certain preferred embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS). Mention is made of US Patent No. 8,709,843; US Patent No. 6,007,845; US Patent No. 5,855,913; US Patent No. 5,985,309; US Patent No. 5,543,158; and the publication by James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84, concerning particles, methods of making and using them and measurements thereof.

[00471] Particles delivery systems within the scope of the present invention may be provided in any form, including but not limited to solid, semi-solid, emulsion, or colloidal particles. As such any of the delivery systems described herein, including but not limited to, e.g., lipid-based systems, liposomes, micelles, microvesicles, exosomes, or gene gun may be provided as particle delivery systems within the scope of the present invention.

Particles

[00472] CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using particles or lipid envelopes; for instance, CRISPR enzyme and RNA of the invention, e.g., as

a complex, can be delivered via a particle as in Dahlman et al., WO2015089419 A2 and documents cited therein, such as 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84), e.g., delivery particle comprising lipid or lipidoid and hydrophilic polymer, e.g., cationic lipid and hydrophilic polymer, for instance wherein the the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium -propane (DOTAP) or 1,2-ditetradecanoyl-s«-glycero-3-phosphocholine (DMPC) and/or wherein the hydrophilic polymer comprises ethylene glycol or polyethylene glycol (PEG); and/or wherein the particle further comprises cholesterol (e.g., particle from formulation 1 = DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; formulation number 2 = DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; formulation number 3 = DOTAP 90, DMPC 0, PEG 5, Cholesterol 5), wherein particles are formed using an efficient, multistep process wherein first, effector protein and RNA are mixed together, e.g., at a 1:1 molar ratio, e.g., at room temperature, e.g., for 30 minutes, e.g., in sterile, nuclease free IX PBS; and separately, DOTAP, DMPC, PEG, and cholesterol as applicable for the formulation are dissolved in alcohol, e.g., 100% ethanol; and, the two solutions are mixed together to form particles containing the complexes).

[00473] Nucleic acid-targeting effector proteins (such as a Type VI protein such as Cas13) mRNA and guide RNA may be delivered simultaneously using particles or lipid envelopes.

[00474] For example, Su X, Fricke J, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun 6;8(3):774-87. doi: 10.1021/impl00390w. Epub 2011 Apr 1) describes biodegradable core-shell structured particles with a poly (P-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[00475] In one embodiment, particles based on self-assembling bioadhesive polymers are contemplated, which may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, all to the brain. Other embodiments, such as oral absorption and ocular delivery of hydrophobic drugs are also contemplated. The molecular envelope technology involves an engineered polymer envelope which is protected and delivered to the site of the disease (see, e.g., Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1): 14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6): 1665-80; Lalatsa, A., et al. Mol

Pharm, 2012. 9(6): 1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X., et al. Biomacromolecules, 2006. 7(12):3452-9 and Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). Doses of about 5 mg/kg are contemplated, with single or multiple doses, depending on the target tissue.

[00476] In one embodiment, particles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan Anderson's lab at MIT may be used/and or adapted to the nucleic acid-targeting system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification, characterization, and formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 and Lee et al., NatNanotechnol. 2012 Jun 3;7(6):389-93.

[00477] US patent application 201 10293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the nucleic acid-targeting system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles, liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The aminoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[00478] US Patent Publication No. 201 10293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid

compound. These primary or secondary amines are left as is or may be reacted with another electrophile such as a different epoxide-terminated compound. As will be appreciated by one skilled in the art, reacting an amine with less than excess of epoxide-terminated compound will result in a plurality of different aminoalcohol lipidoid compounds with various numbers of tails. Certain amines may be fully functionalized with two epoxide-derived compound tails while other molecules will not be completely functionalized with epoxide-derived compound tails. For example, a diamine or polyamine may include one, two, three, or four epoxide-derived compound tails off the various amino moieties of the molecule resulting in primary, secondary, and tertiary amines. In certain embodiments, all the amino groups are not fully functionalized. In certain embodiments, two of the same types of epoxide-terminated compounds are used. In other embodiments, two or more different epoxide-terminated compounds are used. The synthesis of the aminoalcohol lipidoid compounds is performed with or without solvent, and the synthesis may be performed at higher temperatures ranging from 30-100 °C, preferably at approximately 50-90 °C. The prepared aminoalcohol lipidoid compounds may be optionally purified. For example, the mixture of aminoalcohol lipidoid compounds may be purified to yield an aminoalcohol lipidoid compound with a particular number of epoxide-derived compound tails. Or the mixture may be purified to yield a particular stereo- or regioisomer. The aminoalcohol lipidoid compounds may also be alkylated using an alkyl halide (e.g., methyl iodide) or other alkylating agent, and/or they may be acylated.

[00479] US Patent Publication No. 201 10293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[00480] US Patent Publication No. 20130302401 relates to a class of poly(beta-amino alcohols) (PBAs) has been prepared using combinatorial polymerization. The inventive PBAs may be used in biotechnology and biomedical applications as coatings (such as coatings of films or multilayer films for medical devices or implants), additives, materials, excipients, non-biofouling agents, micropatterning agents, and cellular encapsulation agents. When used as surface coatings, these PBAs elicited different levels of inflammation, both in vitro and in vivo, depending on their chemical structures. The large chemical diversity of this class of materials allowed us to identify polymer coatings that inhibit macrophage

activation in vitro. Furthermore, these coatings reduce the recruitment of inflammatory cells, and reduce fibrosis, following the subcutaneous implantation of carboxylated polystyrene microparticles. These polymers may be used to form polyelectrolyte complex capsules for cell encapsulation. The invention may also have many other biological applications such as antimicrobial coatings, DNA or siRNA delivery, and stem cell tissue engineering. The teachings of US Patent Publication No. 20130302401 may be applied to the nucleic acid-targeting system of the present invention.

[00481] In another embodiment, lipid nanoparticles (LNPs) are contemplated. An antitransferrin small interfering RNA has been encapsulated in lipid nanoparticles and delivered to humans (see, e.g., Coelho et al., *N Engl J Med* 2013;369:819-29), and such a system may be adapted and applied to the nucleic acid-targeting system of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetaminophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[00482] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., *Cancer Discovery*, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering RNA encoding nucleic acid-targeting effector protein to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al. demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiving doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[00483] However, the charge of the LNP may be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see,

e.g., Rosin et al, *Molecular Therapy*, vol. 19, no. 12, pages 1286-2200, Dec. 2011). Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilinoyleyl-3-dimethylammonium -propane (DLinDAP), 1,2-dilinoyleoxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoyleoxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoyleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA»DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, *Molecular Therapy*, vol. 19, no. 12, pages 1286-2200, Dec. 2011). A dosage of 1 µg/ml of LNP or CRISPR-Cas RNA in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[00484] Preparation of LNPs and CRISPR-Cas encapsulation may be used/and or adapted from Rosin et al, *Molecular Therapy*, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The cationic lipids 1,2-dilinoyleyl-3-dimethylammonium -propane (DLinDAP), 1,2-dilinoyleoxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoyleoxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoyleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-*o*-[2''-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(*o*-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxylpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific nucleic acid-targeting complex (CRISPR-Cas) RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada).

Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31 °C for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Particle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. RNA encapsulation efficiency may be determined by removal of free RNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted particles and quantified at 260 nm. RNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA). In conjunction with the herein discussion of LNPs and PEG lipids, PEGylated liposomes or LNPs are likewise suitable for delivery of a nucleic acid-targeting system or components thereof.

[00485] Preparation of large LNPs may be used/and or adapted from Rosin et al, *Molecular Therapy*, vol. 19, no. 12, pages 1286-2200, Dec. 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a molar ratio of 0.75:1 (sodium acetate:DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/l, pH 3.0) with vigorous stirring, resulting in spontaneous liposome formation in aqueous buffer containing 35% ethanol. The liposome solution may be incubated at 37 °C to allow for time-dependent increase in particle size. Aliquots may be removed at various times during incubation to investigate changes in liposome size by dynamic light scattering (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, UK). Once the desired particle size is achieved, an aqueous PEG lipid solution (stock = 10 mg/ml PEG-DMG in 35% (vol/vol) ethanol) may be added to the liposome mixture to yield a final PEG molar concentration of 3.5% of total lipid. Upon addition of PEG-lipids, the liposomes should their size, effectively quenching further growth. RNA may then be added to the empty liposomes at a RNA to total lipid ratio of approximately 1:10 (wt:wt), followed by incubation for 30 minutes at 37 °C to form loaded LNPs. The mixture may be subsequently dialyzed overnight in PBS and filtered with a 0.45- μ m syringe filter.

[00486] Spherical Nucleic Acid (SNATM) constructs and other particles (particularly gold particles) are also contemplated as a means to delivery nucleic acid-targeting system to intended targets. Significant data show that AuraSense Therapeutics' Spherical Nucleic Acid (SNATM) constructs, based upon nucleic acid-functionalized gold particles, are useful.

[00487] Literature that may be employed in conjunction with herein teachings include: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, 10:186-192.

[00488] Self-assembling particles with RNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG). This system has been used, for example, as a means to target tumor neovasculature expressing integrins and deliver siRNA inhibiting vascular endothelial growth factor receptor-2 (VEGF R2) expression and thereby achieve tumor angiogenesis (see, e.g., Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19). Nanoplexes may be prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. A dosage of about 100 to 200 mg of nucleic acid-targeting complex RNA is envisioned for delivery in the self-assembling particles of Schiffelers et al.

[00489] The nanoplexes of Bartlett et al. (PNAS, September 25, 2007, vol. 104, no. 39) may also be applied to the present invention. The nanoplexes of Bartlett et al. are prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. The DOTA-siRNA of Bartlett et al. was synthesized as follows: 1,4,7,10-tetraazacyclododecane-1, 4,7,10-tetraacetic acid mono(N-hydroxysuccinimide ester) (DOTA-NHSester) was ordered from Macrocyclics (Dallas, TX).

The amine modified RNA sense strand with a 100-fold molar excess of DOTA-NHS-ester in carbonate buffer (pH 9) was added to a microcentrifuge tube. The contents were reacted by stirring for 4 h at room temperature. The DOTA-RNAsense conjugate was ethanol-precipitated, resuspended in water, and annealed to the unmodified antisense strand to yield DOTA-siRNA. All liquids were pretreated with Chelex-100 (Bio-Rad, Hercules, CA) to remove trace metal contaminants. Tf-targeted and nontargeted siRNA particles may be formed by using cyclodextrin-containing polycations. Typically, particles were formed in water at a charge ratio of 3 (+/-) and an siRNA concentration of 0.5 g/liter. One percent of the adamantane-PEG molecules on the surface of the targeted particles were modified with Tf (adamantane-PEG-Tf). The particles were suspended in a 5% (wt/vol) glucose carrier solution for injection.

[00490] Davis et al. (Nature, Vol 464, 15 April 2010) conducts a RNA clinical trial that uses a targeted particle-delivery system (clinical trial registration number NCT00689065). Patients with solid cancers refractory to standard-of-care therapies are administered doses of targeted particles on days 1, 3, 8 and 10 of a 21-day cycle by a 30-min intravenous infusion. The particles comprise, consist essentially of, or consist of a synthetic delivery system containing: (1) a linear, cyclodextrin-based polymer (CDP), (2) a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells, (3) a hydrophilic polymer (polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids), and (4) siRNA designed to reduce the expression of the RRM2 (sequence used in the clinic was previously denoted siR2B+5). The TFR has long been known to be upregulated in malignant cells, and RRM2 is an established anti-cancer target. These particles (clinical version denoted as CALAA-01) have been shown to be well tolerated in multi-dosing studies in non-human primates. Although a single patient with chronic myeloid leukaemia has been administered siRNA by liposomal delivery, Davis et al.'s clinical trial is the initial human trial to systemically deliver siRNA with a targeted delivery system and to treat patients with solid cancer. To ascertain whether the targeted delivery system can provide effective delivery of functional siRNA to human tumours, Davis et al. investigated biopsies from three patients from three different dosing cohorts; patients A, B and C, all of whom had metastatic melanoma and received CALAA-01 doses of 18, 24 and 30 mg m⁻² siRNA, respectively. Similar doses may also be contemplated for the nucleic acid-targeting system of the present invention. The delivery of the invention may be achieved with particles containing a linear, cyclodextrin-based polymer (CDP), a human transferrin protein (TF) targeting ligand displayed on the exterior of the

particle to engage TF receptors (TFR) on the surface of the cancer cells and/or a hydrophilic polymer (for example, polyethylene glycol (PEG) used to promote particle stability in biological fluids).

[00491] In terms of this invention, it is preferred to have one or more components of nucleic acid-targeting complex, e.g., nucleic acid-targeting effector protein or mRNA, or guide RNA delivered using particles or lipid envelopes. Other delivery systems or vectors are may be used in conjunction with the particle aspects of the invention.

[00492] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain preferred embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other preferred embodiments, particles of the invention have a greatest dimension of 100 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm.

[00493] Particles encompassed in the present invention may be provided in different forms, e.g., as solid particles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of particles, or combinations thereof. Metal, dielectric, and semiconductor particles may be prepared, as well as hybrid structures (e.g., core-shell particles). Particles made of semiconducting material may also be labeled quantum dots if they are small enough (typically sub 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present invention.

[00494] Semi-solid and soft particles have been manufactured, and are within the scope of the present invention. A prototype particle of semi-solid nature is the liposome. Various types of liposome particles are currently used clinically as delivery systems for anticancer drugs and vaccines. Particles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

[00495] US Patent No. 8,709,843, incorporated herein by reference, provides a drug delivery system for targeted delivery of therapeutic agent-containing particles to tissues, cells, and intracellular compartments. The invention provides targeted particles comprising polymer conjugated to a surfactant, hydrophilic polymer or lipid.

[00496] US Patent No. 6,007,845, incorporated herein by reference, provides particles which have a core of a multiblock copolymer formed by covalently linking a multifunctional

compound with one or more hydrophobic polymers and one or more hydrophilic polymers, and contain a biologically active material.

[00497] US Patent No. 5,855,913, incorporated herein by reference, provides a particulate composition having aerodynamically light particles having a tap density of less than 0.4 g/cm³ with a mean diameter of between 5 μm and 30 μm, incorporating a surfactant on the surface thereof for drug delivery to the pulmonary system.

[00498] US Patent No. 5,985,309, incorporated herein by reference, provides particles incorporating a surfactant and/or a hydrophilic or hydrophobic complex of a positively or negatively charged therapeutic or diagnostic agent and a charged molecule of opposite charge for delivery to the pulmonary system.

[00499] US Patent No. 5,543,158, incorporated herein by reference, provides biodegradable injectable particles having a biodegradable solid core containing a biologically active material and poly(alkylene glycol) moieties on the surface.

[00500] WO2012135025 (also published as US20120251560), incorporated herein by reference, describes conjugated polyethyleneimine (PEI) polymers and conjugated azamacrocycles (collectively referred to as "conjugated lipomer" or "lipomers"). In certain embodiments, it can be envisioned that such methods and materials of herein-cited documents, e.g., conjugated lipomers can be used in the context of the nucleic acid-targeting system to achieve in vitro, ex vivo and in vivo genomic perturbations to modify gene expression, including modulation of protein expression.

[00501] In one embodiment, the particle may be epoxide-modified lipid-polymer, advantageously 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84). C71 was synthesized by reacting C15 epoxide-terminated lipids with PEI600 at a 14:1 molar ratio, and was formulated with C14PEG2000 to produce particles (diameter between 35 and 60 nm) that were stable in PBS solution for at least 40 days.

[00502] An epoxide-modified lipid-polymer may be utilized to deliver the nucleic acid-targeting system of the present invention to pulmonary, cardiovascular or renal cells, however, one of skill in the art may adapt the system to deliver to other target organs. Dosage ranging from about 0.05 to about 0.6 mg/kg are envisioned. Dosages over several days or weeks are also envisioned, with a total dosage of about 2 mg/kg.

Exosomes

[00503] Exosomes are endogenous nano-vesicles that transport RNAs and proteins, and which can deliver RNA to the brain and other target organs. To reduce immunogenicity,

Alvarez-Erviti et al. (2011, Nat Biotechnol 29: 341) used self-derived dendritic cells for exosome production. Targeting to the brain was achieved by engineering the dendritic cells to express Lamp2b, an exosomal membrane protein, fused to the neuron-specific RVG peptide. Purified exosomes were loaded with exogenous RNA by electroporation. Intravenously injected RVG-targeted exosomes delivered GAPDH siRNA specifically to neurons, microglia, oligodendrocytes in the brain, resulting in a specific gene knockdown. Pre-exposure to RVG exosomes did not attenuate knockdown, and non-specific uptake in other tissues was not observed. The therapeutic potential of exosome-mediated siRNA delivery was demonstrated by the strong mRNA (60%) and protein (62%) knockdown of BACE1, a therapeutic target in Alzheimer's disease.

[00504] To obtain a pool of immunologically inert exosomes, Alvarez-Erviti et al. harvested bone marrow from inbred C57BL/6 mice with a homogenous major histocompatibility complex (MHC) haplotype. As immature dendritic cells produce large quantities of exosomes devoid of T-cell activators such as MHC-II and CD86, Alvarez-Erviti et al. selected for dendritic cells with granulocyte/macrophage-colony stimulating factor (GM-CSF) for 7 d. Exosomes were purified from the culture supernatant the following day using well-established ultracentrifugation protocols. The exosomes produced were physically homogenous, with a size distribution peaking at 80 nm in diameter as determined by particle tracking analysis (NTA) and electron microscopy. Alvarez-Erviti et al. obtained 6-12 µg of exosomes (measured based on protein concentration) per 10⁶ cells.

[00505] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled RNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated RNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 µT resulted in the greatest retention of RNA and was used for all subsequent experiments.

[00506] Alvarez-Erviti et al. administered 150 µg of each BACE1 siRNA encapsulated in 150 µg of RVG exosomes to normal C57BL/6 mice and compared the knockdown efficiency to four controls: untreated mice, mice injected with RVG exosomes only, mice injected with BACE1 siRNA complexed to an in vivo cationic liposome reagent and mice injected with BACE1 siRNA complexed to RVG-9R, the RVG peptide conjugated to 9 D-arginines that electrostatically binds to the siRNA. Cortical tissue samples were analyzed 3 d after administration and a significant protein knockdown (45%, P < 0.05, versus 62%, P < 0.01) in

both siRNA-RVG-9R-treated and siRNARVG-exosome-treated mice was observed, resulting from a significant decrease in BACE1 mRNA levels (66% [+ or -] 15%, $P < 0.001$ and 61% [+ or -] 13%, respectively, $P < 0.01$). Moreover, Applicants demonstrated a significant decrease (55%, $P < 0.05$) in the total [beta]-amyloid 1-42 levels, a main component of the amyloid plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease observed was greater than the β -amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[00507] Finally, Alvarez-Erviti et al. investigated whether RNA-RVG-exosomes induced immune responses in vivo by assessing IL-6, IP-10, TNF α and IFN- α serum concentrations. Following exosome treatment, nonsignificant changes in all cytokines were registered similar to siRNA-transfection reagent treatment in contrast to siRNA-RVG-9R, which potently stimulated IL-6 secretion, confirming the immunologically inert profile of the exosome treatment. Given that exosomes encapsulate only 20% of siRNA, delivery with RVG-exosome appears to be more efficient than RVG-9R delivery as comparable mRNA knockdown and greater protein knockdown was achieved with fivefold less siRNA without the corresponding level of immune stimulation. This experiment demonstrated the therapeutic potential of RVG-exosome technology, which is potentially suited for long-term silencing of genes related to neurodegenerative diseases. The exosome delivery system of Alvarez-Erviti et al. may be applied to deliver the nucleic acid-targeting system of the present invention to therapeutic targets, especially neurodegenerative diseases. A dosage of about 100 to 1000 mg of nucleic acid-targeting system encapsulated in about 100 to 1000 mg of RVG-exosomes may be contemplated for the present invention.

[00508] El-Andaloussi et al. (Nature Protocols 7,2112-2126(2012)) discloses how exosomes derived from cultured cells can be harnessed for delivery of RNA in vitro and in vivo. This protocol first describes the generation of targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. Next, El-Andaloussi et al. explain how to purify and characterize exosomes from transfected cell supernatant. Next, El-Andaloussi et al. detail crucial steps for loading RNA into exosomes. Finally, El-Andaloussi et al. outline how to use exosomes to efficiently deliver RNA in vitro and in vivo in mouse brain. Examples of anticipated results in which exosome-mediated RNA delivery is evaluated by functional assays and imaging are also provided. The entire

protocol takes ~3 weeks. Delivery or administration according to the invention may be performed using exosomes produced from self-derived dendritic cells. From the herein teachings, this can be employed in the practice of the invention

[00509] In another embodiment, the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) are contemplated. Exosomes are nano-sized vesicles (30-90nm in size) produced by many cell types, including dendritic cells (DC), B cells, T cells, mast cells, epithelial cells and tumor cells. These vesicles are formed by inward budding of late endosomes and are then released to the extracellular environment upon fusion with the plasma membrane. Because exosomes naturally carry RNA between cells, this property may be useful in gene therapy, and from this disclosure can be employed in the practice of the instant invention.

[00510] Exosomes from plasma can be prepared by centrifugation of buffy coat at 900g for 20 min to isolate the plasma followed by harvesting cell supernatants, centrifuging at 300g for 10 min to eliminate cells and at 16 500g for 30 min followed by filtration through a 0.22 mm filter. Exosomes are pelleted by ultracentrifugation at 120 000g for 70 min. Chemical transfection of siRNA into exosomes is carried out according to the manufacturer's instructions in RNAi Human/Mouse Starter Kit (Qiagen, Hilden, Germany). siRNA is added to 100 ml PBS at a final concentration of 2 mmol/ml. After adding HiPerFect transfection reagent, the mixture is incubated for 10 min at RT. In order to remove the excess of micelles, the exosomes are re-isolated using aldehyde/sulfate latex beads. The chemical transfection of nucleic acid-targeting system into exosomes may be conducted similarly to siRNA. The exosomes may be co-cultured with monocytes and lymphocytes isolated from the peripheral blood of healthy donors. Therefore, it may be contemplated that exosomes containing nucleic acid-targeting system may be introduced to monocytes and lymphocytes of and autologously reintroduced into a human. Accordingly, delivery or administration according to the invention may be performed using plasma exosomes.

Liposomes

[00511] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and

transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, *Journal of Drug Delivery*, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1155/201 1/469679 for review).

[00512] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, *Journal of Drug Delivery*, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi:10.1155/201 1/469679 for review).

[00513] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, *Journal of Drug Delivery*, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1155/201 1/469679 for review).

[00514] A liposome formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoyl-sn-glycero-3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, *Journal of Drug Delivery*, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1155/201 1/469679 for review).

[00515] In a particularly advantageous embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at <http://cshprotocols.cshlp.Org/content/2010/4/pdb.prot5407.long>. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to surface allow crossing of the blood brain barrier via endocytosis. Applicant

postulates utilizing Trojan Horse Liposomes to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of DNA or RNA may be contemplated for in vivo administration in liposomes.

[00516] In another embodiment, the nucleic acid-targeting system or components thereof may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., *Nature Biotechnology*, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of a specific nucleic acid-targeting system targeted in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, a specific nucleic acid-targeting system encapsulated SNALP) administered by intravenous injection to at doses of about 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., *Nature Letters*, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(w-methoxypoly(ethylene glycol) 2000) carbamoyl] -1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar per cent ratio (see, e.g., Zimmerman et al., *Nature Letters*, Vol. 441, 4 May 2006).

[00517] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-16 derived liver tumors (see, e.g., Li, *Gene Therapy* (2012) 19, 775-780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80-100 nm in size.

[00518] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, MO, USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., *Lancet* 2010; 375: 1896-905). A dosage of about 2 mg/kg total nucleic acid-targeting system per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[00519] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA) (see, e.g.,

Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for in vivo studies may comprise a final lipid/RNA mass ratio of about 9:1.

[00520] The safety profile of RNAi nanomedicines has been reviewed by Barros and Gollob of Alnylam Pharmaceuticals (see, e.g., *Advanced Drug Delivery Reviews* 64 (2012) 1730-1737). The stable nucleic acid lipid particle (SNALP) is comprised of four different lipids — an ionizable lipid (DLinDMA) that is cationic at low pH, a neutral helper lipid, cholesterol, and a diffusible polyethylene glycol (PEG)-lipid. The particle is approximately 80 nm in diameter and is charge-neutral at physiologic pH. During formulation, the ionizable lipid serves to condense lipid with the anionic RNA during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid also mediates the fusion of SNALP with the endosomal membrane enabling release of RNA into the cytoplasm. The PEG-lipid stabilizes the particle and reduces aggregation during formulation, and subsequently provides a neutral hydrophilic exterior that improves pharmacokinetic properties.

[00521] To date, two clinical programs have been initiated using SNALP formulations with RNA. Tekmira Pharmaceuticals recently completed a phase I single-dose study of SNALP-ApoB in adult volunteers with elevated LDL cholesterol. ApoB is predominantly expressed in the liver and jejunum and is essential for the assembly and secretion of VLDL and LDL. Seventeen subjects received a single dose of SNALP-ApoB (dose escalation across 7 dose levels). There was no evidence of liver toxicity (anticipated as the potential dose-limiting toxicity based on preclinical studies). One (of two) subjects at the highest dose experienced flu-like symptoms consistent with immune system stimulation, and the decision was made to conclude the trial.

[00522] Alnylam Pharmaceuticals has similarly advanced ALN-TTRO1, which employs the SNALP technology described above and targets hepatocyte production of both mutant and wild-type TTR to treat TTR amyloidosis (ATTR). Three ATTR syndromes have been described: familial amyloidotic polyneuropathy (FAP) and familial amyloidotic cardiomyopathy (FAC) — both caused by autosomal dominant mutations in TTR; and senile systemic amyloidosis (SSA) cause by wildtype TTR. A placebo-controlled, single dose-escalation phase I trial of ALN-TTRO1 was recently completed in patients with ATTR. ALN-TTRO1 was administered as a 15-minute IV infusion to 31 patients (23 with study drug and 8 with placebo) within a dose range of 0.01 to 1.0 mg/kg (based on siRNA). Treatment was well tolerated with no significant increases in liver function tests. Infusion-related reactions were noted in 3 of 23 patients at >0.4 mg/kg; all responded to slowing of the infusion rate and

all continued on study. Minimal and transient elevations of serum cytokines IL-6, IP-10 and IL-1ra were noted in two patients at the highest dose of 1 mg/kg (as anticipated from preclinical and NHP studies). Lowering of serum TTR, the expected pharmacodynamics effect of ALN-TTR01, was observed at 1 mg/kg.

[00523] In yet another embodiment, a SNALP may be made by solubilizing a cationic lipid, DSPC, cholesterol and PEG-lipid e.g., in ethanol, e.g., at a molar ratio of 40:10:40:10, respectively (see, Semple et al., Nature Nanotechnology, Volume 28 Number 2 February 2010, pp. 172-177). The lipid mixture was added to an aqueous buffer (50 mM citrate, pH 4) with mixing to a final ethanol and lipid concentration of 30% (vol/vol) and 6.1 mg/ml, respectively, and allowed to equilibrate at 22 °C for 2 min before extrusion. The hydrated lipids were extruded through two stacked 80 nm pore-sized filters (Nuclepore) at 22 °C using a Lipex Extruder (Northern Lipids) until a vesicle diameter of 70-90 nm, as determined by dynamic light scattering analysis, was obtained. This generally required 1-3 passes. The siRNA (solubilized in a 50 mM citrate, pH 4 aqueous solution containing 30% ethanol) was added to the pre-equilibrated (35 °C) vesicles at a rate of ~5 ml/min with mixing. After a final target siRNA/lipid ratio of 0.06 (wt/wt) was reached, the mixture was incubated for a further 30 min at 35 °C to allow vesicle reorganization and encapsulation of the siRNA. The ethanol was then removed and the external buffer replaced with PBS (155 mM NaCl, 3 mM Na₂HPO₄, 1 mM KH₂PO₄, pH 7.5) by either dialysis or tangential flow diafiltration. siRNA were encapsulated in SNALP using a controlled step-wise dilution method process. The lipid constituents of KC2-SNALP were DLin-KC2-DMA (cationic lipid), dipalmitoylphosphatidylcholine (DPPC; Avanti Polar Lipids), synthetic cholesterol (Sigma) and PEG-C-DMA used at a molar ratio of 57.1:7.1:34.3:1.4. Upon formation of the loaded particles, SNALP were dialyzed against PBS and filter sterilized through a 0.2 µm filter before use. Mean particle sizes were 75-85 nm and 90-95% of the siRNA was encapsulated within the lipid particles. The final siRNA/lipid ratio in formulations used for in vivo testing was 0.15 (wt/wt). LNP-siRNA systems containing Factor VII siRNA were diluted to the appropriate concentrations in sterile PBS immediately before use and the formulations were administered intravenously through the lateral tail vein in a total volume of 10 ml/kg. This method and these delivery systems may be extrapolated to the nucleic acid-targeting system of the present invention.

[00524] Other Lipids

[00525] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate nucleic acid-targeting

system or components thereof or nucleic acid molecule(s) coding therefor e.g., similar to siRNA (see, e.g., Jayaraman, *Angew. Chem. Int. Ed.* 2012, 51, 8529 -8533), and hence may be employed in the practice of the invention. A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70-90 nm and a low polydispersity index of 0.11 ± 0.04 ($n=56$), the particles may be extruded up to three times through 80 nm membranes prior to adding the guide RNA. Particles containing the highly potent amino lipid 16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance in vivo activity.

[00526] Michael S D Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: *Nature Biotechnology*, Volume:29, Pages: 154-157 (2011)) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[00527] In another embodiment, lipids may be formulated with the nucleic acid-targeting system of the present invention or component(s) thereof or nucleic acid molecule(s) coding therefor to form lipid nanoparticles (LNPs). Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with RNA-targeting system instead of siRNA (see, e.g., Novobrantseva, *Molecular Therapy-Nucleic Acids* (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid:siRNA weight ratio may be 12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid particles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[00528] Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of LNPs and LNP formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316), all of which may be used and/or adapted to the present invention.

[00529] The nucleic acid-targeting system or components thereof or nucleic acid molecule(s) coding therefor may be delivered encapsulated in PLGA Microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279 (assigned to Moderna Therapeutics) which relate to aspects of formulation of compositions comprising modified nucleic acid molecules which may encode a protein, a protein precursor, or a partially or fully processed form of the protein or a protein precursor. The formulation may have a molar ratio 50:10:38.5:1.5-3.0 (cationic lipid: fusogenic lipid: cholesterol: PEG lipid). The PEG lipid may be selected from, but is not limited to PEG-c-DOMG, PEG-DMG. The fusogenic lipid may be DSPC. See also, Schrum et al., Delivery and Formulation of Engineered Nucleic Acids, US published application 20120251618.

[00530] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumours, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2): 1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[00531] US Patent Publication No. 20050019923 describes cationic dendrimers for delivering bioactive molecules, such as polynucleotide molecules, peptides and polypeptides and/or pharmaceutical agents, to a mammalian body. The dendrimers are suitable for targeting the delivery of the bioactive molecules to, for example, the liver, spleen, lung, kidney or heart (or even the brain). Dendrimers are synthetic 3-dimensional macromolecules that are prepared in a step-wise fashion from simple branched monomer units, the nature and functionality of which can be easily controlled and varied. Dendrimers are synthesized from the repeated addition of building blocks to a multifunctional core (divergent approach to synthesis), or towards a multifunctional core (convergent approach to synthesis) and each addition of a 3-dimensional shell of building blocks leads to the formation of a higher generation of the dendrimers. Polypropylenimine dendrimers start from a diaminobutane core to which is added twice the number of amino groups by a double Michael addition of acrylonitrile to the primary amines followed by the hydrogenation of the nitriles. This results in a doubling of the amino groups. Polypropylenimine dendrimers contain 100% protonable nitrogens and up to 64 terminal amino groups (generation 5, DAB 64). Protonable groups are usually amine groups which are able to accept protons at neutral pH. The use of dendrimers as gene delivery agents has largely focused on the use of the polyamidoamine. and

phosphorous containing compounds with a mixture of amine/amide or N~P(02)S as the conjugating units respectively with no work being reported on the use of the lower generation polypropylenimine dendrimers for gene delivery. Polypropylenimine dendrimers have also been studied as pH sensitive controlled release systems for drug delivery and for their encapsulation of guest molecules when chemically modified by peripheral amino acid groups. The cytotoxicity and interaction of polypropylenimine dendrimers with DNA as well as the transfection efficacy of DAB 64 has also been studied.

[00532] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterised by undesirable cellular proliferation such as neoplasms and tumours, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumour activity may complement the activity of the agent to be delivered." The disclosures of these patent publications may be employed in conjunction with herein teachings for delivery of nucleic acid-targetingsystem(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Supercharged proteins

[00533] Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge and may be employed in delivery of nucleic acid-targetingsystem(s) or component(s) thereof or nucleic acid molecule(s) coding therefor. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, RNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both in vitro and in vivo. David Liu's lab reported the creation and characterization of supercharged proteins in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110-10112).

[00534] The nonviral delivery of RNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications (Akinc et al., 2010, Nat. Biotech. 26, 561-569). Purified +36 GFP protein (or other superpositively charged protein) is mixed with RNAs in the appropriate serum-free media and allowed to complex prior addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-RNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111-6116). However, pilot experiments varying the dose of protein and RNA should be performed to optimize the procedure for specific cell lines.

(1) One day before treatment, plate 1×10^5 cells per well in a 48-well plate.

(2) On the day of treatment, dilute purified +36 GFP protein in serumfree media to a final concentration 200nM. Add RNA to a final concentration of 50nM. Vortex to mix and incubate at room temperature for 10min.

(3) During incubation, aspirate media from cells and wash once with PBS.

(4) Following incubation of +36 GFP and RNA, add the protein-RNA complexes to cells.

(5) Incubate cells with complexes at 37 °C for 4h.

(6) Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48h or longer depending upon the assay for activity.

(7) Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

[00535] David Liu's lab has further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery Applicants have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications.

(1) One day before treatment, plate 1×10^5 per well in a 48-well plate.

(2) On the day of treatment, dilute purified p36 GFP protein in serumfree media to a final concentration 2 mM. Add 1mg of plasmid DNA. Vortex to mix and incubate at room temperature for 10min.

(3) During incubation, aspirate media from cells and wash once with PBS.

(4) Following incubation of p36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells.

(5) Incubate cells with complexes at 37 C for 4h.

(6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24-48h.

(7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[00536] See also, e.g., McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2010); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). The methods of the super charged proteins may be used and/or adapted for delivery of the nucleic acid-targeting system of the present invention. These systems of Dr. Lui and documents herein in conjunction with herein teachings can be employed in the delivery of nucleic acid-targeting system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Cell Penetrating Peptides (CPPs)

[00537] In yet another embodiment, cell penetrating peptides (CPPs) are contemplated for the delivery of the CRISPR Cas system. CPPs are short peptides that facilitate cellular uptake of various molecular cargo (from nanosize particles to small chemical molecules and large fragments of DNA). The term "cargo" as used herein includes but is not limited to the group consisting of therapeutic agents, diagnostic probes, peptides, nucleic acids, antisense oligonucleotides, plasmids, proteins, particles including nanoparticles, liposomes, chromophores, small molecules and radioactive materials. In aspects of the invention, the cargo may also comprise any component of the CRISPR Cas system or the entire functional CRISPR Cas system. Aspects of the present invention further provide methods for delivering a desired cargo into a subject comprising: (a) preparing a complex comprising the cell penetrating peptide of the present invention and a desired cargo, and (b) orally, intraarticularly, intraperitoneally, intrathecally, intrarterially, intranasally, intraparenchymally, subcutaneously, intramuscularly, intravenously, dermally, intrarectally, or topically administering the complex to a subject. The cargo is associated with the peptides either through chemical linkage via covalent bonds or through non-covalent interactions.

[00538] The function of the CPPs are to deliver the cargo into cells, a process that commonly occurs through endocytosis with the cargo delivered to the endosomes of living

mammalian cells. Cell-penetrating peptides are of different sizes, amino acid sequences, and charges but all CPPs have one distinct characteristic, which is the ability to translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPP translocation may be classified into three main entry mechanisms: direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure. CPPs have found numerous applications in medicine as drug delivery agents in the treatment of different diseases including cancer and virus inhibitors, as well as contrast agents for cell labeling. Examples of the latter include acting as a carrier for GFP, MRI contrast agents, or quantum dots. CPPs hold great potential as *in vitro* and *in vivo* delivery vectors for use in research and medicine. CPPs typically have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. One of the initial CPPs discovered was the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1) which was found to be efficiently taken up from the surrounding media by numerous cell types in culture. Since then, the number of known CPPs has expanded considerably and small molecule synthetic analogues with more effective protein transduction properties have been generated. CPPs include but are not limited to Penetratin, Tat (48-60), Transportan, and (R-Ahx-R4) (Ahx=aminohexanoyl).

[00539] US Patent 8,372,951, provides a CPP derived from eosinophil cationic protein (ECP) which exhibits highly cell-penetrating efficiency and low toxicity. Aspects of delivering the CPP with its cargo into a vertebrate subject are also provided. Further aspects of CPPs and their delivery are described in U. S. patents 8,575,305; 8,614,194 and 8,044,019. CPPs can be used to deliver the CRISPR-Cas system or components thereof. That CPPs can be employed to deliver the CRISPR-Cas system or components thereof is also provided in the manuscript "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", by Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, et al. *Genome Res.* 2014 Apr 2. [Epub ahead of print], incorporated by reference in its entirety, wherein it is demonstrated that treatment with CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs lead to endogenous gene disruptions in human cell lines. In the paper the Cas9 protein was conjugated to CPP via a thioether bond, whereas the guide RNA

was complexed with CPP, forming condensed, positively charged particles. It was shown that simultaneous and sequential treatment of human cells, including embryonic stem cells, dermal fibroblasts, HEK293T cells, HeLa cells, and embryonic carcinoma cells, with the modified Cas9 and guide RNA led to efficient gene disruptions with reduced off-target mutations relative to plasmid transfections.

Implantable devices

[00540] In another embodiment, implantable devices are also contemplated for delivery of the nucleic acid-targeting system or component(s) thereof or nucleic acid molecule(s) coding therefor. For example, US Patent Publication 20110195123 discloses an implantable medical device which elutes a drug locally and in prolonged period is provided, including several types of such a device, the treatment modes of implementation and methods of implantation. The device comprising of polymeric substrate, such as a matrix for example, that is used as the device body, and drugs, and in some cases additional scaffolding materials, such as metals or additional polymers, and materials to enhance visibility and imaging. An implantable delivery device can be advantageous in providing release locally and over a prolonged period, where drug is released directly to the extracellular matrix (ECM) of the diseased area such as tumor, inflammation, degeneration or for symptomatic objectives, or to injured smooth muscle cells, or for prevention. One kind of drug is RNA, as disclosed above, and this system may be used/and or adapted to the nucleic acid-targeting system of the present invention. The modes of implantation in some embodiments are existing implantation procedures that are developed and used today for other treatments, including brachytherapy and needle biopsy. In such cases the dimensions of the new implant described in this invention are similar to the original implant. Typically a few devices are implanted during the same treatment procedure.

[00541] US Patent Publication 201 10195123, provides a drug delivery implantable or insertable system, including systems applicable to a cavity such as the abdominal cavity and/or any other type of administration in which the drug delivery system is not anchored or attached, comprising a biostable and/or degradable and/or bioabsorbable polymeric substrate, which may for example optionally be a matrix. It should be noted that the term "insertion" also includes implantation. The drug delivery system is preferably implemented as a "Loder" as described in US Patent Publication 201 10195123.

[00542] The polymer or plurality of polymers are biocompatible, incorporating an agent and/or plurality of agents, enabling the release of agent at a controlled rate, wherein the total volume of the polymeric substrate, such as a matrix for example, in some embodiments is

optionally and preferably no greater than a maximum volume that permits a therapeutic level of the agent to be reached. As a non-limiting example, such a volume is preferably within the range of 0.1 m³ to 1000 mm³, as required by the volume for the agent load. The Loder may optionally be larger, for example when incorporated with a device whose size is determined by functionality, for example and without limitation, a knee joint, an intra-uterine or cervical ring and the like.

[00543] The drug delivery system (for delivering the composition) is designed in some embodiments to preferably employ degradable polymers, wherein the main release mechanism is bulk erosion; or in some embodiments, non degradable, or slowly degraded polymers are used, wherein the main release mechanism is diffusion rather than bulk erosion, so that the outer part functions as membrane, and its internal part functions as a drug reservoir, which practically is not affected by the surroundings for an extended period (for example from about a week to about a few months). Combinations of different polymers with different release mechanisms may also optionally be used. The concentration gradient at the surface is preferably maintained effectively constant during a significant period of the total drug releasing period, and therefore the diffusion rate is effectively constant (termed "zero mode" diffusion). By the term "constant" it is meant a diffusion rate that is preferably maintained above the lower threshold of therapeutic effectiveness, but which may still optionally feature an initial burst and/or may fluctuate, for example increasing and decreasing to a certain degree. The diffusion rate is preferably so maintained for a prolonged period, and it can be considered constant to a certain level to optimize the therapeutically effective period, for example the effective silencing period.

[00544] The drug delivery system optionally and preferably is designed to shield the nucleotide based therapeutic agent from degradation, whether chemical in nature or due to attack from enzymes and other factors in the body of the subject.

[00545] The drug delivery system of US Patent Publication 201 10195123 is optionally associated with sensing and/or activation appliances that are operated at and/or after implantation of the device, by non and/or minimally invasive methods of activation and/or acceleration/deceleration, for example optionally including but not limited to thermal heating and cooling, laser beams, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices.

[00546] According to some embodiments of US Patent Publication 201 10195 123, the site for local delivery may optionally include target sites characterized by high abnormal proliferation of cells, and suppressed apoptosis, including tumors, active and or chronic

inflammation and infection including autoimmune diseases states, degenerating tissue including muscle and nervous tissue, chronic pain, degenerative sites, and location of bone fractures and other wound locations for enhancement of regeneration of tissue, and injured cardiac, smooth and striated muscle.

[00547] The site for implantation of the composition, or target site, preferably features a radius, area and/or volume that is sufficiently small for targeted local delivery. For example, the target site optionally has a diameter in a range of from about 0.1 mm to about 5 cm.

[00548] The location of the target site is preferably selected for maximum therapeutic efficacy. For example, the composition of the drug delivery system (optionally with a device for implantation as described above) is optionally and preferably implanted within or in the proximity of a tumor environment, or the blood supply associated thereof.

[00549] For example the composition (optionally with the device) is optionally implanted within or in the proximity to pancreas, prostate, breast, liver, via the nipple, within the vascular system and so forth.

[00550] The target location is optionally selected from the group comprising, consisting essentially of, or consisting of (as non-limiting examples only, as optionally any site within the body may be suitable for implanting a Loder): 1. brain at degenerative sites like in Parkinson or Alzheimer disease at the basal ganglia, white and gray matter; 2. spine as in the case of amyotrophic lateral sclerosis (ALS); 3. uterine cervix to prevent UPV infection; 4. active and chronic inflammatory joints; 5. dermis as in the case of psoriasis; 6. sympathetic and sensoric nervous sites for analgesic effect; 7. Intra osseous implantation; 8. acute and chronic infection sites; 9. Intra vaginal; 10. Inner ear—auditory system, labyrinth of the inner ear, vestibular system; 11. Intra tracheal; 12. Intra-cardiac; coronary, epicardiac; 13. urinary bladder; 14. biliary system; 15. parenchymal tissue including and not limited to the kidney, liver, spleen; 16. lymph nodes; 17. salivary glands; 18. dental gums; 19. Intra-articular (into joints); 20. Intra-ocular; 21. Brain tissue; 22. Brain ventricles; 23. Cavities, including abdominal cavity (for example but without limitation, for ovary cancer); 24. Intra esophageal and 25. Intra rectal.

[00551] Optionally insertion of the system (for example a device containing the composition) is associated with injection of material to the ECM at the target site and the vicinity of that site to affect local pH and/or temperature and/or other biological factors affecting the diffusion of the drug and/or drug kinetics in the ECM, of the target site and the vicinity of such a site.

[00552] Optionally, according to some embodiments, the release of said agent could be associated with sensing and/or activation appliances that are operated prior and/or at and/or after insertion, by non and/or minimally invasive and/or else methods of activation and/or acceleration/deceleration, including laser beam, radiation, thermal heating and cooling, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices, and chemical activators.

[00553] According to other embodiments of US Patent Publication 201 10195123, the drug preferably comprises a RNA, for example for localized cancer cases in breast, pancreas, brain, kidney, bladder, lung, and prostate as described below. Although exemplified with RNAi, many drugs are applicable to be encapsulated in Loder, and can be used in association with this invention, as long as such drugs can be encapsulated with the Loder substrate, such as a matrix for example, and this system may be used and/or adapted to deliver the nucleic acid-targeting system of the present invention.

[00554] As another example of a specific application, neuro and muscular degenerative diseases develop due to abnormal gene expression. Local delivery of RNAs may have therapeutic properties for interfering with such abnormal gene expression. Local delivery of anti apoptotic, anti inflammatory and anti degenerative drugs including small drugs and macromolecules may also optionally be therapeutic. In such cases the Loder is applied for prolonged release at constant rate and/or through a dedicated device that is implanted separately. All of this may be used and/or adapted to the nucleic acid-targeting system of the present invention.

[00555] As yet another example of a specific application, psychiatric and cognitive disorders are treated with gene modifiers. Gene knockdown is a treatment option. Loders locally delivering agents to central nervous system sites are therapeutic options for psychiatric and cognitive disorders including but not limited to psychosis, bi-polar diseases, neurotic disorders and behavioral maladies. The Loders could also deliver locally drugs including small drugs and macromolecules upon implantation at specific brain sites. All of this may be used and/or adapted to the nucleic acid-targeting system of the present invention.

[00556] As another example of a specific application, silencing of innate and/or adaptive immune mediators at local sites enables the prevention of organ transplant rejection. Local delivery of RNAs and immunomodulating reagents with the Loder implanted into the transplanted organ and/or the implanted site renders local immune suppression by repelling immune cells such as CD8 activated against the transplanted organ. All of this may be used/and or adapted to the nucleic acid-targeting system of the present invention.

[00557] As another example of a specific application, vascular growth factors including VEGFs and angiogenin and others are essential for neovascularization. Local delivery of the factors, peptides, peptidomimetics, or suppressing their repressors is an important therapeutic modality; silencing the repressors and local delivery of the factors, peptides, macromolecules and small drugs stimulating angiogenesis with the Loder is therapeutic for peripheral, systemic and cardiac vascular disease.

[00558] The method of insertion, such as implantation, may optionally already be used for other types of tissue implantation and/or for insertions and/or for sampling tissues, optionally without modifications, or alternatively optionally only with non-major modifications in such methods. Such methods optionally include but are not limited to brachytherapy methods, biopsy, endoscopy with and/or without ultrasound, such as ERCP, stereotactic methods into the brain tissue, Laparoscopy, including implantation with a laparoscope into joints, abdominal organs, the bladder wall and body cavities.

[00559] Implantable device technology herein discussed can be employed with herein teachings and hence by this disclosure and the knowledge in the art, CRISPR-Cas system or components thereof or nucleic acid molecules thereof or encoding or providing components may be delivered via an implantable device.

Patient-specific screening methods

[00560] A nucleic acid-targeting system that targets RNA, e.g., trinucleotide repeats can be used to screen patients or patient samples for the presence of such repeats. The repeats can be the target of the RNA of the nucleic acid-targeting system, and if there is binding thereto by the nucleic acid-targeting system, that binding can be detected, to thereby indicate that such a repeat is present. Thus, a nucleic acid-targeting system can be used to screen patients or patient samples for the presence of the repeat. The patient can then be administered suitable compound(s) to address the condition; or, can be administered a nucleic acid-targeting system to bind to and cause insertion, deletion or mutation and alleviate the condition.

[00561] The invention uses nucleic acids to bind target RNA sequences.

CRISPR effector protein mRNA and guide RNA

[00562] CRISPR effector protein mRNA and guide RNA might also be delivered separately. CRISPR effector protein mRNA can be delivered prior to the guide RNA to give time for CRISPR effector protein to be expressed. CRISPR effector protein mRNA might be

administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA.

[00563] Alternatively, CRISPR effector protein mRNA and guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of CRISPR effector protein mRNA + guide RNA.

[00564] The CRISPR effector protein of the present invention, i.e. a Cas13 effector protein is sometimes referred to herein as a CRISPR Enzyme. It will be appreciated that the effector protein is based on or derived from an enzyme, so the term 'effector protein' certainly includes 'enzyme' in some embodiments. However, it will also be appreciated that the effector protein may, as required in some embodiments, have DNA or RNA binding, but not necessarily cutting or nicking, activity, including a dead-Cas effector protein function.

[00565] Additional administrations of CRISPR effector protein mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification. In some embodiments, phenotypic alteration is preferably the result of genome modification when a genetic disease is targeted, especially in methods of therapy and preferably where a repair template is provided to correct or alter the phenotype.

[00566] In some embodiments diseases that may be targeted include those concerned with disease-causing splice defects.

[00567] In some embodiments, cellular targets include Hemopoietic Stem/Progenitor Cells (CD34+); Human T cells; and Eye (retinal cells) - for example photoreceptor precursor cells.

[00568] In some embodiments Gene targets include: Human Beta Globin - HBB (for treating Sickle Cell Anemia, including by stimulating gene-conversion (using closely related HBD gene as an endogenous template)); CD3 (T-Cells); and CEP920 - retina (eye).

[00569] In some embodiments disease targets also include: cancer; Sickle Cell Anemia (based on a point mutation); HIV; Beta-Thalassemia; and ophthalmic or ocular disease - for example Leber Congenital Amaurosis (LCA)-causing Splice Defect.

[00570] In some embodiments delivery methods include: Cationic Lipid Mediated "direct" delivery of Enzyme-Guide complex (RiboNucleoProtein) and electroporation of plasmid DNA.

[00571] Inventive methods can further comprise delivery of templates, such as repair templates, which may be dsODN or ssODN, see below. Delivery of templates may be via the contemporaneous or separate from delivery of any or all the CRISPR effector protein or guide and via the same delivery mechanism or different. In some embodiments, it is preferred that

the template is delivered together with the guide, and, preferably, also the CRISPR effector protein. An example may be an AAV vector.

[00572] Inventive methods can further comprise: (a) delivering to the cell a double-stranded oligodeoxynucleotide (dsODN) comprising overhangs complimentary to the overhangs created by said double strand break, wherein said dsODN is integrated into the locus of interest; or -(b) delivering to the cell a single-stranded oligodeoxynucleotide (ssODN), wherein said ssODN acts as a template for homology directed repair of said double strand break. Inventive methods can be for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest. Inventive methods can be conducted *in vivo* in the individual or *ex vivo* on a cell taken from the individual, optionally wherein said cell is returned to the individual.

[00573] For minimization of toxicity and off-target effect, it will be important to control the concentration of CRISPR effector protein mRNA and guide RNA delivered. Optimal concentrations of CRISPR effector protein mRNA and guide RNA can be determined by testing different concentrations in a cellular or animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci. For example, for the guide sequence targeting 5'-GAGTCCGAGC AGAAGAAGAA-3' (SEQ ID NO. 107) in the EMX1 gene of the human genome, deep sequencing can be used to assess the level of modification at the following two off-target loci, 1: 5'-GAGTCCTAGCAGGAGAAGAA-3' (SEQ ID NO. 108) and 2: 5'-GAGTCTAAGC AGAAGAAGAA-3' (SEQ ID NO. 109). The concentration that gives the highest level of on-target modification while minimizing the level of off-target modification should be chosen for *in vivo* delivery.

Inducible Systems

[00574] In some embodiments, a CRISPR effector protein may form a component of an inducible system. The inducible nature of the system would allow for spatiotemporal control of gene editing or gene expression using a form of energy. The form of energy may include but is not limited to electromagnetic radiation, sound energy, chemical energy and thermal energy. Examples of inducible system include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome). In one embodiment, the CRISPR effector protein may be a part of a Light Inducible Transcriptional Effector (LITE) to direct changes in transcriptional activity in a sequence-specific manner. The components of a light may include a CRISPR effector protein, a light-responsive cytochrome heterodimer (e.g. from *Arabidopsis thaliana*), and a transcriptional activation/repression

domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283, and WO 2014018423 A2 which is hereby incorporated by reference in its entirety.

Exemplary Methods of Using of CRISPR Cas System

[00575] The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell *in vivo*, *ex vivo* or *in vitro* and, may be conducted in a manner alters the cell such that once modified the progeny or cell line of the CRISPR modified cell retains the altered phenotype. The modified cells and progeny may be part of a multi-cellular organism such as a plant or animal with *ex vivo* or *in vivo* application of CRISPR system to desired cell types. The CRISPR invention may be a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

Modifying a Target with CRISPR Cas System or Complex (e.g., Cas13-RNA Complex)

[00576] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be *in vivo*, *ex vivo* or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[00577] In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR effector protein complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide.

[00578] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR effector protein complexed with a guide sequence hybridized or hybridizable to a target sequence within said polynucleotide. Similar considerations and conditions apply as

above for methods of modifying a target polynucleotide. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention.

[00579] Indeed, in any aspect of the invention, the CRISPR complex may comprise a CRISPR effector protein complexed with a guide sequence hybridized or hybridizable to a target sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide.

[00580] Thus in any of the non-naturally-occurring CRISPR effector proteins described herein comprise at least one modification and whereby the effector protein has certain improved capabilities. In particular, any of the effector proteins are capable of forming a CRISPR complex with a guide RNA. When such a complex forms, the guide RNA is capable of binding to a target polynucleotide sequence and the effector protein is capable of modifying a target locus. In addition, the effector protein in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme/effector protein.

[00581] In addition, the modified CRISPR enzymes described herein encompass enzymes whereby in the CRISPR complex the effector protein has increased capability of modifying the one or more target loci as compared to an unmodified enzyme/effector protein. Such function may be provided separate to or provided in combination with the above-described function of reduced capability of modifying one or more off-target loci. Any such effector proteins may be provided with any of the further modifications to the CRISPR effector protein as described herein, such as in combination with any activity provided by one or more associated heterologous functional domains, any further mutations to reduce nuclease activity and the like.

[00582] In advantageous embodiments of the invention, the modified CRISPR effector protein is provided with reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme/effector protein and increased capability of modifying the one or more target loci as compared to an unmodified enzyme/effector protein. In combination with further modifications to the effector protein, significantly enhanced specificity may be achieved. For example, combination of such advantageous embodiments with one or more additional mutations is provided wherein the one or more additional mutations are in one or more catalytically active domains. In such effector proteins, enhanced specificity may be achieved due to an improved specificity in terms of effector protein activity.

[00583] Additional functionalities which may be engineered into modified CRISPR effector proteins as described herein include the following. 1. modified CRISPR effector proteins that disrupt RNA:protein interactions without affecting protein tertiary or secondary structure. This includes residues that contact any part of the RNA:RNA duplex. 2. modified CRISPR effector proteins that weaken intra-protein interactions holding Cas3 in conformation essential for nuclease cutting in response to RNA binding (on or off target). For example: a modification that mildly inhibits, but still allows, the nuclease conformation of the HNH domain (positioned at the scissile phosphate). 3. modified CRISPR effector proteins that strengthen intra-protein interactions holding Cas3 in a conformation inhibiting nuclease activity in response to RNA binding (on or off targets). For example: a modification that stabilizes the HNH domain in a conformation away from the scissile phosphate. Any such additional functional enhancement may be provided in combination with any other modification to the CRISPR effector protein as described in detail elsewhere herein.

[00584] Any of the herein described improved functionalities may be made to any CRISPR effector protein, such as a Cas3 effector protein. However, it will be appreciated that any of the functionalities described herein may be engineered into Cas3 effector proteins from other orthologs, including chimeric effector proteins comprising fragments from multiple orthologs.

[00585] The invention uses nucleic acids to bind target DNA sequences. This is advantageous as nucleic acids are much easier and cheaper to produce than proteins, and the specificity can be varied according to the length of the stretch where homology is sought. Complex 3-D positioning of multiple fingers, for example is not required. The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan,

1993; WO 97/0321 1; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. A "wild type" can be a base line. As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature. The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions. As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), *Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid*

Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Where reference is made to a polynucleotide sequence, then complementary or partially complementary sequences are also envisaged. These are preferably capable of hybridizing to the reference sequence under highly stringent conditions. Generally, in order to maximize the hybridization rate, relatively low-stringency hybridization conditions are selected: about 20 to 25° C lower than the thermal melting point (T_m). The T_m is the temperature at which 50% of specific target sequence hybridizes to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridized sequences, highly stringent washing conditions are selected to be about 5 to 15° C lower than the T_m . In order to require at least about 70% nucleotide complementarity of hybridized sequences, moderately-stringent washing conditions are selected to be about 15 to 30° C lower than the T_m . Highly permissive (very low stringency) washing conditions may be as low as 50° C below the T_m , allowing a high level of mis-matching between hybridized sequences. Those skilled in the art will recognize that other physical and chemical parameters in the hybridization and wash stages can also be altered to affect the outcome of a detectable hybridization signal from a specific level of homology between target and probe sequences. Preferred highly stringent conditions comprise incubation in 50% formamide, 5×SSC, and 1% SDS at 42° C, or incubation in 5×SSC and 1% SDS at 65° C, with wash in 0.2×SSC and 0.1% SDS at 65° C. "Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence. As used herein, the term "genomic locus" or "locus" (plural loci) is the specific location of a gene or DNA sequence on a chromosome. A "gene" refers to stretches of DNA or RNA that encode a polypeptide or an RNA chain that has functional role to play in an organism and hence is the molecular unit of heredity in living organisms. For the purpose of this invention it may be considered that genes include regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed

sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. As used herein, "expression of a genomic locus" or "gene expression" is the process by which information from a gene is used in the synthesis of a functional gene product. The products of gene expression are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is functional RNA. The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses to generate functional products to survive. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context. As used herein, "expression" also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. As used herein, the term "domain" or "protein domain" refers to a part of a protein sequence that may exist and function independently of the rest of the protein chain. As described in aspects of the invention, sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences.

[00586] In aspects of the invention the term "guide RNA", refers to the polynucleotide sequence comprising one or more of a putative or identified tracr sequence and a putative or

identified crRNA sequence or guide sequence. In particular embodiments, the "guide RNA" comprises a putative or identified crRNA sequence or guide sequence. In further embodiments, the guide RNA does not comprise a putative or identified tracr sequence.

[00587] As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. A "wild type" can be a base line.

[00588] As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature.

[00589] The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. In all aspects and embodiments, whether they include these terms or not, it will be understood that, preferably, they may be optional and thus preferably included or not preferably not included. Furthermore, the terms "non-naturally occurring" and "engineered" may be used interchangeably and so can therefore be used alone or in combination and one or other may replace mention of both together. In particular, "engineered" is preferred in place of "non-naturally occurring" or "non-naturally occurring and/or engineered."

[00590] Sequence homologies may be generated by any of a number of computer programs known in the art, for example BLAST or FASTA, etc. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A; Devereux et al., 1984, *Nucleic Acids Research* 12:387). Examples of other software that may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *ibid* - Chapter 18), FASTA (Atschul et al., 1990, *J. Mol. Biol.*, 403-410) and the GENWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999 *ibid*, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program. Percentage (%) sequence homology may be calculated over contiguous sequences, i.e., one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues. Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise

identical pair of sequences, one insertion or deletion may cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without unduly penalizing the overall homology or identity score. This is achieved by inserting "gaps" in the sequence alignment to try to maximize local homology or identity. However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - may achieve a higher score than one with many gaps. "Affinity gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties may, of course, produce optimized alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example, when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension. Calculation of maximum % homology therefore first requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (Devereux et al., 1984 *Nuc. Acids Research* 12 p387). Examples of other software that may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *Short Protocols in Molecular Biology*, 4th Ed. - Chapter 18), FASTA (Altschul et al., 1990 *J Mol. Biol.* 403-410) and the GENWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999, *Short Protocols in Molecular Biology*, pages 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequences (see *FEMS Microbiol Lett.* 1999 174(2): 247-50; *FEMS Microbiol Lett.* 1999 177(1): 187-8 and the website of the National Center for Biotechnology information at the website of the National Institutes for Health). Although the final % homology may be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pair-wise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the

default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table, if supplied (see user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62. Alternatively, percentage homologies may be calculated using the multiple alignment feature in DNASIS™ (Hitachi Software), based on an algorithm, analogous to CLUSTAL (Higgins DG & Sharp PM (1988), *Gene* 73(1), 237-244). Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result. The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in amino acid properties (such as polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues) and it is therefore useful to group amino acids together in functional groups. Amino acids may be grouped together based on the properties of their side chains alone. However, it is more useful to include mutation data as well. The sets of amino acids thus derived are likely to be conserved for structural reasons. These sets may be described in the form of a Venn diagram (Livingstone CD. and Barton G.J. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" *Comput. Appl. Biosci.* 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" *J. Theor. Biol.* 119; 205-218). Conservative substitutions may be made, for example according to the table below which describes a generally accepted Venn diagram grouping of amino acids.

Set		Sub-set	
Hydrophobic	F W Y H K M I L V A G C (SEQ ID No. 110)	Aromatic	F W Y H (SEQ ID No. 113)
		Aliphatic	I L V

Polar	W Y H K R E D C S T N Q (SEQ ID No. I11)	Charged	H K R E D (SEQ ID No. 114)
		Positively charged	H K R
		Negatively charged	E D
Small	V C A G S P T N D (SEQ ID No. 112)	Tiny	A G S

[00591] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[00592] The terms "therapeutic agent", "therapeutic capable agent" or "treatment agent" are used interchangeably and refer to a molecule or compound that confers some beneficial effect upon administration to a subject. The beneficial effect includes enablement of diagnostic determinations; amelioration of a disease, symptom, disorder, or pathological condition; reducing or preventing the onset of a disease, symptom, disorder or condition; and generally counteracting a disease, symptom, disorder or pathological condition.

[00593] As used herein, "treatment" or "treating," or "palliating" or "ameliorating" are used interchangeably. These terms refer to an approach for obtaining beneficial or desired results including but not limited to a therapeutic benefit and/or a prophylactic benefit. By therapeutic benefit is meant any therapeutically relevant improvement in or effect on one or more diseases, conditions, or symptoms under treatment. For prophylactic benefit, the compositions may be administered to a subject at risk of developing a particular disease, condition, or symptom, or to a subject reporting one or more of the physiological symptoms of a disease, even though the disease, condition, or symptom may not have yet been manifested.

[00594] The term "effective amount" or "therapeutically effective amount" refers to the amount of an agent that is sufficient to effect beneficial or desired results. The therapeutically effective amount may vary depending upon one or more of: the subject and disease condition being treated, the weight and age of the subject, the severity of the disease

condition, the manner of administration and the like, which can readily be determined by one of ordinary skill in the art. The term also applies to a dose that will provide an image for detection by any one of the imaging methods described herein. The specific dose may vary depending on one or more of: the particular agent chosen, the dosing regimen to be followed, whether it is administered in combination with other compounds, timing of administration, the tissue to be imaged, and the physical delivery system in which it is carried.

[00595] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See Sambrook, Fritsch and Maniatis, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd edition (1989); *CURRENT PROTOCOLS IN MOLECULAR BIOLOGY* (F. M. Ausubel, et al. eds., (1987)); the series *METHODS IN ENZYMOLOGY* (Academic Press, Inc.): *PCR 2: A PRACTICAL APPROACH* (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) *ANTIBODIES, A LABORATORY MANUAL*, and *ANIMAL CELL CULTURE* (R.I. Freshney, ed. (1987)).

[00596] Several aspects of the invention relate to vector systems comprising one or more vectors, or vectors as such. Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY* 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[00597] Embodiments of the invention include sequences (both polynucleotide or polypeptide) which may comprise homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue or nucleotide, with an alternative residue or nucleotide) that may occur i.e., like-for-like substitution in the case of amino acids such as basic for basic, acidic for acidic, polar for polar, etc. Non-homologous substitution may also occur i.e., from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyrilylalanine, thienylalanine, naphthylalanine and phenylglycine. Variant amino acid sequences may include suitable spacer groups that may be

inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β -alanine residues. A further form of variation, which involves the presence of one or more amino acid residues in peptoid form, may be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α -carbon substituent group is on the residue's nitrogen atom rather than the α -carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon RJ et al., *PNAS* (1992) 89(20), 9367-9371 and Horwell DC, *Trends Biotechnol.* (1995) 13(4), 132-134.

[00598] Homology modelling: Corresponding residues in other Cas13 orthologs can be identified by the methods of Zhang et al., 2012 (*Nature*; 490(7421): 556-60) and Chen et al., 2015 (*PLoS Comput Biol*; 11(5): e1004248) —a computational protein-protein interaction (PPI) method to predict interactions mediated by domain-motif interfaces. PrePPI (Predicting PPI), a structure based PPI prediction method, combines structural evidence with non-structural evidence using a Bayesian statistical framework. The method involves taking a pair a query proteins and using structural alignment to identify structural representatives that correspond to either their experimentally determined structures or homology models. Structural alignment is further used to identify both close and remote structural neighbors by considering global and local geometric relationships. Whenever two neighbors of the structural representatives form a complex reported in the Protein Data Bank, this defines a template for modelling the interaction between the two query proteins. Models of the complex are created by superimposing the representative structures on their corresponding structural neighbor in the template. This approach is further described in Dey et al., 2013 (*Prot Sci*; 22: 359-66).

[00599] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E.coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR.

[00600] In certain aspects the invention involves vectors. A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is

capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00601] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171 156 A1, the contents of which are herein incorporated by reference in their entirety.

[00602] Aspects of the invention relate to bicistronic vectors for guide RNA and wild type, modified or mutated CRISPR effector proteins/enzymes (e.g. Cas13). Bicistronic expression vectors guide RNA and wild type, modified or mutated CRISPR effector proteins/enzymes

(e.g. Cas13) are preferred. In general and particularly in this embodiment and wild type, modified or mutated CRISPR effector proteins/enzymes (e.g. Cas13) is preferably driven by the CBh promoter. The RNA may preferably be driven by a Pol III promoter, such as a U6 promoter. Ideally the two are combined.

[00603] In some embodiments, a loop in the guide RNA is provided. This may be a stem loop or a tetra loop. The loop is preferably GAAA, but it is not limited to this sequence or indeed to being only 4bp in length. Indeed, preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG.

[00604] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection.

[00605] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent

manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and HI promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). With regards to regulatory sequences, mention is made of U.S. patent application 10/491,026, the contents of which are incorporated by reference herein in their entirety. With regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application 12/511,940, the contents of which are incorporated by reference herein in their entirety.

[00606] Vectors can be designed for expression of CRISPR transcripts (e.g., nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[00607] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a

eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g., amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[00608] Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

[00609] In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, et al., 1987. *EMBO J.* 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz et al., 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif), and picZ (InVitrogen Corp, San Diego, Calif).

[00610] In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. *Mol. Cell Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

[00611] In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of

mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., *MOLECULAR CLONING: A LABORATORY MANUAL*. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY., 1989.

[00612] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety.

[00613] In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic

Repeats), also known as SPIDRs (SPacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in *E. coli* (Ishino et al., *J. Bacterid.*, 169:5429-5433 [1987]; and Nakata et al., *J. Bacterid.*, 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in *Haloferax mediterranei*, *Streptococcus pyogenes*, *Anabaena*, and *Mycobacterium tuberculosis* (See, Groenen et al., *Mol. Microbiol.*, 10:1057-1065 [1993]; Hoe et al., *Emerg. Infect. Dis.*, 5:254-263 [1999]; Masepohl et al., *Biochim. Biophys. Acta* 1307:26-30 [1996]; and Mojica et al., *Mol. Microbiol.*, 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al., *OMICS J. Integ. Biol.*, 6:23-33 [2002]; and Mojica et al., *Mol. Microbiol.*, 36:244-246 [2000]). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., *J. Bacterid.*, 182:2393-2401 [2000]). CRISPR loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., *Mol. Microbiol.*, 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to *Aeropyrum*, *Pyrobaculum*, *Sulfolobus*, *Archaeoglobus*, *Halocarcula*, *Methanobacterium*, *Methanococcus*, *Methanosarcina*, *Methanopyrus*, *Pyrococcus*, *Picrophilus*, *Thermoplasma*, *Corynebacterium*, *Mycobacterium*, *Streptomyces*, *Aquifex*, *Porphyromonas*, *Chlorobium*, *Thermus*, *Bacillus*, *Listeria*, *Staphylococcus*, *Clostridium*, *Thermoanaerobacter*, *Mycoplasma*, *Fusobacterium*, *Azarcus*, *Chromobacterium*, *Neisseria*, *Nitrosomonas*, *Desulfovibrio*, *Geobacter*, *Myxococcus*, *Campylobacter*, *Wolinella*, *Acinetobacter*, *Erwinia*, *Escherichia*, *Legionella*, *Methylococcus*, *Pasteurella*, *Photobacterium*, *Salmonella*, *Xanthomonas*, *Yersinia*, *Treponema*, and *Thermotoga*.

[00614] In general, "nucleic acid-targeting system" as used in the present application refers collectively to transcripts and other elements involved in the expression of or directing the activity of nucleic acid-targeting CRISPR-associated ("Cas") genes (also referred to herein as an effector protein), including sequences encoding a nucleic acid-targeting Cas (effector) protein and a guide RNA (comprising crRNA sequence and a trans-activating CRISPR/Cas system RNA (tracrRNA) sequence), or other sequences and transcripts from a nucleic acid-targeting CRISPR locus. In some embodiments, one or more elements of a

nucleic acid-targeting system are derived from a Type V/Type VI nucleic acid-targeting CRISPR system. In some embodiments, one or more elements of a nucleic acid-targeting system is derived from a particular organism comprising an endogenous nucleic acid-targeting CRISPR system. In general, a nucleic acid-targeting system is characterized by elements that promote the formation of a nucleic acid-targeting complex at the site of a target sequence. In the context of formation of a nucleic acid-targeting complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide RNA promotes the formation of a DNA or RNA-targeting complex. Full complementarity is not necessarily required, provided there is sufficient complementarity to cause hybridization and promote formation of a nucleic acid-targeting complex. A target sequence may comprise RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. In some embodiments, the target sequence may be within an organelle of a eukaryotic cell, for example, mitochondrion or chloroplast. A sequence or template that may be used for recombination into the targeted locus comprising the target sequences is referred to as an "editing template" or "editing RNA" or "editing sequence". In aspects of the invention, an exogenous template RNA may be referred to as an editing template. In an aspect of the invention the recombination is homologous recombination.

[00615] Typically, in the context of an endogenous nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in cleavage of one or both RNA strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. In some embodiments, one or more vectors driving expression of one or more elements of a nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct formation of a nucleic acid-targeting complex at one or more target sites. For example, a nucleic acid-targeting effector protein and a guide RNA could each be operably linked to separate regulatory elements on separate vectors. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the nucleic acid-targeting system not included in the first vector. nucleic acid-targeting system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or

opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-targeting effector protein and a guide RNA embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the nucleic acid-targeting effector protein and guide RNA are operably linked to and expressed from the same promoter.

[00616] In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies, ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. The ability of a guide sequence to direct sequence-specific binding of a nucleic acid-targeting complex to a target sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting CRISPR sequence, followed by an assessment of preferential cleavage within or in the vicinity of the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between

the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00617] A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a gene transcript or mRNA.

[00618] In some embodiments, the target sequence is a sequence within a genome of a cell.

[00619] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide sequence. Secondary structure may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (*Nucleic Acids Res.* 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A.R. Gruber *et al.*, 2008, *Cell* 106(1): 23-24; and PA Carr and GM Church, 2009, *Nature Biotechnology* 27(12): 1151-62). Further algorithms may be found in U.S. application Serial No. TBA (attorney docket 44790.1 1.2022; Broad Reference BI-20 13/004A); incorporated herein by reference.

[00620] In some embodiments, the nucleic acid-targeting effector protein is part of a fusion protein comprising one or more heterologous protein domains (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the nucleic acid-targeting effector protein). In some embodiments, the CRISPR effector protein/enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR effector protein/enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to an effector protein include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase,

green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A nucleic acid-targeting effector protein may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a nucleic acid-targeting effector protein are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged nucleic acid-targeting effector protein is used to identify the location of a target sequence.

[00621] In some embodiments, a CRISPR enzyme may form a component of an inducible system. The inducible nature of the system would allow for spatiotemporal control of gene editing or gene expression using a form of energy. The form of energy may include but is not limited to electromagnetic radiation, sound energy, chemical energy and thermal energy. Examples of inducible system include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cytochrome). In one embodiment, the CRISPR enzyme may be a part of a Light Inducible Transcriptional Effector (LITE) to direct changes in transcriptional activity in a sequence-specific manner. The components of a light may include a CRISPR enzyme, a light-responsive cytochrome heterodimer (e.g. from *Arabidopsis thaliana*), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283 and WO 2014/018423 and US8889418, US8895308, US20140186919, US20140242700, US20140273234, US20140335620, WO2014093635, which is hereby incorporated by reference in its entirety.

[00622] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and organisms (such as animals, plants, or fungi) comprising or produced from such cells. In some embodiments, a nucleic acid-targeting effector protein in combination with (and optionally complexed with) a guide RNA is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a nucleic acid-

targeting system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, *Science* 256:808-813 (1992); Nabel & Feigner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada et al., in *Current Topics in Microbiology and Immunology*, Doerfler and Bohm (eds) (1995); and Yu et al., *Gene Therapy* 1:13-26 (1994).

[00623] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[00624] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., *Crystal*, *Science* 270:404-410 (1995); Blaese et al., *Cancer Gene Ther.* 2:291-297 (1995); Behr et al., *Bioconjugate Chem.* 5:382-389 (1994); Remy et al., *Bioconjugate Chem.* 5:647-654 (1994); Gao et al., *Gene Therapy* 2:710-722 (1995); Ahmad et al., *Cancer Res.* 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[00625] The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of

the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[00626] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., *J. Virol.* 66:2731-2739 (1992); Johann et al., *J. Virol.* 66:1635-1640 (1992); Somnerfelt et al., *Virol.* 176:58-59 (1990); Wilson et al., *J. Virol.* 63:2374-2378 (1989); Miller et al., *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700). In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989).

Models of Genetic and Epigenetic Conditions

[00627] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid

sequences associated with a disease, or a plant, animal or cell in which the expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[00628] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[00629] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[00630] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a CRISPR enzyme, and a direct repeat sequence linked to a guide sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[00631] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study

the effects of a genome sequence modified by the CRISPR complex of the invention on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[00632] Several disease models have been specifically investigated. These include *de novo* autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but serve to show the broad applicability of the invention across genes and corresponding models.

[00633] An altered expression of one or more genome sequences associated with a signalling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the

corresponding genes between the test model cell and a control cell, when they are contacted with

a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[00634] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[00635] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E.coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase

chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[00636] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[00637] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Patent No. 5,210,015.

[00638] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[00639] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for

example, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Patent No. 5,445,934.

[00640] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, β -galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[00641] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphorimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[00642] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent-protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[00643] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the

complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agentpolypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and hence generating a detectable signal.

[00644] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[00645] The amount of agentpolypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agentpolypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[00646] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[00647] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example, anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Antiphosphotyrosine antibodies are particularly useful in detecting proteins

that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2a). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

[00648] In practicing the subject method, it may be desirable to discern the expression pattern of a protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[00649] An altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a signaling biochemical pathway will depend on the biological activity and/or the signal transduction pathway that is under investigation. For example, where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays such as AlphaScreen™ (available from Perkin Elmer) and eTag™ assay (Chan-Hui, et al. (2003) *Clinical Immunology* 111: 162-174).

[00650] Where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPR™ (Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a minisecond.

[00651] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection.

[00652] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA).

[00653] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

[00654] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent

motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme.

[00655] The target polynucleotide of a CRISPR complex may include a number of disease associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in U.S. provisional patent applications 61/736,527 and 61/748,427 both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on December 12, 2012 and January 2, 2013, respectively, and PCT Application PCT/US20 13/074667, entitled DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS, filed December 12, 2013, the contents of all of which are herein incorporated by reference in their entirety.

[00656] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

Transcriptome Wide Knock-down Screening

[00657] The CRISPR effector protein complexes described herein can be used to perform efficient and cost effective functional transcriptomic screens. Such screens can utilize CRISPR effector protein based transcriptome wide libraries. Such screens and libraries can provide for determining the function of genes, cellular pathways genes are involved in, and how any alteration in gene expression can result in a particular biological

process. An advantage of the present invention is that the CRISPR system avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA. In preferred embodiments of the invention, the CRISPR effector protein complexes are Cas13 effector protein complexes.

[00658] In embodiments of the invention, a transcriptome wide library may comprise a plurality of Cas13 guide RNAs, as described herein, comprising guide sequences that are capable of targeting a plurality of target sequences in a plurality of loci in a population of eukaryotic cells. The population of cells may be a population of embryonic stem (ES) cells. The target sequence in the locus may be a non-coding sequence. The non-coding sequence may be an intron, regulatory sequence, splice site, 3' UTR, 5' UTR, or polyadenylation signal. Gene function of one or more gene products may be altered by said targeting. The targeting may result in a knockout of gene function. The targeting of a gene product may comprise more than one guide RNA. A gene product may be targeted by 2, 3, 4, 5, 6, 7, 8, 9, or 10 guide RNAs, preferably 3 to 4 per gene. Off-target modifications may be minimized by exploiting the staggered double strand breaks generated by Cas13 effector protein complexes or by utilizing methods analogous to those used in CRISPR-Cas9 systems (See, e.g., DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013)), incorporated herein by reference. The targeting may be of about 100 or more sequences. The targeting may be of about 1000 or more sequences. The targeting may be of about 20,000 or more sequences. The targeting may be of the entire genome. The targeting may be of a panel of target sequences focused on a relevant or desirable pathway. The pathway may be an immune pathway. The pathway may be a cell division pathway.

[00659] One aspect of the invention comprehends a transcriptome wide library that may comprise a plurality of Cas13 guide RNAs that may comprise guide sequences that are capable of targeting a plurality of target sequences in a plurality of loci, wherein said targeting results in a knockdown of gene function. This library may potentially comprise guide RNAs that target each and every gene in the genome of an organism.

[00660] In some embodiments of the invention the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In some embodiments, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of

the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the organism or subject is algae, including microalgae, or is a fungus.

[00661] The knockdown of gene function may comprise: introducing into each cell in the population of cells a vector system of one or more vectors comprising an engineered, non-naturally occurring Cas13 effector protein system comprising I. a Cas13 effector protein, and II. one or more guide RNAs, wherein components I and II may be same or on different vectors of the system, integrating components I and II into each cell, wherein the guide sequence targets a unique gene in each cell, wherein the Cas13 effector protein is operably linked to a regulatory element, wherein when transcribed, the guide RNA comprising the guide sequence directs sequence-specific binding of the Cas13 effector protein system to a target sequence in the genomic loci of the unique gene, inducing cleavage of the genomic loci by the Cas13 effector protein, and confirming different knockdown events in a plurality of unique genes in each cell of the population of cells thereby generating a gene knockdown cell library. The invention comprehends that the population of cells is a population of eukaryotic cells, and in a preferred embodiment, the population of cells is a population of embryonic stem (ES) cells.

[00662] The one or more vectors may be plasmid vectors. The vector may be a single vector comprising a Cas13 effector protein, a sgRNA, and optionally, a selection marker into target cells. Not being bound by a theory, the ability to simultaneously deliver a Cas13 effector protein and sgRNA through a single vector enables application to any cell type of interest, without the need to first generate cell lines that express the Cas13 effector protein. The regulatory element may be an inducible promoter. The inducible promoter may be a doxycycline inducible promoter. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter and is driven by the expression of T7 polymerase. The confirming of different knockdown events may be by whole transcriptome sequencing. The knockdown event may be achieved in 100 or more unique genes. The knockdown event may be achieved in 1000 or more unique genes. The knockdown event may be achieved in 20,000 or more unique genes. The knockdown event may be achieved in the entire transcriptome. The knockdown of gene function may be achieved in a plurality of unique genes which function in a particular physiological

pathway or condition. The pathway or condition may be an immune pathway or condition. The pathway or condition may be a cell division pathway or condition.

[00663] The invention also provides kits that comprise the transcriptome wide libraries mentioned herein. The kit may comprise a single container comprising vectors or plasmids comprising the library of the invention. The kit may also comprise a panel comprising a selection of unique Cas13 effector protein system guide RNAs comprising guide sequences from the library of the invention, wherein the selection is indicative of a particular physiological condition. The invention comprehends that the targeting is of about 100 or more sequences, about 1000 or more sequences or about 20,000 or more sequences or the entire transcriptome. Furthermore, a panel of target sequences may be focused on a relevant or desirable pathway, such as an immune pathway or cell division.

[00664] In an additional aspect of the invention, the Cas13 effector protein may comprise one or more mutations and may be used as a generic RNA binding protein with or without fusion to a functional domain. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations. The mutations have been characterized as described herein. In one aspect of the invention, the functional domain may be a transcriptional activation domain, which may be VP64. In other aspects of the invention, the functional domain may be a transcriptional repressor domain, which may be KRAB or SID4X. Other aspects of the invention relate to the mutated Cas13 effector protein being fused to domains which include but are not limited to a transcriptional activator, repressor, a recombinase, a transposase, a histone remodeler, a demethylase, a DNA methyltransferase, a cryptochrome, a light inducible/controllable domain or a chemically inducible/controllable domain. Some methods of the invention can include inducing expression of targeted genes. In one embodiment, inducing expression by targeting a plurality of target sequences in a plurality of genomic loci in a population of eukaryotic cells is by use of a functional domain.

[00665] Useful in the practice of the instant invention utilizing Cas13 3effector protein complexes are methods used in CRISPR-Cas9 systems and reference is made to:

[00666] Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [Epub ahead of print]; Published in final edited form as: Science. 2014 Jan 3; 343(6166): 84-87.

[00667] Shalem et al. involves a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both

negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED 12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.

[00668] Reference is also made to US patent publication number US20140357530; and PCT Patent Publication WO20 14093 701, hereby incorporated herein by reference.

Functional Alteration and Screening

[00669] In another aspect, the present invention provides for a method of functional evaluation and screening of genes. The use of the CRISPR system of the present invention to precisely deliver functional domains, to activate or repress genes or to alter epigenetic state by precisely altering the methylation site on a specific locus of interest, can be with one or more guide RNAs applied to a single cell or population of cells or with a library applied to genome in a pool of cells *ex vivo* or *in vivo* comprising the administration or expression of a library comprising a plurality of guide RNAs (sgRNAs) and wherein the screening further comprises use of a Cas13 effector protein, wherein the CRISPR complex comprising the Cas13 effector protein is modified to comprise a heterologous functional domain. In an aspect the invention provides a method for screening a genome/transcriptome comprising the administration to a host or expression in a host *in vivo* of a library. In an aspect the invention provides a method as herein discussed further comprising an activator administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a Cas13 effector protein. In an aspect the invention provides a method as herein discussed wherein the activator is attached to the N terminus or the C terminus of the Cas13 effector protein. In an aspect the invention provides a method as herein discussed wherein the activator is attached to a sgRNA loop. In an aspect the invention provides a method as herein discussed further comprising a repressor administered to the host or expressed in the host. In an aspect the invention provides a method as herein discussed, wherein the screening comprises affecting and detecting gene activation, gene inhibition, or cleavage in the locus.

[00670] In an aspect, the invention provides efficient on-target activity and minimizes off target activity. In an aspect, the invention provides efficient on-target cleavage by Cas13 effector protein and minimizes off-target cleavage by the Cas13 effector protein. In an aspect, the invention provides guide specific binding of Cas13 effector protein at a gene locus without DNA cleavage. Accordingly, in an aspect, the invention provides target-specific gene regulation. In an aspect, the invention provides guide specific binding of Cas13 effector protein at a gene locus without DNA cleavage. Accordingly, in an aspect, the invention provides for cleavage at one locus and gene regulation at a different locus using a single Cas13 effector protein. In an aspect, the invention provides orthogonal activation and/or inhibition and/or cleavage of multiple targets using one or more Cas13 effector protein and/or enzyme.

[00671] In an aspect the invention provides a method as herein discussed, wherein the host is a eukaryotic cell. In an aspect the invention provides a method as herein discussed, wherein the host is a mammalian cell. In an aspect the invention provides a method as herein discussed, wherein the host is a non-human eukaryote. In an aspect the invention provides a method as herein discussed, wherein the non-human eukaryote is a non-human mammal. In an aspect the invention provides a method as herein discussed, wherein the non-human mammal is a mouse. An aspect the invention provides a method as herein discussed comprising the delivery of the Cas13 effector protein complexes or component(s) thereof or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed *in vivo*. In an aspect the invention provides a method as herein discussed wherein the expressing *in vivo* is via a lentivirus, an adenovirus, or an AAV. In an aspect the invention provides a method as herein discussed wherein the delivery is via a particle, a nanoparticle, a lipid or a cell penetrating peptide (CPP).

[00672] In an aspect the invention provides a pair of CRISPR complexes comprising Cas13 effector protein, each comprising a guide RNA (sgRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein at least one loop of each sgRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains, wherein each sgRNA of each Cas13 effector protein complex comprises a functional domain having a DNA cleavage activity.

[00673] In an aspect the invention provides a method for cutting a target sequence in a locus of interest comprising delivery to a cell of the Cas13 effector protein complexes or

component(s) thereof or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed in vivo. In an aspect the invention provides a method as herein-discussed wherein the delivery is via a lentivirus, an adenovirus, or an AAV.

[00674] In an aspect the invention provides a library, method or complex as herein-discussed wherein the sgRNA is modified to have at least one non-coding functional loop, e.g., wherein the at least one non-coding functional loop is repressive; for instance, wherein the at least one non-coding functional loop comprises Alu.

[00675] In one aspect, the invention provides a method for altering or modifying expression of a gene product. The said method may comprise introducing into a cell containing and expressing a DNA molecule encoding the gene product an engineered, non-naturally occurring CRISPR system comprising a Cas13 effector protein and guide RNA that targets the RNA molecule, whereby the guide RNA targets the RNA target molecule encoding the gene product and the Cas13 effector protein cleaves the RNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas13 effector protein and the guide RNA do not naturally occur together. The invention comprehends the guide RNA comprising a guide sequence linked to a direct repeat sequence. The invention further comprehends the Cas13 effector protein being codon optimized for expression in a Eukaryotic cell. In a preferred embodiment the Eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[00676] In some embodiments, one or more functional domains are associated with the Cas13 effector protein. In some embodiments, one or more functional domains are associated with an adaptor protein, for example as used with the modified guides of Konnerman et al. (Nature 517, 583-588, 29 January 2015). In some embodiments, one or more functional domains are associated with an dead sgRNA (dRNA). In some embodiments, a dRNA complex with active Cas13 effector protein directs gene regulation by a functional domain at on gene locus while an sgRNA directs DNA cleavage by the active Cas13 effector protein at another locus, for example as described analogously in CRISPR-Cas9 systems by Dahlman et al., 'Orthogonal gene control with a catalytically active Cas9 nuclease,' Nature Biotechnology 33, p. 1159-61 (November, 2015). In some embodiments, dRNAs are selected to maximize selectivity of regulation for a gene locus of interest compared to off-target regulation. In some embodiments, dRNAs are selected to maximize target gene regulation and minimize target cleavage

[00677] For the purposes of the following discussion, reference to a functional domain could be a functional domain associated with the Cas13 effector protein or a functional domain associated with the adaptor protein.

[00678] In some embodiments, the one or more functional domains is an NLS (Nuclear Localization Sequence) or an NES (Nuclear Export Signal). In some embodiments, the one or more functional domains is a transcriptional activation domain comprises VP64, p65, MyoD1, HSF1, RTA, SET7/9 and a histone acetyltransferase. Other references herein to activation (or activator) domains in respect of those associated with the CRISPR enzyme include any known transcriptional activation domain and specifically VP64, p65, MyoD1, HSF1, RTA, SET7/9 or a histone acetyltransferase.

[00679] In some embodiments, the one or more functional domains is a transcriptional repressor domain. In some embodiments, the transcriptional repressor domain is a KRAB domain. In some embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[00680] In some embodiments, the one or more functional domains have one or more activities comprising translation activation activity, translation repression activity, methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, DNA integration activity or nucleic acid binding activity.

[00681] Histone modifying domains are also preferred in some embodiments. Exemplary histone modifying domains are discussed below. Transposase domains, HR (Homologous Recombination) machinery domains, recombinase domains, and/or integrase domains are also preferred as the present functional domains. In some embodiments, DNA integration activity includes HR machinery domains, integrase domains, recombinase domains and/or transposase domains. Histone acetyltransferases are preferred in some embodiments.

[00682] In some embodiments, the DNA cleavage activity is due to a nuclease. In some embodiments, the nuclease comprises a FokI nuclease. See, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung *Nature Biotechnology* 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[00683] In some embodiments, the one or more functional domains is attached to the Cas13 effector protein so that upon binding to the sgRNA and target the functional domain is

in a spatial orientation allowing for the functional domain to function in its attributed function.

[00684] In some embodiments, the one or more functional domains is attached to the adaptor protein so that upon binding of the Cas13 effector protein to the sgRNA and target, the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[00685] In an aspect the invention provides a composition as herein discussed wherein the one or more functional domains is attached to the Cas13 effector protein or adaptor protein via a linker, optionally a GlySer linker, as discussed herein.

[00686] It is also preferred to target endogenous (regulatory) control elements, such as involved in translation, stability, etc. Targeting of known control elements can be used to activate or repress the gene of interest. Targeting of putative control elements on the other hand can be used as a means to verify such elements (by measuring the translation of the gene of interest) or to detect novel control elements (. In addition, targeting of putative control elements can be useful in the context of understanding genetic causes of disease. Many mutations and common SNP variants associated with disease phenotypes are located outside coding regions. Targeting of such regions with either the activation or repression systems described herein can be followed by readout of transcription of either a) a set of putative targets (e.g. a set of genes located in closest proximity to the control element) or b) whole-transcriptome readout by e.g. RNAseq or microarray. This would allow for the identification of likely candidate genes involved in the disease phenotype. Such candidate genes could be useful as novel drug targets.

[00687] Histone acetyltransferase (HAT) inhibitors are mentioned herein. However, an alternative in some embodiments is for the one or more functional domains to comprise an acetyltransferase, preferably a histone acetyltransferase. These are useful in the field of epigenomics, for example in methods of interrogating the epigenome. Methods of interrogating the epigenome may include, for example, targeting epigenomic sequences. Targeting epigenomic sequences may include the guide being directed to an epigenomic target sequence. Epigenomic target sequence may include, in some embodiments, include a promoter, silencer or an enhancer sequence.

[00688] Use of a functional domain linked to a Cas13 effector protein as described herein, preferably a dead- Cas13 effector protein, more preferably a dead-FnCas13 effector protein, to target epigenomic sequences can be used to activate or repress promoters, silencer or enhancers.

[00689] Examples of acetyltransferases are known but may include, in some embodiments, histone acetyltransferases. In some embodiments, the histone acetyltransferase may comprise the catalytic core of the human acetyltransferase p300 (Gerbasch & Reddy, Nature Biotech 6th April 2015).

[00690] In some preferred embodiments, the functional domain is linked to a dead- Cas13 effector protein to target and activate epigenomic sequences such as promoters or enhancers. One or more guides directed to such promoters or enhancers may also be provided to direct the binding of the CRISPR enzyme to such promoters or enhancers.

[00691] In certain embodiments, the RNA targeting effector protein of the invention can be used to interfere with co-transcriptional modifications of DNA/chromatin structure, RNA-directed DNA methylation, or RNA-directed silencing/activation of DNA/chromatin. RNA-directed DNA methylation (RdDM) is an epigenetic process first discovered in plants. During RdDM, double-stranded RNAs (dsRNAs) are processed to 21-24 nucleotide small interfering RNAs (siRNAs) and guide methylation of homologous DNA loci. Besides RNA molecules, a plethora of proteins are involved in the establishment of RdDM, like Argonautes, DNA methyltransferases, chromatin remodelling complexes and the plant-specific PolIV and PolV. All these act in concert to add a methyl-group at the 5' position of cytosines. Small RNAs can modify the chromatin structure and silence transcription by guiding Argonaute-containing complexes to complementary nascent (non-coding) RNA transcripts. Subsequently the recruitment of chromatin-modifying complexes, including histone and DNA methyltransferases, is mediated. The RNA targeting effector protein of the invention may be used to target such small RNAs and interfere in interactions between these small RNAs and the nascent non-coding transcripts.

[00692] The term "associated with" is used here in relation to the association of the functional domain to the Cas13 effector protein or the adaptor protein. It is used in respect of how one molecule 'associates' with respect to another, for example between an adaptor protein and a functional domain, or between the Cas13 effector protein and a functional domain. In the case of such protein-protein interactions, this association may be viewed in terms of recognition in the way an antibody recognizes an epitope. Alternatively, one protein may be associated with another protein via a fusion of the two, for instance one subunit being fused to another subunit. Fusion typically occurs by addition of the amino acid sequence of one to that of the other, for instance via splicing together of the nucleotide sequences that encode each protein or subunit. Alternatively, this may essentially be viewed as binding between two molecules or direct linkage, such as a fusion protein. In any event, the fusion

protein may include a linker between the two subunits of interest (i.e. between the enzyme and the functional domain or between the adaptor protein and the functional domain). Thus, in some embodiments, the Cas13 effector protein or adaptor protein is associated with a functional domain by binding thereto. In other embodiments, the Cas13 effector protein or adaptor protein is associated with a functional domain because the two are fused together, optionally via an intermediate linker.

Saturating Mutagenesis

[00693] The Cas13 effector protein system(s) described herein can be used to perform saturating or deep scanning mutagenesis of genomic loci in conjunction with a cellular phenotype—for instance, for determining critical minimal features and discrete vulnerabilities of functional elements required for gene expression, drug resistance, and reversal of disease. By saturating or deep scanning mutagenesis is meant that every or essentially every RNA base is cut within the genomic loci. A library of Cas13 effector protein guide RNAs may be introduced into a population of cells. The library may be introduced, such that each cell receives a single guide RNA (sgRNA). In the case where the library is introduced by transduction of a viral vector, as described herein, a low multiplicity of infection (MOI) is used. The library may include sgRNAs targeting every sequence upstream of a (protospacer adjacent motif) (PAM) sequence in a genomic locus. The library may include at least 100 non-overlapping genomic sequences upstream of a PAM sequence for every 1000 base pairs within the genomic locus. The library may include sgRNAs targeting sequences upstream of at least one different PAM sequence. The Cas13 effector protein systems may include more than one Cas13 protein. Any Cas13 effector protein as described herein, including orthologues or engineered Cas13 effector proteins that recognize different PAM sequences may be used. The frequency of off target sites for a sgRNA may be less than 500. Off target scores may be generated to select sgRNAs with the lowest off target sites. Any phenotype determined to be associated with cutting at a sgRNA target site may be confirmed by using sgRNAs targeting the same site in a single experiment. Validation of a target site may also be performed by using a modified Cas13 effector protein, as described herein, and two sgRNAs targeting the genomic site of interest. Not being bound by a theory, a target site is a true hit if the change in phenotype is observed in validation experiments.

[00694] The Cas13 effector protein system(s) for saturating or deep scanning mutagenesis can be used in a population of cells. The Cas13 effector protein system(s) can be used in eukaryotic cells, including but not limited to mammalian and plant cells. The population of cells may be prokaryotic cells. The population of eukaryotic cells may be a population of

embryonic stem (ES) cells, neuronal cells, epithelial cells, immune cells, endocrine cells, muscle cells, erythrocytes, lymphocytes, plant cells, or yeast cells.

[00695] In one aspect, the present invention provides for a method of screening for functional elements associated with a change in a phenotype. The library may be introduced into a population of cells that are adapted to contain a Cas13 effector protein. The cells may be sorted into at least two groups based on the phenotype. The phenotype may be expression of a gene, cell growth, or cell viability. The relative representation of the guide RNAs present in each group are determined, whereby genomic sites associated with the change in phenotype are determined by the representation of guide RNAs present in each group. The change in phenotype may be a change in expression of a gene of interest. The gene of interest may be upregulated, downregulated, or knocked out. The cells may be sorted into a high expression group and a low expression group. The population of cells may include a reporter construct that is used to determine the phenotype. The reporter construct may include a detectable marker. Cells may be sorted by use of the detectable marker.

[00696] In another aspect, the present invention provides for a method of screening for loci associated with resistance to a chemical compound. The chemical compound may be a drug or pesticide. The library may be introduced into a population of cells that are adapted to contain a Cas13 effector protein, wherein each cell of the population contains no more than one guide RNA; the population of cells are treated with the chemical compound; and the representation of guide RNAs are determined after treatment with the chemical compound at a later time point as compared to an early time point, whereby genomic sites associated with resistance to the chemical compound are determined by enrichment of guide RNAs. Representation of sgRNAs may be determined by deep sequencing methods.

[00697] Useful in the practice of the instant invention utilizing Cas13 effector protein complexes are methods used in CRISPR-Cas9 systems and reference is made to the article entitled BCL1 1A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., Yuan, G., Zhang, F., Orkin, S.H., & Bauer, D.E. DOI:10.1038/nature15521, published online September 16, 2015, the article is herein incorporated by reference and discussed briefly below:

[00698] Canver et al. involves novel pooled CRISPR-Cas9 guide RNA libraries to perform in situ saturating mutagenesis of the human and mouse BCL1 1A erythroid enhancers previously identified as an enhancer associated with fetal hemoglobin (HbF) level and whose

mouse ortholog is necessary for erythroid BCL1 1A expression. This approach revealed critical minimal features and discrete vulnerabilities of these enhancers. Through editing of primary human progenitors and mouse transgenesis, the authors validated the BCL1 1A erythroid enhancer as a target for HbF reinduction. The authors generated a detailed enhancer map that informs therapeutic genome editing.

Method of Using Cas13 Systems to Modify a Cell or Organism

[00699] The invention in some embodiments comprehends a method of modifying a cell or organism. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The mammalian cell may be a non-human primate, bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish or shrimp. The cell may also be a plant cell. The plant cell may be of a crop plant such as cassava, corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree or vegetable. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

[00700] The system may comprise one or more different vectors. In an aspect of the invention, the effector protein is codon optimized for expression the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell.

[00701] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ 2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producing a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also be infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper

plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[00702] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huhl, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr ^{-/-}, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-IOA, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1 A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a

nucleic acid-targeting system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a nucleic acid-targeting complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[00703] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. In certain embodiments, the organism or subject is a plant. In certain embodiments, the organism or subject or plant is algae. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

[00704] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a nucleic acid-targeting complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said target polynucleotide.

[00705] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a nucleic acid-targeting complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said polynucleotide.

Cas13 Effector Protein Complexes Can Be Used In Plants

[00706] The Cas13 effector protein system(s) (e.g., single or multiplexed) can be used in conjunction with recent advances in crop genomics. The systems described herein can be used to perform efficient and cost effective plant gene or genome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The Cas13 effector protein system(s) can be used with regard to plants in Site-Directed

Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques. Aspects of utilizing the herein described Cas13 effector protein systems may be analogous to the use of the CRISPR-Cas (e.g. CRISPR-Cas9) system in plants, and mention is made of the University of Arizona website "CRISPR-PLANT" (<http://www.genome.arizona.edu/crispr/>) (supported by Penn State and AGI). Embodiments of the invention can be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; see, e.g., Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system," *Plant Methods* 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system," *Plant Physiology* September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system," *Nature Biotechnology* 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR/Cas system," *Cell Research* (2013) 23:1229-1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system," *Mol Plant*. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the *Agrobacterium tumefaciens*-mediated CRISPR-Cas system in rice," *Rice* 2014, 7:5 (2014), Zhou et al., "Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate: CoA ligase specificity and Redundancy," *New Phytologist* (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliendo et al, "Targeted DNA degradation using a CRISPR device stably carried in the host genome," *NATURE COMMUNICATIONS* 6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - *Agrobacterium*-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications," *Nat Rev Genet*. 2011 Dec 29;13(2):85-96; each of which is incorporated by reference herein including as to how herein embodiments may be used as to plants. Accordingly, reference herein to animal cells may also apply, *mutatis mutandis*, to plant cells unless otherwise apparent; and, the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein.

[00707] Sugano et al. (Plant Cell Physiol. 2014 Mar;55(3):475-81. doi: 10.1093/pcp/pcu014. Epub 2014 Jan 18) reports the application of CRISPR-Cas9 to targeted mutagenesis in the liverwort *Marchantia polymorpha* L., which has emerged as a model species for studying land plant evolution. The U6 promoter of *M. polymorpha* was identified and cloned to express the gRNA. The target sequence of the gRNA was designed to disrupt the gene encoding auxin response factor 1 (ARF1) in *M. polymorpha*. Using *Agrobacterium*-mediated transformation, Sugano et al. isolated stable mutants in the gametophyte generation of *M. polymorpha*. CRISPR-Cas9-based site-directed mutagenesis *in vivo* was achieved using either the Cauliflower mosaic virus 35S or *M. polymorpha* EFla promoter to express Cas9. Isolated mutant individuals showing an auxin-resistant phenotype were not chimeric. Moreover, stable mutants were produced by asexual reproduction of T1 plants. Multiple arf1 alleles were easily established using CRISPR/Cas9-based targeted mutagenesis. The Cas13 systems of the present invention can be used to regulate the same as well as other genes, and like expression control systems such as RNAi and siRNA, the method of the invention can be inducible and reversible.

[00708] Kabadi et al. (Nucleic Acids Res. 2014 Oct 29;42(19):e147. doi: 10.1093/nar/gku749. Epub 2014 Aug 13) developed a single lentiviral system to express a Cas9 variant, a reporter gene and up to four sgRNAs from independent RNA polymerase III promoters that are incorporated into the vector by a convenient Golden Gate cloning method. Each sgRNA was efficiently expressed and can mediate multiplex gene editing and sustained transcriptional activation in immortalized and primary human cells. The instant invention can be used to regulate the plant genes of Kabadi.

[00709] Xing et al. (BMC Plant Biology 2014, 14:327) developed a CRISPR-Cas9 binary vector set based on the pGreen or pCambia backbone, as well as a gRNA. This toolkit requires no restriction enzymes besides BsaI to generate final constructs harboring maize-codon optimized Cas9 and one or more gRNAs with high efficiency in as little as one cloning step. The toolkit was validated using maize protoplasts, transgenic maize lines, and transgenic *Arabidopsis* lines and was shown to exhibit high efficiency and specificity. More importantly, using this toolkit, targeted mutations of three *Arabidopsis* genes were detected in transgenic seedlings of the T1 generation. Moreover, the multiple-gene mutations could be inherited by the next generation. (guide RNA)module vector set, as a toolkit for multiplex genome editing in plants. The Cas13 systems and proteins of the instant invention may be used to target the genes targeted by Xing.

[00710] The Cas13 CRISPR systems of the invention may be used in the detection of plant viruses. Gambino et al. (Phytopathology. 2006 Nov;96(11):1223-9. doi: 10.1094/PHYTO-96-1223) relied on amplification and multiplex PCR for simultaneous detection of nine grapevine viruses. The Cas13 systems and proteins of the instant invention may similarly be used to detect multiple targets in a host. Moreover, the systems of the invention can be used to simultaneously knock down viral gene expression in valuable cultivars, and prevent activation or further infection by targeting expressed viral RNA.

[00711] Murray et al. (Proc Biol Sci. 2013 Jun 26;280(1765):20130965. doi: 10.1098/rspb.2013.0965; published 2013 Aug 22) analyzed 12 plant RNA viruses to investigate evolutionary rates and found evidence of episodic selection possibly due to shifts between different host genotypes or species. The Cas13 systems and proteins of the instant invention may be used to target or immunize against such viruses in a host. For example, the systems of the invention can be used to block viral RNA expression hence replication. Also, the invention can be used to target nucleic acids for cleavage as well as to target expression or activation. Moreover, the systems of the invention can be multiplexed so as to hit multiple targets or multiple isolates of the same virus.

[00712] Ma et al. (Mol Plant. 2015 Aug 3;8(8):1274-84. doi: 10.1016/j.molp.2015.04.007) reports a robust CRISPR-Cas9 vector system, utilizing a plant codon optimized Cas9 gene, for convenient and high-efficiency multiplex genome editing in monocot and dicot plants. Ma et al. designed PCR-based procedures to rapidly generate multiple sgRNA expression cassettes, which can be assembled into the binary CRISPR-Cas9 vectors in one round of cloning by Golden Gate ligation or Gibson Assembly. With this system, Ma et al. edited 46 target sites in rice with an average 85.4% rate of mutation, mostly in biallelic and homozygous status. Ma et al. provide examples of loss-of-function gene mutations in rice and Arabidopsis plants by simultaneous targeting of multiple (up to eight) members of a gene family, multiple genes in a biosynthetic pathway, or multiple sites in a single gene. Similarly, the Cas13 systems of the instant invention can efficiently target expression of multiple genes simultaneously.

[00713] Lowder et al. (Plant Physiol. 2015 Aug 21. pii: pp.00636.2015) also developed a CRISPR-Cas9 toolbox that enables multiplex genome editing and transcriptional regulation of expressed, silenced or non-coding genes in plants. This toolbox provides researchers with a protocol and reagents to quickly and efficiently assemble functional CRISPR-Cas9 T-DNA constructs for monocots and dicots using Golden Gate and Gateway cloning methods. It comes with a full suite of capabilities, including multiplexed gene editing and transcriptional

activation or repression of plant endogenous genes. T-DNA based transformation technology is fundamental to modern plant biotechnology, genetics, molecular biology and physiology. As such, we developed a method for the assembly of Cas9 (WT, nickase or dCas9) and gRNA(s) into a T-DNA destination-vector of interest. The assembly method is based on both Golden Gate assembly and MultiSite Gateway recombination. Three modules are required for assembly. The first module is a Cas9 entry vector, which contains promoterless Cas9 or its derivative genes flanked by attL1 and attR5 sites. The second module is a gRNA entry vector which contains entry gRNA expression cassettes flanked by attL5 and attL2 sites. The third module includes attR1-attR2-containing destination T-DNA vectors that provide promoters of choice for Cas9 expression. The toolbox of Lowder et al. may be applied to the Cas13 effector protein system of the present invention.

[00714] Organisms such as yeast and microalgae are widely used for synthetic biology. Stovicek et al. (Metab. Eng. Comm., 2015; 2:13 describes genome editing of industrial yeast, for example, *Saccharomyces cerevisiae*, to efficiently produce robust strains for industrial production. Stovicek used a CRISPR-Cas9 system codon-optimized for yeast to simultaneously disrupt both alleles of an endogenous gene and knock in a heterologous gene. Cas9 and gRNA were expressed from genomic or episomal 2 μ -based vector locations. The authors also showed that gene disruption efficiency could be improved by optimization of the levels of Cas9 and gRNA expression. Hlavova et al. (Biotechnol. Adv. 2015) discusses development of species or strains of microalgae using techniques such as CRISPR to target nuclear and chloroplast genes for insertional mutagenesis and screening. The same plasmids and vectors can be applied to the Cas13 systems of the instant invention.

[00715] Petersen ("Towards precisely glycol engineered plants," Plant Biotech Denmark Annual meeting 2015, Copenhagen, Denmark) developed a method of using CRISPR/Cas9 to engineer genome changes in Arabidopsis, for example to glyco engineer Arabidopsis for production of proteins and products having desired posttranslational modifications. Hebelstrup et al. (Front Plant Sci. 2015 Apr 23; 6:247) outlines *in planta* starch bioengineering, providing crops that express starch modifying enzymes and directly produce products that normally are made by industrial chemical and/or physical treatments of starches. The methods of Petersen and Hebelstrup may be applied to the Cas13 effector protein system of the present invention.

[00716] Kurthe t al, J Virol. 2012 Jun;86(1 1):6002-9. doi: 10.1128/JVI.00436-12. Epub 2012 Mar 21) developed an RNA virus-based vector for the introduction of desired traits into grapevine without heritable modifications to the genome. The vector provided the ability to

regulate expression of of endogenous genes by virus-induced gene silencing. The Cas13 systems and proteins of the instant invention can be used to silence genes and proteins without heritable modification to the genome.

[00717] In an embodiment, the plant may be a legume. The present invention may utilize the herein disclosed CRISPR-Cas system for exploring and modifying, for example, without limitation, soybeans, peas, and peanuts. Curtin et al. provides a toolbox for legume function genomics. (See Curtin et al., "A genome engineering toolbox for legume Functional genomics," International Plant and Animal Genome Conference XXII 2014). Curtin used the genetic transformation of CRISPR to knock-out/down single copy and duplicated legume genes both in hairy root and whole plant systems. Some of the target genes were chosen in order to explore and optimize the features of knock-out/down systems (e.g., phytoene desaturase), while others were identified by soybean homology to Arabidopsis Dicer-like genes or by genome-wide association studies of nodulation in Medicago. The Cas13 systems and proteins of the instant invention can be used to knockout/knockdown systems.

[00718] Peanut allergies and allergies to legumes generally are a real and serious health concern. The Cas13 effector protein system of the present invention can be used to identify and then edit or silence genes encoding allergenic proteins of such legumes. Without limitation as to such genes and proteins, Nicolaou et al. identifies allergenic proteins in peanuts, soybeans, lentils, peas, lupin, green beans, and mung beans. See, Nicolaou et al., *Current Opinion in Allergy and Clinical Immunology* 201 1;1 1(3):222).

[00719] In an advantageous embodiment, the plant may be a tree. The present invention may also utilize the herein disclosed CRISPR Cas system for herbaceous systems (see, e.g., Belhaj et al., *Plant Methods* 9: 39 and Harrison et al., *Genes & Development* 28: 1859-1872). In a particularly advantageous embodiment, the CRISPR Cas system of the present invention may target single nucleotide polymorphisms (SNPs) in trees (see, e.g., Zhou et al., *New Phytologist*, Volume 208, Issue 2, pages 298-301, October 2015). In the Zhou et al. study, the authors applied a CRISPR Cas system in the woody perennial *Populus* using the 4-coumarate:CoA ligase (4CL) gene family as a case study and achieved 100% mutational efficiency for two 4CL genes targeted, with every transformant examined carrying biallelic modifications. In the Zhou et al., study, the CRISPR-Cas9 system was highly sensitive to single nucleotide polymorphisms (SNPs), as cleavage for a third 4CL gene was abolished due to SNPs in the target sequence. These methods may be applied to the Cas13 effector protein system of the present invention.

[00720] The methods of Zhou et al. (New Phytologist, Volume 208, Issue 2, pages 298-301, October 2015) may be applied to the present invention as follows. Two 4CL genes, 4CL1 and 4CL2, associated with lignin and flavonoid biosynthesis, respectively are targeted for CRISPR-Cas9 editing. The *Populus tremula* x *alba* clone 717-1B4 routinely used for transformation is divergent from the genome-sequenced *Populus trichocarpa*. Therefore, the 4CL1 and 4CL2 gRNAs designed from the reference genome are interrogated with in-house 717 RNA-Seq data to ensure the absence of SNPs which could limit Cas efficiency. A third gRNA designed for 4CL5, a genome duplicate of 4CL1, is also included. The corresponding 717 sequence harbors one SNP in each allele near/within the PAM, both of which are expected to abolish targeting by the 4CL5-gRNA. All three gRNA target sites are located within the first exon. For 717 transformation, the gRNA is expressed from the *Medicago* U6.6 promoter, along with a human codon-optimized Cas under control of the CaMV 35S promoter in a binary vector. Transformation with the Cas-only vector can serve as a control. Randomly selected 4CL1 and 4CL2 lines are subjected to amplicon-sequencing. The data is then processed and biallelic mutations are confirmed in all cases. These methods may be applied to the Cas13 effector protein system of the present invention.

[00721] In plants, pathogens are often host-specific. For example, *Fusarium oxysporum* f. sp. *lycopersici* causes tomato wilt but attacks only tomato, and *F. oxysporum* f. *dianthii* *Puccinia graminis* f. sp. *tritici* attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present

invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[00722] Aside from the plants otherwise discussed herein and above, engineered plants modified by the effector protein and suitable guide, and progeny thereof, as provided. These may include disease or drought resistant crops, such as wheat, barley, rice, soybean or corn; plants modified to remove or reduce the ability to self-pollinate (but which can instead, optionally, hybridise instead); and allergenic foods such as peanuts and nuts where the immunogenic proteins have been disabled, destroyed or disrupted by targeting via a effector protein and suitable guide.

Therapeutic Treatment

[00723] The system of the invention can be applied in areas of former RNA cutting technologies, without undue experimentation, from this disclosure, including therapeutic, assay and other applications, because the present application provides the foundation for informed engineering of the system. The present invention provides for therapeutic treatment of a disease caused by overexpression of RNA, toxic RNA and/or mutated RNA (such as, for example, splicing defects or truncations). Expression of the toxic RNA may be associated with formation of nuclear inclusions and late-onset degenerative changes in brain, heart or skeletal muscle. In the best studied example, myotonic dystrophy, it appears that the main pathogenic effect of the toxic RNA is to sequester binding proteins and compromise the regulation of alternative splicing (Hum. Mol. Genet. (2006) 15 (suppl 2): R162-R169). Myotonic dystrophy [dystrophia myotonica (DM)] is of particular interest to geneticists because it produces an extremely wide range of clinical features. A partial listing would include muscle wasting, cataracts, insulin resistance, testicular atrophy, slowing of cardiac conduction, cutaneous tumors and effects on cognition. The classical form of DM, which is now called DM type 1 (DM1), is caused by an expansion of CTG repeats in the 3'-untranslated region (UTR) of DMPK, a gene encoding a cytosolic protein kinase.

[00724] The below table presents a list of exons shown to have misregulated alternative splicing in DM1 skeletal muscle, heart or brain.

Tissue/gene Target	Reference
<i>Skeletal muscle</i>	
ALP ex 5a, 5b	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. Hum. Mol. Genet 2006; 15:2087-2097
CAPN3 ex 16	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. Hum. Mol. Genet 2006;15:2087-2097

Tissue/gene Target	Reference
CLCN1 int 2, ex 7a, 8a	Mankodi A., et al. Expanded CUG repeats trigger aberrant splicing of CIC-1 chloride channel pre-mRNA and hyperexcitability of skeletal muscle in myotonic dystrophy. <i>Mol. Cell</i> 2002;10:35-44 Charlet-B N., et al. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. <i>Mol. Cell</i> 2002;10:45-53
FHOS ex 11a	Lin X., et al. Failure of MBNL1-dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006;15:2087-2097
GFAT1 ex 10	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006;15:2087-2097
IR ex 11	Savkur R.S., et al. Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. <i>Nat. Genet.</i> 2001;29:40-47
MBNL1 ex 7	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
MBNL2 ex 7	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
MTMR1 ex 2.1, 2.2	Buj-Bello A., et al. Muscle-specific alternative splicing of myotubularin-related 1 gene is impaired in DM1 muscle cells. <i>Hum. Mol. Genet.</i> 2002;11:2297-2307
NRAP ex 12	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
RYR1 ex 70	Kimura T., et al. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca ²⁺ -ATPase in myotonic dystrophy type 1. <i>Hum. Mol. Genet.</i> 2005;14:2189-2200
SERCA1 ex 22	Kimura T., et al. Altered mRNA splicing of the skeletal muscle ryanodine receptor and sarcoplasmic/endoplasmic reticulum Ca ²⁺ -ATPase in myotonic dystrophy type 1. <i>Hum. Mol. Genet.</i> 2005;14:2189-2200 Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
z-Titin ex Zr4, Zr5	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
m-Titin M-line ex5	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing transitions in myotonic dystrophy. <i>Hum. Mol. Genet</i> 2006; 15:2087-2097
TNNT3 tetai ex	Kanadia R.N., et al. A muscleblind knockout model for myotonic dystrophy. <i>Science</i> 2003;302:1978-1980
ZASP ex 11	Lin X., et al. Failure of MBNL1 -dependent postnatal splicing

Tissue/gene Target	Reference
	transitions in myotonic dystrophy. Hum. Mol. Genet 2006; 15:2087-2097
<i>Heart</i>	
TNNT2 ex 5	Philips A.V., et al. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. Science 1998;280:737-741
ZASP ex 11	Mankodi A., et al. Nuclear RNA foci in the heart in myotonic dystrophy. Circ. Res. 2005;97:1 152-1 155
m-Titin M-line ex 5	Mankodi A., et al. Nuclear RNA foci in the heart in myotonic dystrophy. Circ. Res. 2005;97:1 152-1 155
KCNAB1 ex 2	Mankodi A., et al. Nuclear RNA foci in the heart in myotonic dystrophy. Circ. Res. 2005;97:1 152-1 155
ALP ex 5	(Mankodi A., et al. Nuclear RNA foci in the heart in myotonic dystrophy. Circ. Res. 2005;97:1 152-1 155

Brain

	Sergeant N., et al. Dysregulation of human brain microtubule-associated tau mRNA maturation in myotonic dystrophy type 1. Hum. Mol. Genet. 2001;10:2143-2155
TAU ex 2, ex 10	Jiang H., et al. Myotonic dystrophy type 1 associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins, and deregulated alternative splicing in neurons. Hum. Mol. Genet. 2004;13:3079-3088
APP ex 7	Jiang H., et al. Myotonic dystrophy type 1 associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins, and deregulated alternative splicing in neurons. Hum. Mol. Genet. 2004;13:3079-3088
NMDAR1 ex 5	Jiang H., et al. Myotonic dystrophy type 1 associated with nuclear foci of mutant RNA, sequestration of muscleblind proteins, and deregulated alternative splicing in neurons. Hum. Mol. Genet. 2004;13:3079-3088

[00725] The enzymes of the present invention may target overexpressed RNA or toxic RNA, such as for example, the DMPK gene or any of the misregulated alternative splicing in DM1 skeletal muscle, heart or brain in, for example, the above table.

[00726] The enzymes of the present invention may also target trans-acting mutations affecting RNA-dependent functions that cause disease (summarized in Cell. 2009 Feb 20; 136(4): 777-793) as indicated in the below table.

DISEASE	GENE / MUTATION	FUNCTION
Prader Willi syndrome	SNORD 116	ribosome biogenesis
Spinal muscular atrophy (SMA)	SMN2	splicing
Dyskeratosis congenita (X-linked)	DKC1	telomerase / translation
Dyskeratosis congenita (autosomal dominant)	H:RC	telomerase

DISEASE	GENE / MUTATION	FUNCTION
Dyskeratosis congenita (autosomal dominant)	TERT	telomerase
Diamond-Blackfan anemia	RPS19, RPS24	ribosome biogenesis
Shwachman-Diamond syndrome	SBDS	ribosome biogenesis
Treacher-Collins syndrome	TCOF1	ribosome biogenesis
Prostate cancer	SNHG5	ribosome biogenesis
Myotonic dystrophy, type 1 (DM1)	DMPK (RNA gain-of-function)	protein kinase
Myotonic dystrophy type 2 (DM2)	ZNF9 (RNA gain-of-function)	RNA binding
Spinocerebellar ataxia 8 (SCA8)	ATXN8 / ATXN8OS (RNA gain-of-function)	unknown / noncoding RNA
Huntington's disease-like 2 (HDL2)	JPH3 (RNA gain-of-function)	ion channel function
Fragile X-associated tremor ataxia syndrome (FXTAS)	FMR1 (RNA gain-of-function)	translation / mRNA localization
Fragile X syndrome	FMR1	translation / mRNA localization
X-linked mental retardation	UPF3B	translation/ nonsense mediated decay
Oculopharyngeal muscular dystrophy (OPMD)	PABPN1	3' end formation
Human pigmentary genodermatosis	DSRAD	editing
Retinitis pigmentosa	PRPF31	splicing
Retinitis pigmentosa	PRPF8	splicing
Retinitis pigmentosa	HPRP3	splicing
Retinitis pigmentosa	PAP1	splicing
Cartilage-hair hypoplasia (recessive)	RMRP	splicing
Autism	7q22-q33 locus breakpoint	noncoding RNA
Beckwith-Wiedemann syndrome (BWS)	H19	noncoding RNA
Charcot-Marie-Tooth (CMT) Disease	GRS	translation
Charcot-Marie-Tooth (CMT) Disease	YRS	translation
Amyotrophic lateral sclerosis (ALS)	TARDBP	splicing, transcription

DISEASE	GENE / MUTATION	FUNCTION
Leukoencephalopathy with vanishing white matter	EIF2B1	translation
Wolcott–Rallison syndrome	EIF2AK3	translation (protease)
Mitochondrial myopathy and sideroblastic anemia (MLASA)	PUS1	translation
Encephalomyopathy and hypertrophic cardiomyopathy	TSFM	translation (mitochondrial)
Hereditary spastic paraplegia	SPG7	ribosome biogenesis
Leukoencephalopathy	DARS2	translation (mitochondrial)
Susceptibility to diabetes mellitus	LARS2	translation (mitochondrial)
Deafness	MTRNR1	ribosome biogenesis (mitochondrial)
MELAS syndrome, deafness	MTRNR2	ribosome biogenesis (mitochondrial)
Cancer	SFRS1	splicing, translation, export
Cancer	RBM5	splicing
Multiple disorders	mitochondrial tRNA mutations	translation (mitochondrial)
Cancer	miR-17-92 cluster	RNA interference
Cancer	miR-372 / miR-373	RNA interference

[00727] The enzyme of the present invention may also be used in the treatment of various tauopathies, including primary and secondary tauopathies, such as primary age-related tauopathy (PART)/Neurofibrillary tangle-predominant senile dementia, with NFTs similar to AD, but without plaques, dementia pugilistica (chronic traumatic encephalopathy), progressive supranuclear palsy, corticobasal degeneration, frontotemporal dementia and parkinsonism linked to chromosome 17, lytico-Bodig disease (Parkinson-dementia complex of Guam), ganglioglioma and gangliocytoma, meningioangiomas, postencephalitic parkinsonism, subacute sclerosing panencephalitis, as well as lead encephalopathy, tuberous sclerosis, Hallervorden-Spatz disease, and lipofuscinosis, alzheimers disease. The enzymes of the present invention may also target mutations disrupting the cis-acting splicing code cause splicing defects and disease (summarized in Cell. 2009 Feb 20; 136(4): 777-793). The motor neuron degenerative disease SMA results from deletion of the SMN1 gene. The remaining

SMN2 gene has a C->T substitution in exon 7 that inactivates an exonic splicing enhancer (ESE), and creates an exonic splicing silencer (ESS), leading to exon 7 skipping and a truncated protein (SMNA7). A T->A substitution in exon 31 of the dystrophin gene simultaneously creates a premature termination codon (STOP) and an ESS, leading to exon 31 skipping. This mutation causes a mild form of DMD because the mRNA lacking exon 31 produces a partially functional protein. Mutations within and downstream of exon 10 of the MAPT gene encoding the tau protein affect splicing regulatory elements and disrupt the normal 1:1 ratio of mRNAs including or excluding exon 10. This results in a perturbed balance between tau proteins containing either four or three microtubule-binding domains (4R-tau and 3R-tau, respectively), causing the neuropathological disorder FTDP-17. The example shown is the N279K mutation which enhances an ESE function promoting exon 10 inclusion and shifting the balance toward increased 4R-tau. Polymorphic (UG)m(U)n tracts within the 3' splice site of the CFTR gene exon 9 influence the extent of exon 9 inclusion and the level of full-length functional protein, modifying the severity of cystic fibrosis (CF) caused by a mutation elsewhere in the CFTR gene.

[00728] The innate immune system detects viral infection primarily by recognizing viral nucleic acids inside an infected cell, referred to as DNA or RNA sensing. In vitro RNA sensing assays can be used to detect specific RNA substrates. The RNA targeting effector protein can for instance be used for RNA-based sensing in living cells. Examples of applications are diagnostics by sensing of, for examples, disease-specific RNAs.

[00729] The RNA targeting effector protein of the invention can further be used for antiviral activity, in particular against RNA viruses. The effector protein can be targeted to the viral RNA using a suitable guide RNA selective for a selected viral RNA sequence. In particular, the effector protein may be an active nuclease that cleaves RNA, such as single stranded RNA. provided is therefore the use of an RNA targeting effector protein of the invention as an antiviral agent.

[00730] Therapeutic dosages of the enzyme system of the present invention to target RNA the above-referenced RNAs are contemplated to be about 0.1 to about 2 mg/kg the dosages may be administered sequentially with a monitored response, and repeated dosages if necessary, up to about 7 to 10 doses per patient. Advantageously, samples are collected from each patient during the treatment regimen to ascertain the effectiveness of treatment. For example, RNA samples may be isolated and quantified to determine if expression is reduced or ameliorated. Such a diagnostic is within the purview of one of skill in the art.

Transcript Detection Methods

[00731] The effector proteins and systems of the invention are useful for specific detection of RNAs in a cell or other sample. In the presence of an RNA target of interest, guide-dependent Cas13 nuclease activity may be accompanied by non-specific RNase activity against collateral targets. To take advantage of the RNase activity, all that is needed is a reporter substrate that can be detectably cleaved. For example, a reporter molecule can comprise RNA, tagged with a fluorescent reporter molecule (fluor) on one end and a quencher on the other. In the absence of Cas13 RNase activity, the physical proximity of the quencher dampens fluorescence from the fluor to low levels. When Cas13 target specific cleavage is activated by the presence of an RNA target-of-interest and suitable guide RNA, the RNA-containing reporter molecule is non-specifically cleaved and the fluor and quencher are spatially separated. This causes the fluor to emit a detectable signal when excited by light of the appropriate wavelength.

[00732] In an aspect, the invention relates to a (target) RNA detection system comprising an RNA targeting effector; one or more guide RNAs designed to bind to the corresponding RNA target; and an RNA-based cleavage inducible reporter construct. In another aspect, the invention relates to a method for (target) RNA detection in a sample, comprising adding an RNA targeting effector, one or more guide RNAs designed to bind to said (target) RNA, and an RNA-based cleavage inducible reporter construct to said sample. In a further aspect, the invention relates to a kit or device comprising the (target) RNA detection system as defined herein, or a kit or device comprising at least the RNA targeting effector and the RNA-based cleavage inducible reporter construct. In a further aspect, the invention relates to the use of the RNA targeting system or kit or device as defined herein for (target) RNA detection. The RNA targeting effector in certain embodiments is an RNA guided RNase. In certain embodiments, the RNA targeting effector is a CRISPR effector. In certain embodiments, the RNA targeting effector is a class 2 CRISPR effector. In certain embodiments, the RNA targeting effector is a class 2, type VI CRISPR effector. In a preferred embodiment, the RNA targeting effector is Cas13. In certain embodiments, the RNA targeting effector, preferably Cas13, is derived from a species as described herein elsewhere. It will be understood that the guide RNA designed to bind to said (target) RNA as described herein is capable of forming a complex with the RNA targeting effector and wherein the guide RNA in said complex is capable of binding to a target RNA molecule and whereby the target RNA is cleaved, as also described herein elsewhere. It will be understood that the guide RNA typically comprises a guide sequence and a direct repeat, as described herein elsewhere. In certain embodiments,

the one or more guide RNAs are designed to bind to one or more target molecules that are diagnostic for a disease state. In certain embodiments, the disease state is infection, such as viral, bacterial, fungal, or parasitic infection. In certain embodiments, the disease state is characterised by aberrant (target) RNA expression. In certain embodiments, the disease state is cancer. In certain embodiments, the disease state is autoimmune disease. The RNA-based cleavage inducible reporter construct comprises RNA and cleavage of the RNA results in a detectable readout, i.e. a detectable signal is generated upon cleavage of the RNA. In certain embodiments, the RNA-based cleavage inducible reporter construct comprises a fluorochrome and a quencher. The skilled person will understand that different types of fluorochromes and corresponding quenchers may be used. The skilled person will readily envisage other types of inducible reporter systems which may be adapted for use in the present RNA cleavage reporter constructs.

[00733] In one exemplary assay method, Cas13 effector, target-of-interest-specific guide RNA, and reporter molecule are added to a cellular sample. An increase in fluorescence indicates the presence of the RNA target-of-interest. In another exemplary method, a detection array is provided. Each location of the array is provided with Cas13 effector, reporter molecule, and a target-of-interest-specific guide RNA. Depending on the assay to be performed, the target-of-interest-specific guide RNAs at each location of the array can be the same, different, or a combination thereof. Different target-of-interest-specific guide RNAs might be provided, for example when it is desired to test for one or more targets in a single source sample. The same target-of-interest-specific guide RNA might be provided at each location, for example when it is desired to test multiple samples for the same target.

[00734] As used herein, a "masking construct" refers to a molecule that can be cleaved or otherwise deactivated by an activated CRISPR system effector protein described herein. In certain example embodiments, the masking construct is a RNA-based masking construct. The masking construct prevents the generation or detection of a positive detectable signal. A positive detectable signal may be any signal that can be detected using optical, fluorescent, chemiluminescent, electrochemical or other detection methods known in the art. The masking construct may prevent the generation of a detectable positive signal or mask the presence of a detectable positive signal until the masking construct is removed or otherwise silenced. The term "positive detectable signal" is used to differentiate from other detectable signals that may be detectable in the presence of the masking construct. For example, in certain embodiments a first signal may be detected when the masking agent is present (*i.e.* a negative detectable signal), which then converts to a second signal (*e.g.* the positive detectable signal)

upon detection of the target molecules and cleavage or deactivation of the masking agent by the activated CRISPR effector protein.

[00735] In certain example embodiments, the masking construct may suppress generation of a gene product. The gene product may be encoded by a reporter construct that is added to the sample. The masking construct may be an interfering RNA involved in a RNA interference pathway, such as a shRNA or siRNA. The masking construct may also comprise microRNA (miRNA). While present, the masking construct suppresses expression of the gene product. The gene product may be a fluorescent protein or other RNA transcript or proteins that would otherwise be detectable by a labeled probe or antibody but for the presence of the masking construct. Upon activation of the effector protein the masking construct is cleaved or otherwise silenced allowing for expression and detection of the gene product as the positive detectable signal.

[00736] In certain example embodiments, the masking construct may sequester one or more reagents needed to generate a detectable positive signal such that release of the one or more reagents from the masking construct results in generation of the detectable positive signal. The one or more reagents may combine to produce a colorimetric signal, a chemiluminescent signal, a fluorescent signal, or any other detectable signal and may comprise any reagents known to be suitable for such a purpose. In certain example embodiments, the one or more reagents are sequestered by RNA aptamers that bind the one or more reagents. The one or more reagents are released when the effector protein is activated upon detection of a target molecule. In certain example embodiments, the one or more reagents is a protein, such as an enzyme, capable of facilitating generation of a detectable signal, such as a colorimetric, chemiluminescent, or fluorescent signal, that is inhibited or sequestered such that the protein cannot generate the detectable signal by the binding of one or more RNA aptamers to the protein. Upon activation of the effector proteins disclosed herein, the RNA aptamers are cleaved or degraded to the extent they no longer inhibit the protein's ability to generate the detectable signal.

[00737] In one embodiment, thrombin is used as a signal amplification enzyme with an inhibitory aptamer, for example having the following sequence: GGAACAAAGCUGAAGUACUUACCC (SEQ ID No. 115). When this aptamer is cleaved, thrombin becomes active and will cleave a peptide colorimetric substrate (see, e.g., www.sigmaaldrich.com/catalog/product/sigma/t3068?lang=en®ion=US) or fluorescent substrate (see, e.g., www.sigmaaldrich.com/catalog/product/sigma/b9385?lang=en®ion=US). The

colorimetric substrate, para-nitroanilide (pNA), is covalently linked to the peptide substrate for thrombin. Upon cleavage by thrombin, pNA is released and becomes yellow in color and easily visible by eye. The fluorescent substrate operates by a similar principle and, upon cleavage by thrombin, releases 7-amino-4-methylcoumarin, a blue fluorophore that can be detected using a fluorescence detector. Alternatives to thrombin include horseradish peroxidase (HRP), β -galactosidase, and calf alkaline phosphatase (CAP) which can similarly be used to generate a colorimetric or fluorescent signal, and be inhibited by an inhibitory aptamer.

[00738] In certain example embodiments, the masking construct may be immobilized on a solid substrate in an individual discrete volume (defined further below) and sequesters a single reagent. For example, the reagent may be a bead comprising a dye. When sequestered by the immobilized reagent, the individual beads are too diffuse to generate a detectable signal, but upon release from the masking construct are able to generate a detectable signal, for example by aggregation or simple increase in solution concentration. In certain example embodiments, the immobilized masking agent is a RNA-based aptamer that can be cleaved by the activated effector protein upon detection of a target molecule.

[00739] In certain other example embodiments, the masking construct binds to an immobilized reagent in solution thereby blocking the ability of the reagent to bind to a separate labeled binding partner that is free in solution. Thus, upon application of a washing step to a sample, the labeled binding partner can be washed out of the sample in the absence of a target molecule. However, if the effector protein is activated, the masking construct is cleaved to a degree sufficient to interfere with the ability of the masking construct to bind the reagent thereby allowing the labeled binding partner to bind to the immobilized reagent. Thus, the labeled binding partner remains after the wash step indicating the presence of the target molecule in the sample. In certain aspects, the masking construct that binds the immobilized reagent is a RNA aptamer. The immobilized reagent may be a protein and the labeled binding partner may be a labeled antibody. Alternatively, the immobilized reagent may be a streptavidin and the labeled binding partner may be labeled biotin. The label on the binding partner used in the above embodiments may be any detectable label known in the art. In addition, other known binding partners may be used in accordance with the overall design described here.

[00740] In certain example embodiments, the masking construct may comprise a ribozyme. Ribozymes are RNA molecules having catalytic properties. As ribozymes, both naturally and engineered, comprise or consist of RNA, that may be targeted by the effector

proteins disclosed herein. The ribozyme may be selected or engineered to catalyze a reaction that either generates a negative detectable signal or prevents generation of a positive control signal. Upon deactivation of the ribozyme by the activated effector protein molecule the reaction generating a negative controls signal or preventing generation of a positive detectable signal is removed, thereby allowing a positive detectable signal to be detected. In one example embodiment, the ribozyme may catalyze a colorimetric reaction causing a solution to appear as a first color. When the ribozyme is deactivated the solution then turns to a second color, the second color being the detectable positive signal. An example of how ribozymes can be used to catalyze a colorimetric reaction are described in Zhao et al. "Signal amplification of glucosamine-6-phosphate based on ribozyme glmS," *Biosens Bioelectron.* 2014; 16:337-42, and provide an example of how such a system could be modified to work in the context of the embodiments disclosed herein. Alternatively, ribozymes, when present can generate cleavage products of, for example, RNA transcripts. Thus, detection of a positive detectable signal may comprise detection of non-cleaved RNA transcripts that are only generated in the absence of the ribozyme.

[00741] In one example embodiment, the masking construct comprises a detection agent that changes color depending on whether the detection agent is aggregated or dispersed in solution. For example, certain nanoparticles, such as colloidal gold, undergo a visible purple to red color shift as they move from aggregates to dispersed particles. Accordingly, in certain example embodiments, such detection agents may be held in aggregate by one or more bridge molecules. At least a portion of the bridge molecule comprises RNA. Upon activation of the effector proteins disclosed herein, the RNA portion of the bridge molecule is cleaved allowing the detection agent to disperse and resulting in the corresponding change in color. In certain example embodiments the, bridge molecule is a RNA molecule. In certain example embodiments, the detection agent is a colloidal metal. The colloidal metal material may include water-insoluble metal particles or metallic compounds dispersed in a liquid, a hydrosol, or a metal sol. The colloidal metal may be selected from the metals in groups IA, IB, IIB and IIIB of the periodic table, as well as the transition metals, especially those of group VIII. Preferred metals include gold, silver, aluminum, ruthenium, zinc, iron, nickel and calcium. Other suitable metals also include the following in all of their various oxidation states: lithium, sodium, magnesium, potassium, scandium, titanium, vanadium, chromium, manganese, cobalt, copper, gallium, strontium, niobium, molybdenum, palladium, indium, tin, tungsten, rhenium, platinum, and gadolinium. The metals are preferably provided in

ionic form, derived from an appropriate metal compound, for example the Al^{3+} , Ru^{3+} , Zn^{2+} , Fe^{3+} , Ni^{2+} and Ca^{2+} ions

[00742] In certain other example embodiments, the masking construct may comprise an RNA oligonucleotide to which are attached a detectable label and a masking agent of that detectable label. An example of such a detectable label/masking agent pair is a fluorophore and a quencher of the fluorophore. Quenching of the fluorophore can occur as a result of the formation of a non-fluorescent complex between the fluorophore and another fluorophore or non-fluorescent molecule. This mechanism is known as ground-state complex formation, static quenching, or contact quenching. Accordingly, the RNA oligonucleotide may be designed so that the fluorophore and quencher are in sufficient proximity for contact quenching to occur. Fluorophores and their cognate quenchers are known in the art and can be selected for this purpose by one having ordinary skill in the art. The particular fluorophore/quencher pair is not critical in the context of this invention, only that selection of the fluorophore/quencher pairs ensures masking of the fluorophore. Upon activation of the effector proteins disclosed herein, the RNA oligonucleotide is cleaved thereby severing the proximity between the fluorophore and quencher needed to maintain the contact quenching effect. Accordingly, detection of the fluorophore may be used to determine the presence of a target molecule in a sample.

[00743] In one example embodiment, the masking construct may comprise a quantum dot. The quantum dot may have multiple linker molecules attached to the surface. At least a portion of the linker molecule comprises RNA. The linker molecule is attached to the quantum dot at one end and to one or more quenchers along the length or at terminal ends of the linker such that the quenchers are maintained in sufficient proximity for quenching of the quantum dot to occur. The linker may be branched. As above, the quantum dot/quencher pair is not critical, only that selection of the quantum dot/quencher pair ensures masking of the fluorophore. Quantum dots and their cognate quenchers are known in the art and can be selected for this purpose by one having ordinary skill in the art. Upon activation of the effector proteins disclosed herein, the RNA portion of the linker molecule is cleaved thereby eliminating the proximity between the quantum dot and one or more quenchers needed to maintain the quenching effect. In one embodiment, the quantum dot is streptavidin conjugated, such as Qdot® 625 Streptavidin Conjugate (www.thermofisher.com/order/catalog/product/A10196). RNA are attached via biotin linkers and recruit quenching molecules, with the sequence /5Biosg/UCUCGUACGUUC/3IAbRQSp/ (SEQ ID No. 116) or

/5Biosg/UCUCGUACGUUCUCUCGUACGUUC/3IAbRQSp/ (SEQ ID No 117) where /5Biosg/ is a biotin tag and /3IAbRQSp/ is an Iowa black quencher. Upon cleavage, the quencher will be released and the quantum dot will fluoresce visibly.

[00744] In a similar fashion, fluorescence energy transfer (FRET) may be used to generate a detectable positive signal. FRET is a non-radiative process by which a photon from an energetically excited fluorophore (i.e. "donor fluorophore") raises the energy state of an electron in another molecule (i.e. "the acceptor") to higher vibrational levels of the excited singlet state. The donor fluorophore returns to the ground state without emitting a fluorescence characteristic of that fluorophore. The acceptor can be another fluorophore or non-fluorescent molecule. If the acceptor is a fluorophore, the transferred energy is emitted as fluorescence characteristic of that fluorophore. If the acceptor is a non-fluorescent molecule the absorbed energy is lost as heat. Thus, in the context of the embodiments disclosed herein, the fluorophore/quencher pair is replaced with a donor fluorophore/acceptor pair attached to the oligonucleotide molecule. When intact, the masking construct generates a first signal (negative detectable signal) as detected by the fluorescence or heat emitted from the acceptor. Upon activation of the effector proteins disclosed herein the RNA oligonucleotide is cleaved and FRET is disrupted such that fluorescence of the donor fluorophore is now detected (positive detectable signal).

[00745] One mode of colorimetric readout for the detection of RNAses is based upon intercalating dyes, which change their absorbance in response to cleavage of long RNAs to short nucleotides. Several existing dyes with these properties exist. From Wagner (1983), Pyronine-Y will complex with RNA and form a complex that has an absorbance at 572nm; cleavage of RNA results in loss of absorbance and a color change. Greiner-Stoeffele (1996) used methylene blue in a similar fashion, with changes in absorbance at 688nm upon RNase activity.

[00746] Another mode of colorimetric readout involves nucleic acid substrates that change color upon cleavage. Witmer (1991) utilized a synthetic ribonucleotide substrate, U-3'-BCIP, that releases a reporter group after cleavage, resulting in generation of absorbance at 650nm.

[00747] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, AAV, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,945,839, 8,932,814, 8,906,616, 8,895,308, 8,889,418,

8,889,356, 8,871,445, 8,865,406, 8,795,965, 8,771,945 and 8,697,359; US Patent Publications US 2014-0310830 (US APP. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); European Patents EP 2 784 162 B1 and EP 2 771 468 B1; European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP 13824232.6), and EP 2 784 162 (EP 14 1703 83.5); and PCT Patent Publications PCT Patent Publications WO 2014/093661 (PCT/US20 13/074743), WO 2014/093694 (PCT/US20 13/074790), WO 2014/093595 (PCT/US20 13/0746 11), WO 2014/093718 (PCT/US20 13/074825), WO 2014/093709 (PCT/US20 13/0748 12), WO 2014/093622 (PCT/US20 13/074667), WO 2014/093635 (PCT/US20 13/074691), WO 2014/093655 (PCT/US20 13/07473 6), WO 2014/093712 (PCT/US20 13/0748 19), WO 2014/093701 (PCT/US20 13/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US20 14/04 1790), WO 2014/204724 (PCT/US20 14/04 1800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US20 14/04 1804), WO 2014/204727 (PCT/US20 14/04 1806), WO 2014/204728 (PCT/US20 14/04 1808), WO 2014/204729 (PCT/US20 14/04 1809). Reference is also made to US provisional patent applications 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US provisional patent application 61/836,123, filed on June 17, 2013. Reference is additionally made to US provisional patent applications 61/835,931, 61/835,936, 61/836,127, 61/836, 101, 61/836,080 and 61/835,973, each filed June 17, 2013. Further reference is made to US provisional patent applications 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961,980 filed on October 28, 2013. Reference is yet further made to: PCT Patent applications Nos: PCT/US20 14/04 1803,

PCT/US20 14/04 1800, PCT/US20 14/04 1809, PCT/US20 14/04 1804 and PCT/US20 14/04 1806, each filed June 10, 2014 6/10/14; PCT/US20 14/04 1808 filed June 11, 2014; and PCT/US2014/62558 filed October 28, 2014, and US Provisional Patent Applications Serial Nos.: 61/915,150, 61/915,301, 61/915,267 and 61/915,260, each filed December 12, 2013; 61/757,972 and 61/768,959, filed on January 29, 2013 and February 25, 2013; 61/835,936, 61/836,127, 61/836,101, 61/836,080, 61/835,973, and 61/835,931, filed June 17, 2013; 62/010,888 and 62/010,879, both filed June 11, 2014; 62/010,329 and 62/010,441, each filed June 10, 2014; 61/939,228 and 61/939,242, each filed February 12, 2014; 61/980,012, filed April 15, 2014; 62/038,358, filed August 17, 2014; 62/054,490, 62/055,484, 62/055,460 and 62/055,487, each filed September 25, 2014; and 62/069,243, filed October 27, 2014. Reference is also made to US provisional patent applications Nos. 62/055,484, 62/055,460, and 62/055,487, filed September 25, 2014; US provisional patent application 61/980,012, filed April 15, 2014; and US provisional patent application 61/939,242 filed February 12, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to US provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on December 12, 2013. Reference is made to US provisional patent application USSN 61/980,012 filed April 15, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US 14/4 1806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to US provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on December 12, 2013.

[00748] Mention is also made of US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,462, 12-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/096,324, 23-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-

WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/054,675, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH

OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[00749] Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

[00750] Also with respect to general information on CRISPR-Cas Systems, mention is made of the following (also hereby incorporated herein by reference):

- Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N, Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. *Science* Feb 15;339(6121):819-23 (2013);
- > RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. *Nat Biotechnol* Mar;31(3):233-9 (2013);
- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. *Cell* May 9;153(4):910-8 (2013);
- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Piatt RJ, Scott DA, Church GM, Zhang F. *Nature*. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);
- > Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. *Cell* Aug 28. pii: S0092-8674(13)01015-5 (2013-A);
- > DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X.,

- Shalem, O., Cradick, T.J., Marraffini, L.A., Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013);
- Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. *Nature Protocols* Nov;8(11):2281-308 (2013-B);
 - Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelsen, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. *Science* Dec 12. (2013). [Epub ahead of print];
 - Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. *Cell* Feb 27, 156(5):935-49 (2014);
 - Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC, Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. *Nat Biotechnol.* Apr 20. doi: 10.1038/nbt.2889 (2014);
 - CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Piatt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. *Cell* 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);
 - Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., *Cell.* Jun 5;157(6):1262-78 (2014).
 - Genetic screens in human cells using the CRISPR/Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., *Science.* January 3; 343(6166): 80-84. doi: 10.1126/science.1246981 (2014);
 - Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) *Nat Biotechnol.* Dec;32(12): 1262-7 (2014);
 - *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) *Nat Biotechnol.* Jan;33(1): 102-6 (2015);
 - Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu

- PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).
- A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2): 139-42 (2015);
 - Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246-1260, March 12, 2015 (multiplex screen in mouse), and
 - *In vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546): 186-91 (2015).
 - Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015).
 - Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015).
 - Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).
 - Ramanan et al., CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srepl0833 (June 2, 2015)
 - > Nishimasu et al., "Crystal Structure of *Staphylococcus aureus* Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)
 - Zetsche *et al.* (2015), "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system," Cell 163, 759-771 (Oct. 22, 2015) doi: 10.1016/j.cell.2015. 09.038. Epub Sep. 25, 2015
 - Shmakov *et al.* (2015), "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems," Molecular Cell 60, 385-397 (Nov. 5, 2015) doi: 10.1016/j.molcel.2015. 10.008. Epub Oct 22, 2015
 - > Dahlman et al., "Orthogonal gene control with a catalytically active Cas9 nuclease," Nature Biotechnology 33, 1159-1161 (November, 2015)
 - Gao *et al.*, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: <http://dx.doi.org/10.1101/091611> Epub Dec. 4, 2016

- > Smargon et al. (2017), "Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28," *Molecular Cell* 65, 618-630 (Feb. 16, 2017) doi: 10.1016/j.molcel.2016.12.023. Epub Jan 5, 2017

each of which is incorporated herein by reference, may be considered in the practice of the instant invention, and discussed briefly below:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptococcus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.
- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli*, 65% that were recovered contained the mutation.

- Wang *et al.* (2013) used the CRISPR/Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR/Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.
- > Konermann *et al.* (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- > Ran *et al.* (2013-A) described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. This addresses the issue of the Cas9 nuclease from the microbial CRISPR-Cas system being targeted to specific genomic loci by a guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.
- Hsu *et al.* (2013) characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and sgRNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.
- > Ran *et al.* (2013-B) described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in

mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1-2 weeks, and modified clonal cell lines can be derived within 2-3 weeks.

- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED 12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.
- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 Å resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.

- > Wu *et al.* mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targeted dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.
- Piatt *et al.* established a Cre-dependent Cas9 knockin mouse. The authors demonstrated *in vivo* as well as *ex vivo* genome editing using adeno-associated virus (AAV)-, lentivirus-, or particle-mediated delivery of guide RNA in neurons, immune cells, and endothelial cells.
- Hsu *et al.* (2014) is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells.
- Wang *et al.* (2014) relates to a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library.
- > Doench *et al.* created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.
- Swiech *et al.* demonstrate that AAV-mediated SpCas9 genome editing can enable reverse genetic studies of gene function in the brain.
- > Konermann *et al.* (2015) discusses the ability to attach multiple effector domains, e.g., transcriptional activator, functional and epigenomic regulators at appropriate positions on the guide such as stem or tetraloop with and without linkers.
- Zetsche *et al.* demonstrates that the Cas9 enzyme can be split into two and hence the assembly of Cas9 for activation can be controlled.

- > Chen *et al.* relates to multiplex screening by demonstrating that a genome-wide *in vivo* CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis.
- > Ran *et al.* (2015) relates to SaCas9 and its ability to edit genomes and demonstrates that one cannot extrapolate from biochemical assays. Shalem *et al.* (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.
- Shalem *et al.* (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.
- > Xu *et al.* (2015) assessed the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens. The authors explored efficiency of CRISPR/Cas9 knockout and nucleotide preference at the cleavage site. The authors also found that the sequence preference for CRISPRi/a is substantially different from that for CRISPR/Cas9 knockout.
- Parnas *et al.* (2015) introduced genome-wide pooled CRISPR-Cas9 libraries into dendritic cells (DCs) to identify genes that control the induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS). Known regulators of Tlr4 signaling and previously unknown candidates were identified and classified into three functional modules with distinct effects on the canonical responses to LPS.
- Ramanan *et al.* (2015) demonstrated cleavage of viral episomal DNA (cccDNA) in infected cells. The HBV genome exists in the nuclei of infected hepatocytes as a 3.2kb double-stranded episomal DNA species called covalently closed circular DNA (cccDNA), which is a key component in the HBV life cycle whose replication is not inhibited by current therapies. The authors showed that sgRNAs specifically targeting highly conserved regions of HBV robustly suppresses viral replication and depleted cccDNA.
- > Nishimasu *et al.* (2015) reported the crystal structures of SaCas9 in complex with a single guide RNA (sgRNA) and its double-stranded DNA targets, containing the 5'-TTGAAT-3' PAM and the 5'-TTGGGT-3' PAM. A structural comparison of SaCas9

with SpCas9 highlighted both structural conservation and divergence, explaining their distinct PAM specificities and orthologous sgRNA recognition.

- Zetsche *et al.* (2015) reported the characterization of Cpf1, a putative class 2 CRISPR effector. It was demonstrated that Cpf1 mediates robust DNA interference with features distinct from Cas9. Identifying this mechanism of interference broadens our understanding of CRISPR-Cas systems and advances their genome editing applications.
- Shmakov *et al.* (2015) reported the characterization of three distinct Class 2 CRISPR-Cas systems. The effectors of two of the identified systems, C2c1 and C2c3, contain RuvC like endonuclease domains distantly related to Cpf1. The third system, Cas13, contains an effector with two predicted HEPN RNase domains.
- Gao *et al.* (2016) reported using a structure-guided saturation mutagenesis screen to increase the targeting range of Cpf1. AsCpf1 variants were engineered with the mutations S542R/K607R and S542R/K548V/N552R that can cleave target sites with TYCV/CCCC and TATV PAMs, respectively, with enhanced activities *in vitro* and in human cells.

[00751] Also, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung *Nature Biotechnology* 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[00752] In addition, mention is made of PCT application PCT/US 14/70057, Attorney Reference 47627.99.2060 and BI-2013/107 entitled "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS (claiming priority from one or more or all of US provisional patent applications: 62/054,490, filed September 24, 2014; 62/010,441, filed June 10, 2014; and 61/915,118, 61/915,215 and 61/915,148, each filed on December 12, 2013) ("the Particle Delivery PCT"), incorporated herein by reference, with respect to a method of preparing an sgRNA-and-Cas9 protein containing particle comprising admixing a mixture comprising an sgRNA and Cas9 protein (and optionally HDR template) with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol; and particles from such a process. For example, wherein Cas9

protein and sgRNA were mixed together at a suitable, e.g., 3:1 to 1:3 or 2:1 to 1:2 or 1:1 molar ratio, at a suitable temperature, e.g., 15-30C, e.g., 20-25C, e.g., room temperature, for a suitable time, e.g., 15-45, such as 30 minutes, advantageously in sterile, nuclease free buffer, e.g., IX PBS. Separately, particle components such as or comprising: a surfactant, e.g., cationic lipid, e.g., 1,2-dioleoyl-3-trimethylammonium -propane (DOTAP); phospholipid, e.g., dimyristoylphosphatidylcholine (DMPC); biodegradable polymer, such as an ethylene-glycol polymer or PEG, and a lipoprotein, such as a low-density lipoprotein, e.g., cholesterol were dissolved in an alcohol, advantageously a C₁₋₆ alkyl alcohol, such as methanol, ethanol, isopropanol, e.g., 100% ethanol. The two solutions were mixed together to form particles containing the Cas9-sgRNA complexes. Accordingly, sgRNA may be pre-complexed with the Cas9 protein, before formulating the entire complex in a particle. Formulations may be made with a different molar ratio of different components known to promote delivery of nucleic acids into cells (e.g. 1,2-dioleoyl-3-trimethylammonium -propane (DOTAP), 1,2-ditetradecanoyl- α -glycero-3-phosphocholine (DMPC), polyethylene glycol (PEG), and cholesterol) For example DOTAP :DMPC :PEG :Cholesterol Molar Ratios may be DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 5, Cholesterol 5. DOTAP 100, DMPC 0, PEG 0, Cholesterol 0. That application accordingly comprehends admixing sgRNA, Cas9 protein and components that form a particle; as well as particles from such admixing. Aspects of the instant invention can involve particles; for example, particles using a process analogous to that of the Particle Delivery PCT, e.g., by admixing a mixture comprising sgRNA and/or Cas9 as in the instant invention and components that form a particle, e.g., as in the Particle Delivery PCT, to form a particle and particles from such admixing (or, of course, other particles involving sgRNA and/or Cas9 as in the instant invention).

[00753] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1: Luciferase assay of Casl3 (Casl3a) in eukaryotic cells

A luciferase targeting assay was performed with different gRNAs directed against Gluc. Casl3 orthologues *Leptotrichia wadei* F0279 (Lw2) and *Listeria newyorkensis* FSL M6-0635 (LbFSL) were fused to an NLS or NES or alternatively were not fused to a localization signal. Normalized protein expression of luciferase was determined and

compared to non targeting (NT) gRNA. Efficient knockdown was apparent. The spacer sequences used in the experiments were:

Guide 1 ATCAGGGCAAACAGAACTTTGACTCCca

Guide 2 AGATCCGTGGTCGCGAAGTTGCTGGCCA

Guide 3 TCGCCTTCGTAGGTGTGGCAGCGTCCTG

Guide NT tagattgctgttctaccaagtaatccat

Guide 1 TCGCCTTCGTA GGTGTGGCAGCGTCCTG

Guide NT tagattgctgttctaccaagtaatccat

(SEQ ID Nos. 118 - 123)

[00754] A targeting assay based on GFP expression was performed with different gRNAs directed against EGFP. Cas13 orthologues *Leptotrichia wadei* F0279 (Lw2) and *Listeria newyorkensis* FSL M6-0635 (LbFSL) were fused to an NLS or NES or alternatively were not fused to a localization signal. Normalized expression of GFP was determined and compared to non targeting (NT) gRNA. Efficient knockdown was apparent. The spacer sequences used in the experiments were as follows:

Guide 1 tgaacagctcctcgcccttgctcacat

Guide 2 tcagcttgccgtaggtggcatcgccctc

Guide 3 gggtagcggctgaagcactgcacgccgt

Guide 4 ggtctttagttgccgtcgtccttgaag

Guide 5 tactccagcttggtccccaggatgttgc

Guide 6 cacgctgccgtcctcgatgttggcgg

Guide 7 tctttgctcagggcggactgggtgctca

Guide 8 gactgtacagctcgtccatgccgagag

Guide NT tagattgctgttctaccaagtaatccat

Guide 2 tcagcttgccgtaggtggcatcgccctc

Guide 3 gggtagcggctgaagcactgcacgccgt

Guide 4 ggtctttagttgccgtcgtccttgaag

Guide 5 tactccagcttggtccccaggatgttgc

Guide 6 cacgctgccgtcctcgatgttggcgg

Guide 7 tctttgctcagggcggactgggtgctca

Guide 8 gactgtacagctcgtccatgccgagag

Guide NT tagattgctgttctaccaagtaatccat
(SEQ ID Nos. 124 - 140)

[00755] A targeting assay was performed on different endogenous target genes in HEK293 cells with gRNAs directed against endogenous target genes. Cas13 *Leptotrichia wadei* F0279 (Lw2) was fused to an NES. Normalized protein expression of the respective target genes was determined (compared to non targeting (NT) gRNA). The spacer sequences used were:

CTNNA1	ctgctgccacagaccgagaggcttaaaa
PPIB	tccttgattacacgatggaatttgctgt
mAPK14	tcaaggtggggtcacaggagaagccaaa
CXCR4	atgataatgcaatagcaggacaggatga
TINCR	gcgtgagccaccgcgcctggccggctgt
PCATI	ccagctgcagatgctgcagttttggcg
CAPN1	ctggaaatggaagatgccggcatagcca
LETMD1	gatgacacctcacacggaccaccctag
MAPK14	taatactgctccagatatgggtgggcca
RBI	catgaagaccgagttatagaatactata
TP53	ggtgaaatattctccatccagtggttc
KRAS	aattctcgaactaatgtatagaaggca

(SEQ ID Nos. 141 - 152)

Example 2: Translation upregulation with catalytically inactive Cas13 fused to a translation activator/promoter in eukaryotic cells

[00756] Catalytically inactive Cas13 orthologues *Leptotrichia wadei* F0279 (Lw) and *Listeria newyorkensis* FSL M6-0635 (LbFSL) were generated.

[00757] Lw Cas13 was fused to an NES or without localization signal and optionally EIF4E

[00758] LbFSL was fused to an NLS and optionally EIF4E

[00759] Glue was targeted.

[00760] Relative protein expression was evaluated based on comparison between targeting with Cas13 with and without EIF4E.

[00761] Efficient translation upregulation was apparent. The spacer sequence used in the experiment is tagattgctgttctaccaagtaatccat, and target has a 3x binding sites for this spacer.

Example 3: Co-localization of Casl3 and its target beta-actin upon treatment with NaAsO₂

[00762] The localization of Casl3 targeting beta-actin under influence of sodium arsenite (NaAsCh) was investigated. A fusion construct of *Leptotrichia wadei* Casl3 with mCherry and NES was made and cloned in a mammalian expression vector together with a guide targeting beta-actin or a non-targeting guide. Cellular localization of Casl3 was evaluated based on mCherry expression, stress granules were labeled with G3BP1-GFP. Lw Casl3 targeting beta actin was found to localize to stress granules upon treatment with NaAsO₂. This localization was guide-dependent as only seen with beta-acting targeting and not with non-targeting guides.

Example 4: Alternative crRNA promoters are used to boost knockdown activity

[00763] In order to further increase the interference effect, the crRNA was placed under the control of the U6 promoter.

[00764] With the aim to improve efficiency of interference by Casl3, expression of genes targeted using tRNA- crRNA and U6 driven crRNA and shRNA were compared. Reliable target gene knockdown was observed with comparable efficiency as shRNA. Further experiments were performed to determine effect of increasing crRNA transfection amount, increasing protein transfection amount and effect of DR-spacer-DR-spacer constructs. It was found that Casl3 outperformed optimized shRNA for corresponding targets on endogenous genes.

Example 5: Fusion constructs with Casl3

[00765] As demonstrated in dLw2Casl3-EIF4E fusion can upregulate translation of three genes; Protein levels as measured by band intensity on western blot.

Example 6: Target induced non-specific RNase activity

[00766] Casl3 target induced non-specific RNase activity is useful to detect RNA species in samples. In the presence of an RNA target of interest, guide-dependent Casl3 nuclease activity is accompanied by non-specific RNase activity against collateral targets. For example, a reporter RNA comprising a fluorescent moiety and a fluorescence quencher is non-specifically cleaved by activated Casl3. An RNA substrate is tagged with a fluorescent reporter molecule (fluor) on one end and a quencher on the other. In the absence of Casl3 RNase activity, the physical proximity of the quencher dampens fluorescence from the fluor

to low levels. When Cas13 target specific cleavage is activated, the RNA substrate is non-specifically cleaved and the fluor and quencher are spatially separated. This causes the fluor to emit a signal when excited by light of the appropriate wavelength.

Example 7: Biochemical characterization of Lw2C2c2 (LwaCas13a)

[00767] Methodology is essentially as described in Abudayyeh et al. (2016) "Cas13 is a single-component programmable RNA-guided RNA-targeting CRISPR effector"; *Science* 353(6299), DOI: 10.1126/science.aaf5573; and is incorporated herein by reference.

[00768] Because of LwaCas13a's desirable cleavage activity in bacterial cells, Applicants explored further biochemical characterization to better understand its cleavage prior to testing in mammalian cells. *In vitro* cleavage reactions with both LshCas13a and LwaCas13a demonstrated programmable target cleavage with a guide encoding a 28 nt spacer and a requirement for Mg²⁺ as well as confirming the *in vivo* increase of efficiency of LwaCas13a over LshCas13a. Incubation of single stranded RNA target (ssRNA 1) with LwaCas13a and guide showed detectable cleavage within 2 minutes with nearly complete cleavage after 30 minutes of incubation, while LwaCas13a without guide had no observable cleavage, and cleavage was dose-dependent with LwaC2c2-guide complex levels. Given the appearance of stereotyped cleavage products, Applicants hypothesized that LwaCas13a cleavage patterns were target-dependent, similar to LshCas13a (Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573, doi:10.1126/science.aaf5573 (2016)). Incubation with multiple RNA targets with various *in silico*-predicted secondary structures revealed substantially different cleavage patterns. To determine if LwaCas13a cleavage depended on base identity in exposed single stranded regions on the target, Applicants incubated LwaCas13a on a target (modified ssRNA target 4) with homopolymer substitutions in a loop. Applicants found stronger cleavage for targets with C or U substitutions, showing that LwaCas13a has more substrate flexibility than LshCas13a, which preferentially cleaves at U residues (Abudayyeh, O. O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353, aaf5573, doi:10.1126/science.aaf5573 (2016)). In addition to target RNase activity, the Cas13 family has been reported to process its own corresponding pre-crRNA transcript from *L. wadiei* (East-Seletsky, A. et al. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection, *Nature* 538, 270-273 (2016)). Applicants also explored the guide constraints on LwaCas13a cleavage by truncating either the spacer or the direct repeat (DR) sequences. Applicants found that LwaCas13a retained *in vitro* cleavage

activity with spacer lengths as short as 20 nt, and could cleave with DR truncations as short as 27 nt, although one DR length truncation (32 nt) seemed to eliminate activity, possibly due to secondary structure perturbation. Although guide lengths less than 20 nt no longer display catalytic activity, the LwaCas13-guide complex could still retain binding activity, allowing for orthogonal applications with a single catalytic enzyme (Dahlman, J. E. *et al* Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol* **33**, 1159-1161, doi:10.1038/nbt.3390 (2015)).

[00769] The Cas13 family has been found to have a dual RNase activity for processing of full-length CRISPR transcripts (East-Seletsky, A. *et al* Two distinct RNase activities of CRISPR-C2c2 enable guide processing and RNA detection. *Nature* **538**, 270-273, doi:10.1038/nature19802 (2016)), in a manner similar to Cpf1 (Fonfara, L, Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517-521, doi:10.1038/nature17945 (2016); Zetsche, B. *et al* Multiplex gene editing by CRISPR-Cpf1 using a single guide array. *Nat Biotechnol* **35**, 31-34, doi:10.1038/nbt.3737 (2017)). Applicants found that LwaCas13a could cleave the corresponding CRISPR spacer transcript from *L. wadei* and this cleavage was concentration dependent. Furthermore, LwaCas13a showed collateral activity on RNA products separated by gel electrophoresis, confirming previous characterization of collateral activity by cleavage of a quenched fluorophore (Gootenberg, J. S. *et al* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **In press** (2017)).

[00770] *Leptotrichia wadei* F0279 (Lwa2) Cas13 was used in *in vitro* assays to evaluate cleavage kinetics, dependence of cleavage activity on the presence of cations, PFS preference, effect of direct repeat length, effect of spacer length, effect of target RNA sequence and secondary structure and nucleotide cut preference.

Example 8: LwaCas13a can be reprogrammed to knockdown reporter mRNA

[00771] Given LwaCas13a's robust RNA cleavage activity and flexible sequence preference, Applicants decided to evaluate its ability to cleave transcripts in mammalian cells. Applicants first cloned mammalian codon-optimized LwaCas13a into mammalian expression vectors with msfGFP fusions on the C- or N-terminus and either a dual-flanking nuclear export sequence (NES) or nuclear localization sequence (NLS) and evaluated expression and localization. Applicants found that msfGFP-fused LwaCas13a constructs expressed well and localized effectively to the cytoplasm or nucleus according to the localization sequence. To evaluate the *in vivo* cleavage activity of LwaCas13a Applicants developed a dual luciferase reporter system, which expresses both Gaussia luciferase (Glue)

and Cypridinia luciferase (Clue) under different promoters on the same vector, allowing one transcript to serve as the Casl3a target and the other to serve as a dosing control. Applicants then designed guides against Glue and cloned them into a tRNA^{Val}-promoter-expressing guide vector. Applicants transfected the LwaCasl3a expression vector, guide vector, and dual-luciferase construct into HEK293FTs and measured luciferase activity at 48 hours post transfection. Applicants found that nuclear-localized LwaCasl3a-msfGFP resulted in the highest levels of knockdown (75.7% for guide 1, 72.9% for guide 2), comparable to position-matched shRNA controls (78.3% for guide 1, 51.1% for geode 2), which control for accessibility and sequence in the target region. Because of the superior cleavage of the LwaCasl3a-msfGFP-NLS construct, Applicants used this design for all further knockdown experiments. The nuclear localized LwaCasl3a-msfGFP also fared better than mCherry-fused versions, likely due to the enhanced stability offered by the msfGFP. The ability to manipulate LwaCasl3a activity by engineered fusions highlights the flexibility of the Casl3a tool. LwaCasl3a is also capable of knockdown in the A375 melanoma cell line, demonstrating the versatility of Casl3. Applicants also found that LwaCasl3a yields the best Glue knockdown with a spacer length of 28 nt (73.8%) and that knockdown is dose-responsive to both the protein and guide or crRNA transfected vector amounts. Guide expression is not sensitive to promoter choice, and guide or crRNAs expressed from the tRNA^{Val} or U6 promoters result in similar levels of Glue knockdown (66.3% for tRNA^{Val}, 74.5% for U6).

[00772] Applicants next tested knockdown on three endogenous genes: *KRAS*, *CXCR4*, and *PPIB*, and found that varying levels of knockdown, and for 2 of 3 genes, LwaCasl3a knockdown (40.4% for *PPIB*, 83.9% for *CXCR4*, 57.5% for *KRAS*) was similar to RNAi with position-matched shRNAs (63.0% for *PPIB*, 73.9% for *CXCR4*, 44.3% for *KRAS*). Applicants also found that endogenous gene knockdown was flexible to guide promoter choice, with similar levels of knockdown for guides expressed from the tRNA^{Val} or U6 promoters (86.7% for tRNA^{Val}, 77.6% for U6). Applicants also found that LwaCasl3a is capable of knockdown in the A375 melanoma cell line. To expand the versatility of LwaCasl3a knockdown, Applicants designed guides against transcripts for rice (*Oryza sativa*) genes *EPSPS*, *HCT*, and *PDS* and co-transfected the LwaCasl3a and guide vectors into *O. sativa* protoplasts. After transfection, Applicants observed >50% knockdown of for all three genes and 7 out of 9 guides tested, with maximal knockdown of 78.0%. Together, these results suggest that LwaCasl3a is able to mediate similar levels of RNA knockdown as

RNAi. Further exploration of additional members of the Cas13 family may reveal proteins able to achieve even more potent knockdown effect.

Example 9: LwaCas13a knockdown screening of reporter and endogenous transcripts

[00773] To comprehensively characterize the dependence of RNA context on the efficiency of LwaCas13a knockdown, Applicants harnessed the programmability of LwaCas13a to tile guides along the length of the Glue, KRAS, PPIB or Clue transcripts. The Glue and Clue screens revealed guides with greater than 60% knockdown, and that a majority of Glue targeting guides had more than 50% knockdown with up to 83% maximal knockdown. To compare LwaCas13a knockdown with RNAi, Applicants selected the top three performing guides against Glue and Clue and compared them to position-matched shRNAs. Applicants found that five out of six top performing guides achieved significantly higher levels of knockdown ($p < 0.05$) than their matched shRNA.

[00774] Having demonstrated robust knockdown on reporter genes, Applicants next explored whether Cas13a could be engineered to target endogenous transcripts via tiling of two genes, KRAS and PPIB. Applicants found that, while knockdown efficiency was transcript dependent, Applicants could still find guides capable of achieving 50% knockdown on either target with maximal knockdown of 85% and 75% for KRAS and PPIB, respectively. Applicants also found that endogenous gene knockdown was flexible to guide expression design, with similar levels of knockdown for crRNAs expressed from the tRNA^{Val} or U6 promoters.

[00775] To further understand the efficiency of LwaCas13 knockdown versus RNAi, Applicants compared a variety of guides to shRNA constructs that were position matched to the same target region. Applicants selected the top three guides from each of the endogenous tiling screens (*KRAS* and *PPIB*) and observed robust knockdown with Cas13a (53.7%-88.8%) equivalent to levels attained by shRNA knockdown (61.8%-95.2%), with shRNA better for 2 out of 6 guides ($p < 0.01$) and Cas13a better for 2 out of 6 guides ($p < 0.01$).

Example 10: LwaCas13a knockdown is optimal at accessible sites in the target transcript

[00776] Since Applicants found that LshCas13a activity was governed by target accessibility in *E. coli* (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)), Applicants decided to investigate whether LwaCas13a activity was increased for guides located in regions of accessibility along the four transcripts targeted in our guide tiling screens. Applicants first found that the most effective guides seemed to cluster into defined regions and by comparing the pair-wise distances between

effective guides to the null distribution, Applicants observed guides are significantly more closer together than would be expected by chance on all four transcripts. These initial clustering results suggest that regions of accessibility may be enriched for better LwaCasl3a cleavage activity.

[00777] To confirm that transcript accessibility influenced LwaCasl3a activity, Applicants computationally predicted accessibility of all target regions across each of the transcripts and found that these computational predictions were partially correlated to knockdown efficiency. Across the four targeted transcripts, predicted target accessibility could explain some of the variation in targeting efficacy (4.4%-16% of the variation in knockdown), indicating that while accessibility is a determinant of knockdown efficiency, other factors such as base-identity, sequence properties and protein binding to the RNA may also play important roles in targeting efficacy. More extensive screening in the future will likely be able to elucidate these mechanisms more clearly.

Example 11: Comparison of LwaCasl3a knockdown and RNAi on endogenous transcripts

[00778] To further understand the efficiency of LwCasl3 knockdown versus RNAi, Applicants compared a variety of guides to shRNA constructs that were matched to the same target region. Applicants first selected the top three guides from each of the endogenous tiling screens (KRAS and PPIB) and observed robust knockdown with Casl3a with knockdown equivalent to shRNA for almost every guide. To further compare to shRNA, Applicants also designed Casl3a crRNAs in regions of accessibility predicted by the RNAXs algorithm for KRAS, PPIB, and CXCR4 and found comparable levels of knockdown to shRNA.

Example 12: LwaCasl3a knockdown screening of MALAT1 lncRNA

Because LwaCasl3a can be engineered for cellular localization, it has versatility for which compartments of the cell can be targeted for RNA knockdown. Applicants designed 93 guides tiled evenly across the entire lncRNA MALAT1 transcript, which is nuclear localized, and transfected these guides with nuclear-localized LwaCasl3a. Applicants found varying levels of knockdown, with up to as much as about 40% to 50% knockdown in one experiment. Compared against position-matched shRNA, which showed no detectable knockdown ($p > 0.05$), Casl3a achieved significantly higher levels of knockdown (39.0-66.5%, $p < 0.05$). Applicants also tiled the lncRNA XIST transcript, and found an average of 22.0% and a maximum of 83.9% knockdown across all guides).

Example 13: Multiplexed knockdown of endogenous transcripts

[00779] Other CRISPR effectors with CRISPR array processing activity, such as Cpf1, have been leveraged for multiplexed gene editing by expressing many guides under one

promoter (Zetsche, B. *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single guide array. *Nat Biotechnol* **35**, 31-34, doi:10.1038/nbt.3737 (2017)). Because LwaCas13a can process its own array, Applicants decided to test multiplexed delivery of LwaCas13a guides as a CRISPR array expressed under a single promoter. Applicants designed five different guides against the endogenous PPIB, CXCR4, KRAS, TINCR, and PCAT transcripts, and delivered the targeting system as a CRISPR array with 28 nt guides flanked by 36 nt direct repeats (DR), representing an unprocessed DR and a truncated spacer, under expression of the U6 promoter. With this approach, Applicants found levels of knockdown for each gene that were comparable to single or pooled guide controls.

[00780] Because of concerns that off-target LwaCas13a activity might be causing non-specific knockdown of the five transcripts targeted by the CRISPR array, Applicants designed an experiment with multiplexed delivery of three guides against PPIB, CXCR4, and KRAS and three variants where each one of the three guides was replaced with a non-targeting guide. Applicants found that in each case where a guide was absent from the array, there was no significant knockdown of the transcript targeted by the missing guide and only the targeted transcripts were knocked down by LwaCas13a, demonstrating that knockdown is not due to nonspecific degradation of the transcripts, but is in fact due to specific, multiplexed knockdown by LwaCas13a.

Example 14: LwaCas13a knockdown is sensitive to mismatches

[00781] Specificity is a central concern for nucleic acid targeting tools, and the specificity of both RNAi and Cas9 DNA-targeting systems (Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838, doi:10.1038/nbt.2675 (2013); Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826, doi:10.1038/nbt.2623 (2013); Pattanayak, V. *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-843, doi:10.1038/nbt.2673 (2013); Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013); Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016)) has been extensively characterized. The initial characterization of LshCas13a showed that it could be sensitive to as few as two mismatches *in vitro* (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)), and

specificity profiling of LwaCas13a via the collateral effect (Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science In press* (2017)) revealed that discrimination could be achieved at double-nucleotide resolution, with single-nucleotide resolution in the seed region of the guide:target duplex. To investigate the specificity of Cas13a *in vivo*, Applicants introduced mismatches into guides targeting either Glue or the endogenous genes CXCR4, KRAS, and PPIB. Applicants found that for all transcripts, the central region of the guide:target duplex was most sensitive to single mismatches, in agreement with the previous *in vitro* characterizations of Cas13a specificity (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)). While knockdown was reduced in the seed region, Applicants found that single mismatches were not sufficient for substantial levels of specificity on some targets, such as Glue. Tiling of consecutive double mismatches for spacers against Glue revealed that double mismatches resulted in up to 8-fold reduction of activity, showing the promise of Cas13a as a specific *in vivo* targeting tool. Applicants also investigated the effect of non-consecutive double mismatches and found that most double mismatches reduced the knockdown from 80.4% to less than 60%, except for double mismatches located in either the 5' or 3' distal ends of the guide sequence.

[00782] For further characterization of Cas13a across multiple specificity parameters *in vitro*, Applicants used detection of collateral activity (Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science In press* (2017)) as a proxy for direct Cas13a activity. Given results from Cas9 experiments showing that specificity could be increased by shorter spacer lengths (Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284, doi:10.1038/nbt.2808 (2014)), Applicants wondered whether spacer length had an effect on Cas13a specificity against two targets that differ by a single mismatch. Applicants found that while shorter spacers have reduced activity, as expected from our *in vivo* LwaCas13a results, shorter spacers also had improved single base-mismatch distinction. Applicants next explored if specificity could be improved by designing an additional synthetic mismatch in the spacer sequence, as this approach has successfully been used for single-mismatch distinction *in vitro* with LwaCas13a (Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science In press* (2017)). Applicants found that, compared to full-length spacers, spacers truncated to either 23 nt or 20 nt had less overall activity but substantially increased specificity. Taken together, the *in vitro* and *in vivo* engineering of LwaCas13a show promise for its use as a specific knockdown tool. The ability

to engineer guides to confer single-base specificity should facilitate allele-specific transcript knockdown by LwaCasl3a.

Example 15: Transcript knockdown with LwaCasl3a is highly specific

[00783] To comprehensively understand if there are any off-target effects of LwaCasl3a knockdown, Applicants performed transcriptome-wide mRNA sequencing. Applicants targeted the Glue transcript with LwaCasl3a or a position matched-shRNA construct, and found significant knockdown of the target transcript. Similar results were found for the same comparison on two endogenous genes KRAS and PPIB. shRNA conditions had more transcriptome-wide variation and weaker correlation between targeting and non-targeting controls than LwaCasl3a conditions, suggesting more off-targets in the shRNA targeting experiment. Applicants further characterized the number of significant off-targets by differential expression analysis and found hundreds of off-targets in each of the shRNA conditions but zero-off targets in LwaCasl3a conditions, despite comparable levels of knockdown of the target transcripts (30.5%, 43.5%, and 64.7% for shRNA, 62.6%, 27.1%, and 29.2% for Casl3a, for Glue, KRAS, and PPIB, respectively). Applicants performed additional analysis of the Glue targeting RNA-seq comparisons, and found that the dominant source of variability in shRNA conditions was due to differences between targeting and non-targeting conditions in individual replicates (average Kendall's tau = 0.917). When individual replicates of the same condition were compared, there were much higher correlations and less variability (average Kendall's tau = 0.941), indicating that the variation observed is from consistent off-target effects of a given shRNA construct. When this analysis is applied across all RNA-seq libraries analyzed for the three genes, all LwaCasl3a conditions have high correlations with each other despite different guide sequences due to the narrow spreads of the transcript distributions. In contrast, the sets of three replicates for each of the shRNA conditions have higher intra-set correlation than between shRNA conditions due to the amount of off-target variation for each different shRNA sequence. Furthermore, when the distribution of standard deviations for each guide condition is compared against each shRNA condition across the three transcripts, there is significantly more variation observed in the shRNA conditions ($p < 10^{-192}$, 2-sided K-S test).

Example 16: LwaCasl3a displays no observable collateral activity in mammalian cells

[00784] The collateral activity of Casl3a has been directly observed biochemically *in vitro* and indirectly through growth suppression in bacteria. Because the multiplexed leave-one-out and RNA-seq analyses suggested a lack of non-specific RNA degradation and thus collateral activity in mammalian cells, Applicants wanted to see if reduction in global off-target

expression due to collateral activity occurred in knockdown experiments. Applicants analyzed the gene controls in the luciferase and endogenous knockdown experiments to see if there was any variation in the controls as a result of target transcript knockdown. From the initial Glue tiling experiment, it was clear that while many guides displayed significant knockdown of Glue, there was little variation in Clue levels. Applicants then decided to analyze the correlations between on-target knockdown to on-target expression or on-target knockdown to off-target expression (the luciferase control or GAPDH in the case of endogenous targeting). Applicants found that for each of the four targets, there was significant positive correlation between on-target knockdown and on-target expression (Glue: $R = 0.89$, $p < 0.0001$; PPIB: $R = 0.81$, $p < 0.0001$; KRAS: $R = 0.52$, $p < 0.0001$) while much weaker or no correlation between the on-target knockdown and control gene expression (Glue: $R = -0.078$, $p > 0.05$; PPIB: $R = -0.058$, $p > 0.05$; KRAS: $R = -0.51$, $p < 0.0001$), indicating that there was no detectable off-target knockdown.

[00785] The lack of any significant correlation between control expression and knockdown suggests that there is little or no collateral activity of LwaCasl3a in mammalian cells. Applicants wanted to further investigate this by seeing if any growth restriction of cells during transcript knockdown would be seen as previously observed in bacteria (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)). Applicants transfected LwaCasl3a with multiple guides against Glue and either with or without selection for 72 hours and then measured knockdown immediately before measuring cell viability and LwaCasl3a expression via GFP fluorescence. Applicants observed significant levels of knockdown for all five Glue targeting guides, but no significant differences in cell growth or GFP fluorescence between the targeting guides and a non-targeting guide control.

[00786] The collateral activity of Casl3a has been directly observed biochemically in vitro and indirectly through growth suppression in bacteria, but the extent of this activity in mammalian cells is unclear. Applicants saw no sequence-specific off-target LwaCasl3a activity in our RNA sequencing experiments, and LwaCasl3a-mediated knockdown of targeted transcripts did not affect the growth of mammalian cells expressing similar levels of LwaCasl3a. Additionally, there were no detectable gene expression changes, indicating that the presence of LwaCasl3a targeting does not lead to an observable cell stress response at the transcriptomic level (Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc. Natl. Acad. Sci USA* **102**, 15545-15550 (2005). In summary, although Applicants cannot rule out the possibility

that low levels of uniform collateral activity cleavage may be occurring, Applicants see no detectable collateral activity across the four following observations: 1) in all of our tiling experiments, Applicants observed no significant correlation between target transcript knockdown and the in-line control gene knockdown, 2) Applicants see minimal disturbance to the transcriptome in our RNA sequencing analysis and no significant off-targets, 3) in the leave one-out-multiplexing experiments Applicants do not see knockdown of the excluded gene, and 4) Applicants do not see phenotypic effects on cellular growth or stress due to LwaCasl3a targeting.

Example 17: dCasl3a programmably binds transcripts in mammalian cells

[00787] As a programmable RNA-binding protein could serve as the foundation for a wide range of applications, Applicants explored whether LwaCasl3a could be engineered as a catalytically inactive variant (dCasl3a). Previous studies have demonstrated that inactivation of LshCasl3a via mutation of catalytic residues eliminated RNase activity, yet maintained RNA-binding (Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573, doi:10.1126/science.aaf5573 (2016)). Applicants mutated catalytic arginine residues in LwaCasl3a to generate dCasl3a and found that targeting of dCasl3a to a 5' UTR upstream of a reporter coding sequence resulted in reduced translation and reporter gene expression. To quantify RNA binding by dCasl3a, Applicants performed RNA immunoprecipitation (RIP) using guides containing the 36 nt DR and 28 nt spacers and found that pulldown of dCasl3a targeted to either luciferase transcripts or ACTB mRNA resulted in significant enrichment of the corresponding target over non-targeting controls (7.8 - 11.2x enrichment for luciferase and 2.1 - 3x enrichment for ACTB; $p < 0.05$), validating dCasl3a as a reprogrammable RNA binding protein.

Example 18: Negative feedback imaging of transcripts with dCasl3a

[00788] To engineer dCasl3a for *in vivo* imaging and reduce background noise due to unbound protein, Applicants incorporated a negative-feedback system based upon zinc finger self-targeting and KRAB domain repression (Gross, G. G. *et al.* Recombinant probes for visualizing endogenous synaptic proteins in living neurons. *Neuron* **78**, 971-985, doi: 10.1016/j.neuron.2013.04.017 (2013)). Fusing a zinc finger, KRAB domain, and NLS to dCasl3a resulted in a negative feedback construct (dCasl3a-NF). When dCasl3a-NF is not bound to its target transcript, it localizes to nucleus and represses its own expression. Upon transcript binding, dCasl3a-NF is exported into the cytoplasm, thereby increasing expression, although it is also possible that newly translated dCasl3a-NF remains resident in the cytoplasm. In comparison to dCasl3a, which showed modest levels of cytoplasmic

translocation (or retention) as a result of transcript binding, dCas13a-NF effectively translocated or re-localized when targeted to ACTB mRNA. To further characterize the degree of translocation of dCas13a-NF, Applicants targeted ACTB transcripts with two guides and found that both guides increased translocation compared to a non-targeting guide (3.1 - 3.7x cellular/nuclear signal ratio; $p < 0.0001$). Quantification of translocation showed that targeting guides resulted in significantly more fluorescence fraction in the cytoplasm than a non-targeting guide, showing the utility of dCas13a-NF as a transcript imaging tool. To further validate dCas13a-NF imaging, Applicants analyzed the correlation of dCas13a-NF signal to ACTB mRNA fluorescent in situ hybridization (FISH) signal and found that there was significant correlation and signal overlap for the targeting guides versus the non-targeting guide conditions ($R = 0.27$ and 0.30 for guide 1 and 2, respectively, and $R = 0.00$ for the non-targeting guide condition; $p < 0.0001$).

Example 19: dCas13a imaging of stress granules in live cells

[00789] Oxidative stress results in the aggregation of polyadenylated transcripts and proteins into stress granules within the cytoplasm (Nelles, D. A. *et al.* Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**, 488-496, doi:10.1016/j.cell.2016.02.054 (2016); Unsworth, H., Raguz, S., Edwards, H. J., Higgins, C. F. & Yague, E. mRNA escape from stress granule sequestration is dictated by localization to the endoplasmic reticulum. *FASEB J* **24**, 3370-3380, doi: 10.1096/fj.09-151142 (2010)), and the development of stress granules has been associated with many pathologies, including cancer, neurodegenerative disease, and myopathies (Wyss-Coray, T. Ageing, neurodegeneration and brain rejuvenation. *Nature* **539**, 180-186, doi:10.1038/nature20411 (2016); Protter, D. S. & Parker, R. Principles and Properties of Stress Granules. *Trends Cell Biol* **26**, 668-679, doi:10.1016/j.tcb.2016.05.004 (2016)). Applicants investigated the accumulation of mRNA into stress granules by combining dCas13a-NF imaging of transcripts with visualization of a well known marker of stress granules, G3BP1 (Tourriere, H. *et al.* The RasGAP-associated endoribonuclease G3BP assembles stress granules. *J Cell Biol* **160**, 823-831, doi:10.1083/jcb.200212128 (2003)). To confirm mRNA tracking in fixed samples, Applicants co-transfected either of two ACTB targeting crRNAs or guides with dCas13a-NF, induced stress granule formation with sodium arsenite, and visualized G3BP1 with immunofluorescence. dCas13a-NF translocated in or re-localized to the cytoplasm for the targeting conditions as expected, and Applicants found significant correlations between the dCas13a-NF signal and the G3BP1 fluorescence compared to the non-targeting control ($R = 0.49$ and 0.50 for guide 1 and guide 2, respectively, and 0.08 for the non-targeting guide;

p<0.001), suggesting the ability of dCas13a-NF to track stress granule formation. Given co-localization in fixed samples, Applicants next performed stress granule tracking in live cells. Applicants transfected ACTB targeting guide and non-targeting guide with dCas13a-NF, induced stress granule formation with sodium arsenite 24 hours post-transfection, and imaged the live cells over time. Using G3BP1-RFP fusion as a stress granule marker, Applicants found that the dCas13a-NF targeted to ACTB localized to significantly more stress granules per cell over time than the corresponding non-targeting control (p<0.05).

Example 20: Discussion

[00790] The class 2 type VI CRISPR-Cas effector Cas13a can be effectively reprogrammed with crRNAs or guides to knockdown or bind transcripts in mammalian cells. Applicants identified LwaCas13a as the most active of fifteen Cas13a orthologs for RNA cleavage in bacteria and harnessed it for mammalian RNA knockdown with levels comparable to RNAi. Applicants found that there was no detectable PFS through bacterial screening and that guide activity was not influenced by PFS. LwaCas13a is sensitive to mismatches in the spacer-target duplex *in vivo*, and this sensitivity translates into high specificity of knockdown compared to RNAi. Applicants also showcase unique attributes of LwaCas13a as an RNA knockdown tool, including the ability to further engineer and optimize the protein, multiplexed delivery of guides, and knockdown of nuclear lncRNAs. Importantly, Applicants observe no collateral activity of LwaCas13a, a feature that is highly active *in vitro* and useful for many applications, such as diagnostics (Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science In press* (2017)). Furthermore, Applicants show that LwaCas13a can be rendered catalytically inactive, such that it can be used as a programmable RNA binding platform, and Applicants demonstrate its utility for tracking transcript accumulation in stress granules in live cells.

[00791] Importantly, Applicants observe no collateral activity of LwaCas13a in mammalian cells, a feature that Applicants observed *in vitro* and harnessed for diagnostics applications (Gootenberg, J.S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2, *Science*, in press (2017)). Collateral activity has been hypothesized to be part of a programmed cell death/dormancy pathway in native bacterial cells, which would remove infected cells from the population or provide infected cells time to adapt and overcome infection, supplementing the on-target viral transcript cleavage activity of Cas13a. The lack of collateral activity in mammalian cells does not preclude the possibility of its existence in the native cellular context.

[00792] There are numerous opportunities for refinement and diversification of RNA-targeting tools based upon Cas13 family members. *In vivo* characterization of additional Cas13 proteins, such as Cas13b (Smargon, A. A. *et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell* **65**, 618-630 e617, doi:10.1016/j.molcel.2016.12.023 (2017)) or Cas13c (Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182, doi:10.1038/nrmicro.2016.184 (2017)), may yield further improvements in cleavage or binding capacity and enable applications requiring orthogonal RNA binding proteins, including multi-color imaging. Additionally, smaller orthologs would allow for size-constrained delivery options such as Adeno-associated viral vectors (Ran, F. A. *et al.* In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186-191, doi: 10.1038/nature 14299, nature14299 [pii] (2015)). Lastly, exploration of the diversity of Cas13 members coupled with increased structural data (Liu, L. *et al.* Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168**, 121-134 e112, doi:10.1016/j.cell.2016.12.031 (2017)) may allow for either bioinformatics-(Zinn, E. *et al.* In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. *Cell Rep* **12**, 1056-1068, doi: 10.1016/j.celrep.2015.07.019 (2015)) or structure-(Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495, doi:10.1038/nature16526 (2016); Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88, doi:10.1126/science.aad5227 (2016)) guided rational design. Improved RNA binding tools will allow additional functionalizations, including imaging via reconstitution of split fluorophores (Ozawa, T., Natori, Y., Sato, M. & Umezawa, Y. Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nat Methods* **4**, 413-419, doi:10.1038/nmeth1030 (2007)), translational modulation (De Gregorio, E., Preiss, T. & Hentze, M. W. Translation driven by an eIF4G core domain in vivo. *EMBO J* **18**, 4865-4874, doi: 10.1093/emboj/18.17.4865 (1999); Adamala, K. P., Martin-Alarcon, D. A. & Boyden, E. S. Programmable RNA-binding protein composed of repeats of a single modular unit. *Proc Natl Acad Sci U S A* **113**, E2579-2588, doi:10.1073/pnas.1519368113 (2016); Campbell, Z. T., Valley, C. T. & Wickens, M. A protein-RNA specificity code enables targeted activation of an endogenous human transcript. *Nat Struct Mol Biol* **21**, 732-738, doi: 10.1038/nsmb.2847 (2014); Cao, J. *et al.* Light-inducible activation of target mRNA translation in mammalian cells. *Chem Commun (Camb)* **49**, 8338-8340, doi:10.1039/c3cc44866e (2013); Cooke, A., Prigge, A., Opperman, L. & Wickens, M. Targeted translational regulation using the PUF

protein family scaffold. *Proc Natl Acad Sci U S A* **108**, 15870-15875, doi:10.1073/pnas.1105151108 (2011)), RNA base editing (Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* **17**, 83-96, doi:10.1038/nrm.2015.4 (2016); Wedekind, J. E., Dance, G. S., Sowden, M. P. & Smith, H. C. Messenger RNA editing in mammals: new members of the APOBEC family seeking roles in the family business. *Trends Genet* **19**, 207-216, doi:10.1016/S0168-9525(03)00054-4 (2003)), epitranscriptomic perturbation (Harcourt, E. M., Kietrys, A. M. & Kool, E. T. Chemical and structural effects of base modifications in messenger RNA. *Nature* **541**, 339-346, doi:10.1038/nature21351 (2017)), targeted induction of apoptosis based on RNA expression levels (Rider, T. H. *et al.* Broad-spectrum antiviral therapeutics. *PLoS One* **6**, e22572, doi:10.1371/journal.pone.0022572 (2011)), or splicing modulation (Wang, Y., Cheong, C. G., Hall, T. M. & Wang, Z. Engineering splicing factors with designed specificities. *Nat Methods* **6**, 825-830, doi:10.1038/nmeth.1379 (2009)).

[00793] RNA knockdown with Cas13a can be applied to perturbing RNAs in multiple biological contexts, including genome-wide pooled knockdown screening, interrogation of lncRNA and nascent transcript function, allele-specific knockdown, and RNA viral therapeutics. In addition, dCas13a and derivatives enable RNA pulldown to study RNA-protein interactions, tracking of transcripts in live cells, and targeted destruction of cells based on RNA levels, which would be useful for studying specific cell populations or killing cancerous cells. Applicants have shown Cas13 to be a robust platform for both programmable knockdown and binding of RNAs in mammalian and plant cells, and this platform may be extended to other eukaryotic organisms. CRISPR-Cas13 coupled with creative engineering approaches will be a powerful platform for nucleic acid based diagnostics and therapeutics and can usher a revolution for studying the transcriptome.

Example 21: Split designs for Apoptosis

[00794] It is often desirable to deplete or kill cells based on transcriptional signatures or specific gene expression, either for basic biology applications to study the role of specific cell types or for therapeutic applications such as cancer or senescent cell clearance (Baker, D.J., Childs, B.G, Durik, M, Wijers, M.E., Sieben, C.J., Zhong, J., Saltness, R.A., Jeganathan, K.B., Verzosa, G.C., Pezeshki, A., *et al.* (2016). Naturally occurring pl6(Ink4a)-positive cells shorten healthy lifespan. *Nature* **530**, 184-189.). This targeted cell killing can be achieved by fusing split apoptotic domains to Cas13 proteins, which upon binding to the transcript are reconstituted, leading to death of cells specifically expressing targeted genes or sets of genes. In certain embodiments, the apoptotic domains may be split Caspase 3 (Chelur,

D.S., and Chalfie, M. (2007). Targeted cell killing by reconstituted caspases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 2283-2288.). Other possibilities are the assembly of Caspases, such as bringing two Caspase 8 (Pajvani, U.B., Trujillo, M.E., Combs, T.P., Iyengar, P., Jelicks, L., Roth, K.A., Kitsis, R.N., and Scherer, P.E. (2005). Fat apoptosis through targeted activation of caspase 8: a new mouse model of inducible and reversible lipoatrophy. *Nat. Med.* 11, 797-803.) or Caspase 9 (Straathof, K.C., Pule, M.A., Yotnda, P., Dotti, G., Vanin, E.F., Brenner, M.K., Heslop, H.E., Spencer, D.M., and Rooney, CM. (2005). An inducible caspase 9 safety switch for T-cell therapy. *Blood* 105, 4247-4254.) effectors in proximity via Cas13 binding. It is also possible to reconstitute a split TEV (Gray, D.C, Mahrus, S., and Wells, J.A. (2010). Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* 142, 637-646.) via Cas13 binding on a transcript. This split TEV can be used in a variety of readouts, including luminescent and fluorescent readouts (Wehr, M.C, Laage, R., Bolz, U., Fischer, T.M., Grunewald, S., Scheek, S., Bach, A., Nave, K.-A., and Rossner, M.J. (2006). Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods* 3, 985-993.). One embodiment involves the reconstitution of this split TEV to cleave modified pro-caspase 3 or pro-caspase 7 (Gray, D.C, Mahrus, S., and Wells, J.A. (2010). Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* 142, 637-646), resulting in cell death.

[00795] Inducible apoptosis. Guides depicted in Fig. 2 were used to locate Cas13 complexes bearing functional domains to induce apoptosis along a luciferase transcript. The Cas13 (Cas13b) is from *Prevotella* sp. P5-125. Functional domains are fused at the C-terminus of the protein. The Cas13 is catalytically inactive via histidine mutations in both HEPN domains (H133A and H1058A). The adaptability of system was demonstrated by employing various methods of caspase activation and optimization of guide spacing along a transcript. Caspase 8 and caspase 9 (aka "initiator" caspases) activity was induced using Cas13 complex formation on a luciferase transcript to bring together caspase 8 or caspase 9 enzymes associated with Cas13. Alternatively, caspase 3 and caspase 7 (aka "effector" caspases) activity was induced when Cas13 complexes bearing tobacco etch virus (TEV) N-terminal and C-terminal portions ("snipper") were maintained in proximity, activating the TEV protease activity and leading to cleavage and activation of caspase 3 or caspase 7 proteins. The system was also tested using split caspase 3, with heterodimerization of the caspase 3 portions by attachment to Cas13 complexes bound to a luciferase transcript. Figs. 4A-4D show a comparison of apoptosis induction by TEV-dependent activation of caspase 7 or caspase 3, or by dimerization of caspase 8 or caspase 9 was compared. Guide pairs 1-6

indicate the seed guide paired with each of guides 1-6. Exemplary apoptotic components are set forth in the table below. Figs. 5A-C show cell death data normalized to cell survival and relative to the non-targeting condition for all four caspase variants (A) as well as the SNIPPER conditions TEV protease/caspase 7 (B) and TEV protease/caspase 3 (C) separately. Figs. 5 D-F show cell death relative to the non-targeting condition for the variants together (D) and the SNIPPER conditions TEV protease/caspase 7 (E) and TEV protease/caspase 3 (F) separately. Figs. 5 G-I shows overall cell death ratio relative for the caspase variants together (G) and the SNIPPER conditions TEV protease/caspase 7 (H) and TEV protease/caspase 3 (I) separately.

iCasp9 (SEQ ID No. 153)	GFGDVGALSLRGNADLAYILSMPCGHCLIINNVNF CRESGLRTRTGSNIDCEKLRRRFSSLHFMVEVKGDLT AKKMVLALLELARQDHGALDCCVVILSHGCQASH LQFPGAVYGTGCPVSVKIVNIFNGTSCPSLGGKPK LFFIQACGGEQKDHGFEVASTSPEDESPGSNPEPDATP FQEGLRTFDQLDAISSLPTSDIFVSYSTFPGFVSWRD PKSGSWYVETLDDIFEQWAHSEDLQSLLLRVANAVS VKGIYKQMPGCFNFLRKKLFFKTSVD	Straathof, K.C., et al. (2005) Blood 105, 4247-4254
Caspase 8 (SEQ ID No. 154)	SESQTLDKVYQMKSKPRGYCLIINNHNF AKAREKVP KLHSIRDRNGTHLDAGALTTTFFELHFEIKPHDDCTV EQIYEILKIYQLMDHSNMDCFICILSHGDKGIIYGTD GQEAPIYELTSQFTGLKCPFLAGKPKVFFIQACQDN YQKGIPVETDSEEQPYLEMDLSSPQTRYIPDEADFLG MATVNNCVSYRNPAEGTWYIQLCQSLRERCPRGDD ILTILTEVNIEVSNKDDKKNMGKQMPQPTFTLRKKL VFPSD	Pajvani, U.B., et al. (2005). Nat. Med. 11, 797- 803
Split caspase 3 (p12) (SEQ ID No. 155)	SGVDDDMACHKIPVEADFLYAYSTAPGYYSWRNSK DGSWFIQSLCAMLKQYADKLEFMHILTRVNRKVATE FESFSFDATFHAKKQIPCIVSMLTKELYFYH	Chelur, D.S., and Chalfie, M. (2007). Proc. Natl. Acad. Sci. U. S. A. 104, 2283-2288
Split caspase 3 (p17) (SEQ ID No. 156)	SGISLDNSYKMDYPEMGLCIIINKNFHKSTGMTSRS GTDVDAANLRETFRNLKYEVRNKNDLTREEIVELMR DVSKEHDHRSRSSFVCLVLSHGEEGIIFGTNGPVDLKKI TNFFRGDRCRSLTGKPKLFIQACRGTELDGCIETD	Chelur, D.S., and Chalfie, M. (2007). Proc. Natl. Acad. Sci. U. S. A. 104, 2283-2288
SNIPPER N- TEV (SEQ ID No. 157)	GESLFKGPRDYNPISSTICHLTNE SDGHTTSLYGIGFGP FIITNKHLFRRNNGTLLVQSLHGVFKVKNTTTLQQHL IDGRDMIIRMPKDFPPFPQKLKFRPQREERICLVTTN FQT	Gray, D.C., et al. (2010) Cell 142, 637-646
SNIPPER C- TEV (SEQ ID No. 158)	KSMSSMVS DTSCTFPSSDGIFWKHWIQTKDGQCGSPL VSTRDGFIVGIHSASNFTNTN NYFTSVPKNFMELLTN QEAQQWVSGWRLNADSVLWGGHKVFMV	Gray, D.C., et al. (2010) Cell 142, 637-646

158)	<p>SNIPPER Caspase 7 (SEQ ID No. 159)</p> <p>ATGGCAGATGATCAGGGCTGTATTGAAGAGCAGG GGGTTGAGGATTCAGCAAATGAAGATTCAGTGGAA AATCTCTACTTCCAGGCTAAGCCAGACCCGGTCCTC GTTTGTACCGTCCCTCTTCAGTAAGAAGAAGAAAA ATGTCACCATGCGATCCATCAAGACCACCCGGGAC CGAGTGCCTACATATCAGTACAACATGAATTTTGA AAAGCTGGGCAAATGCATCATAATAACAACAAG AACTTTGATAAAGTGACAGGTATGGGCGTTCGAAA CGGAACAGACAAAGATGCCGAGGCGCTCTTCAAGT GCTTCCGAAGCCTGGGTTTTGACGTGATTGTCTATA ATGACTGCTCTTGTGCCAAGATGCAAGATCTGCTT AAAAAAGCTTCTGAAGAGGACCATACAAATGCCG CCTGCTTCGCCTGCATCCTCTTAAGCCATGGAGAA GAAAATGTAATTTATGGGAAAGATGGTGTACACC AATAAAGGATTTGACAGCCACTTTAGGGGGGATA GATGCAAAACCCTTTTAGAGAAACCCAAACTCTTC TTCATTCAGGCTTGCCGAGGGACCGAGCTTGATGA TGGCATCCAGGCCGAAAATCTCTACTTCCAGTCGG GGCCCATCAATGACACAGATGCTAATCCTCGATAC AAGATCCCAGTGGAAGCTGACTTCCTCTTCGCCTA TTCCACGGTTCCAGGCTATTACTCGTGGAGGAGCC CAGGAAGAGGCTCCTGGTTTGTGCAAGCCCTCTGC TCCATCCTGGAGGAGCACGGAAAAGACCTGGAAA TCATGCAGATCCTCACCAGGGTGAATGACAGAGTT GCCAGGCACTTTGAGTCTCAGTCTGATGACCCACA CTTCCATGAGAAGAAGCAGATCCCCTGTGTGGTCT CCATGCTCACCAAGGAACTCTACTTCAGTCAA</p>	<p>Gray, D.C., et al. (2010) Cell 142, 637-646</p>
<p>SNIPPER Caspase 3 (SEQ ID No. 160)</p>	<p>ATGGAGAACACTGAAAACCTCAGTGGATTCAAATC CATTAAAAATTTGGAACCAAAGATCATACATGGAA GCGAATCAATGGAAAATCTCTACTTCCAGTCTGGA ATATCCCTGGACAACAGTTATAAAATGGATTATCC TGAGATGGGTTTATGTATAATAATTAATAAAGA ATTTTCATAAAAGCACTGGAATGACATCTCGGTCT GGTACAGATGTTCGATGCAGCAAACCTCAGGGAAA CATTACAGAACTTGAAATATGAAGTCAGGAATAAA AATGATCTTACACGTGAAGAAATTGTGGAATTGAT GCGTGATGTTTCTAAAGAAGATCACAGCAAAAGGA GCAGTTTTGTTTGTGTGCTTCTGAGCCATGGTGAAG AAGGAATAATTTTTGGAACAAATGGACCTGTTGAC CTGAAAAAATAACAAACTTTTTTCAGAGGGGATCG TTGTAGAAGTCTAACTGGAAAACCCAACTTTTCA TTATTCAGGCCTGCCGTGGTACAGAACTGGACTGT GGCATTGAGACAGAAAATCTCTACTTCCAGAGTGG TGTTGATGATGACATGGCGTGTGCATAAAATACCAG TGGAGGCCGACTTCTTGTATGCATACTCCACAGCA CCTGGTTATTATTCTTGGCGAAATTCAAAGGATGG CTCCTGGTTCATCCAGTCGCTTTGTGCCATGCTGAA ACAGTATGCCGACAAGCTTGAATTTATGCACATTC TTACCCGGGTTAACC GAAAGGTGGCAACAGAATTT</p>	<p>Gray, D.C., et al. (2010) Cell 142, 637-646</p>

	GAGTCCTTTTCCTTTGACGCTACTTTTCATGCAAAG AAACAGATTCCATGTATTGTTTCCATGCTCACAAA AGAACTCTATTTTTATCAC	
--	--	--

Example 22: Split designs for imaging

[00796] Split-fluorophore constructs were designed for imaging with reduced background via reconstitution of a split fluorophore upon binding of two Cas13 proteins to a transcript. These split proteins include iSplit {Filonov, G.S., and Verkhusha, V.V. (2013). *A near-infrared BiFC reporter for in vivo imaging of protein-protein interactions. Chem. Biol. 20, 1078–1086.*}, Split Venus (Wu, B., Chen, J., and Singer, R.H. (2014). *Background free imaging of single mRNAs in live cells using split fluorescent proteins. Sci. Rep. 4, 3615.*), and Split superpositive GFP (Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). *Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions in vivo. Mol. Biosyst. 8, 2036–2040.*). Such proteins are set forth in the table below:

iSplit PAS domain of iRFP (N-term) (SEQ ID No. 161)	MAEGSVARQPDLLTCDDEPIHIPGAIQPHG LLLALAADMIVAGSDNLPGLTGLAIGALI GRSAADVDFDSETHNRLTIALAEPGAAVGA PITVGFTMRKDAGFIGSWHRHDQLIFLELE PPQRRGGSEVSALEKEVSALEKEVSALEKE VSALEKEVSALEKGGGS*	Filonov, G.S., and Verkhusha, V.V. (2013). <i>Chem. Biol. 20, 1078–1086</i>
iSplit GAFm domain of iRFP (C-term) (SEQ ID No. 162)	MGGSKVSALKEKVSALKEKVSALKEKVS ALKEKVSALKEGGSPQRDVAEPQAFFRR TNSAIRRLQAAETLESACAAAQEVKRT GYDRVMIYRFASDFSGEVIAEDRCAEVES KLGLHYPASTVPAQARRLYTINPVRIIPDIN YRPVPVTPYLNPTGRPIDLSFAILRSVSPV HLEFMRNIGMHGTMSISILRGERLWGLIVC HHRTPYYVDLDGRQACELVAQVLARQIG VMEE*	Filonov, G.S., and Verkhusha, V.V. (2013). <i>Chem. Biol. 20, 1078–1086</i>
Split Venus N-term (SEQ ID No. 163)	MVSKGEELFTGVVPILVELDGDVNGHKFS VSGEGEGDATYGKLTCLKICTTGKLPVPW PTLVTTLGYGLQCFARYPDHMKQHDFFKS AMPEGYVQERTIFFKDDGNYKTRAEVKFE GDTLVNRIELKGIDFKEDGNILGHKLEYN YNSHNVYIT*	Wu, B., Chen, J., and Singer, R.H. (2014). <i>Sci. Rep. 4, 3615.</i>
Split Venus C-term (SEQ ID No. 164)	ADKQKNGIKANFKIRHNIEDGGVQLADHY QQNTPIGDGPVLLPDNHLYSYQSALSNDP NEKRDHMLVLEFVTAAGITLGMDELYK	Wu, B., Chen, J., and Singer, R.H. (2014). <i>Sci. Rep. 4, 3615.</i>
Split superpositive GFP N-term (SEQ ID No. 165)	SKGERLFRGKVPILVELKGDVNGHKFSVR GEGKGDATRGLTLKFICTTGKLPVPWPT LVTTLTYGVQCFSRYPKHMKRHDFKSA MPKGYVQERTISFKKDGKYKTRAEVKFE GRTLNVNRIKLGKGRDFKEKGNILGHKLRYN	Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). <i>Mol. Biosyst. 8, 2036–2040.</i>

	FNSHKVYITADKR	
Split superpositive GFP C-term (SEQ ID No. 166)	KNGIKAKFKIRHNVKDGSVQLADHYQQN TPIGRGPVLLPRNHYLSTRSKLSKDPKEKR DHMVLLEFVTAAGIKHGRDERYK	Blakeley, B.D., Chapman, A.M., and McNaughton, B.R. (2012). <i>Mol. Biosyst.</i> 8, 2036-2040.

[00797] Tobacco etch virus can be used for modulating components of imaging systems such as, but not limited to fluorophores, and including spatial and temporal control. TEV can be adapted to cleavage of blocking groups that inhibit fluorescence. TEV can be adapted to cleave degrons from proteins such as transcription factors or other proteins to promote expression. TEV can be used to cleave localization factors, for example to induce relocation of imaging components within a cell, including but not limited to nucleus, membranes, and organelles.

Example 23: Split designs for additional transcriptional activity, luciferase

[00798] Additional possible split fusions that could be constituted by Cas13 proteins could include luciferase for luminescent imaging (Kim, S.B., Ozawa, T., Watanabe, S., and Umezawa, Y. (2004). *High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split Renilla luciferase. Proc. Natl. Acad. Sci. U. S. A.* 101, 11542–11547.) or split transcription factors to drive expression of genes of genetic circuits in an RNA-sensing based manner. Possible split transcription factors include split-ubiquitin based systems, such as the split-ubiquitin-LexA system (Petschnigg, J., Groisman, B., Kotlyar, M., Taipale, M., Zheng, Y., Kurat, C.F., Sayad, A., Sierra, J.R., Mattiazzi Usaj, M., Snider, J., et al. (2014). *The mammalian-membrane two-hybrid assay (MaMTH) for probing membrane-protein interactions in human cells. Nat. Methods* 11, 585–592)

* * *

[00799] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the

invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. A non-naturally occurring or engineered composition comprising:
a CRIPSR protein linked to an inactive first portion of an enzyme or reporter moiety, wherein the enzyme or reporter moiety is reconstituted when contacted with a complementary portion of the enzyme or reporter moiety.
2. The composition of claim 1, wherein the enzyme or reporter moiety comprises a proteolytic enzyme.
3. The composition of any one of claims 1 to 2, wherein the CRISPR protein is a first CRISPR protein and the composition further comprises a second CRISPR protein linked to the complementary portion of the enzyme or reporter moiety.
4. The composition of claim 3, wherein the first and second CRISPR proteins comprise RNA binding proteins.
5. The composition of any one of claims 1 to 4, wherein the first CRISPR protein and the second CRISPR protein are the same or different.
6. The composition of any one of claims 3 to 5, which further comprises:
 - i) a first guide capable of binding to the first CRISPR protein and hybridizing to a first target sequence of a nucleic acid, and
 - ii) a second guide capable of binding to the second CRISPR protein and hybridizing to a second target sequence of the nucleic acid.
7. The composition of any one of claims 2 to 6, wherein the proteolytic enzyme comprises a caspase.
8. The composition of any one of claims 2 to 6, wherein the proteolytic enzyme comprises caspase 8 or caspase 9.
9. The composition of any one of claims 2 to 6, wherein the proteolytic enzyme comprises caspase 3 or caspase 7.

10. The composition of any one of claims 2 to 6, wherein the first portion of the proteolytic enzyme comprises caspase 3 p12 and the complementary portion of the proteolytic enzyme comprises caspase 3 p17.

11. The composition of any one of claims 2 to 6, wherein the proteolytic enzyme comprises tobacco etch virus (TEV).

12. The composition of any one of claims 2 to 11, which further comprises a substrate of the proteolytic enzyme.

13. The composition of claim 12, wherein the substrate is engineered to comprise a cleavage site for the proteolytic enzyme.

14. The composition of claim 12 or 13, wherein the substrate comprises a procaspase and a TEV cleavage site.

15. The composition of claim 12 or 13, wherein the substrate comprises a fluorescent protein and a TEV cleavage site.

16. The composition of claim 12 or 13, wherein the substrate comprises a luminescent protein and a TEV cleavage site.

17. The composition of any one of claims 14 to 16, wherein cleavage at the TEV cleavage site activates the substrate.

18. The composition of any one of claims 14 to 16, wherein cleavage at the TEV cleavage site inactivates the substrate.

19. A method of providing a proteolytic activity in a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises:

- i) a first CRISPR protein linked to an inactive first portion of a proteolytic enzyme;
- ii) a second CRISPR protein linked to a complementary portion of the proteolytic enzyme wherein proteolytic activity of the proteolytic enzyme is reconstituted when the first portion and the complementary portion of the proteolytic enzyme are contacted;

iii) a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA;

iv) a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA,

whereby the first portion and a second portion of the proteolytic enzyme are contacted and the proteolytic activity of the proteolytic enzyme is reconstituted.

20. A method of inducing cell death in a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises:

i) a first CRISPR protein linked to an inactive first portion of an enzyme capable of inducing cell death;

ii) a second CRISPR protein linked to a complementary portion of the enzyme wherein activity of the enzyme is reconstituted when the first portion and the complementary portion of the enzyme are contacted;

iii) a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA; and

iv) a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA,

whereby the first portion and a second portion of the enzyme are contacted and the activity of the enzyme is reconstituted and induces cell death.

21. The method of claim 20, wherein the enzyme is a proteolytic enzyme.

22. The method of claim 20, wherein the enzyme is a caspase.

23. The method of claim 21, wherein the proteolytic enzyme is TEV protease, wherein the proteolytic activity of the TEV protease is reconstituted, whereby a TEV substrate is cleaved and activated.

24. The method of claim 23, wherein the TEV substrate is a procaspase engineered to contain TEV target sequences whereby cleavage by the TEV protease activates the procaspase.

25. A method of identifying a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises:

i) a first CRISPR protein linked to an inactive first portion of a proteolytic enzyme;

- ii) a second CRISPR protein linked to a complementary portion of the proteolytic enzyme wherein activity of the proteolytic enzyme is reconstituted when the first portion and the complementary portion of the proteolytic enzyme are contacted;
 - iii) a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA;
 - iv) a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA; and
 - v) a reporter which is detectably cleaved,
- wherein the first portion and a second portion of the proteolytic enzyme are contacted when the RNA of interest is present in the cell, whereby the activity of the proteolytic enzyme is reconstituted and detectably cleaves the reporter.

26. A method of identifying a cell which contains an RNA of interest, which comprises contacting the RNA in the cell with a composition which comprises:

- i) a first CRISPR protein linked to an inactive first portion of a reporter;
- ii) a second CRISPR protein linked to a complementary portion of the reporter wherein activity of the reporter is reconstituted when the first portion and the complementary portion of the reporter are contacted;
- iii) a first guide that binds to the first CRISPR protein and hybridizes to a first target sequence of the RNA;
- iv) a second guide that binds to the second CRISPR protein and hybridizes to a second target sequence of the RNA; and
- v) the reporter,

wherein the first portion and a second portion of the reporter are contacted when the RNA of interest is present in the cell, whereby the activity of the reporter is reconstituted.

27. The method of claim 25 or 26, wherein the reporter is a fluorescent protein.

28. The method of claim 27, wherein the fluorescent protein is a green fluorescent protein.

29. The method of claim 25 or 26, wherein the reporter is a luminescent protein.

30. The method of claim 29, wherein the luminescent protein is a luciferase.

1/39

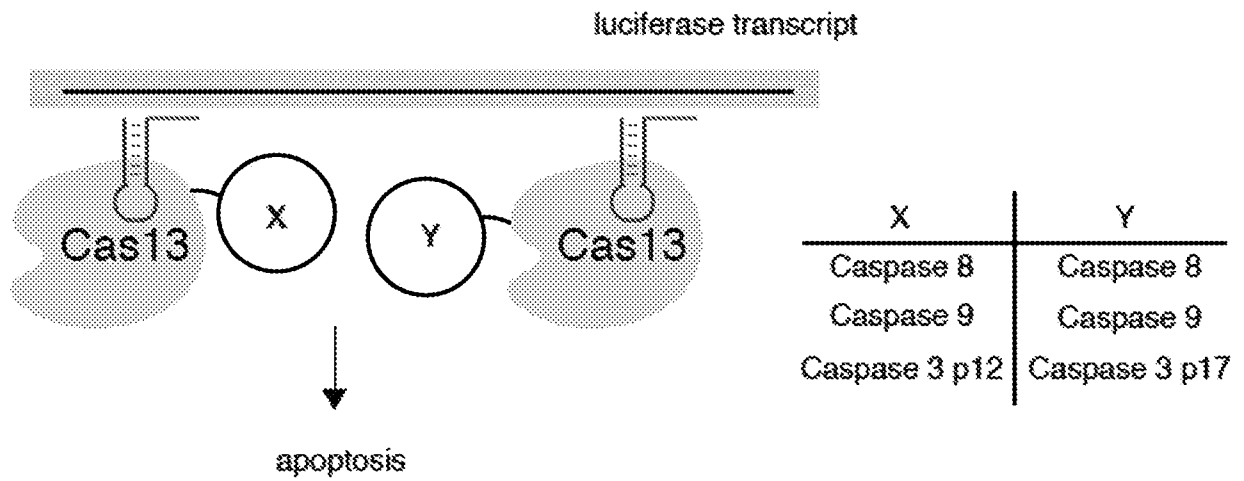


FIG. 1A

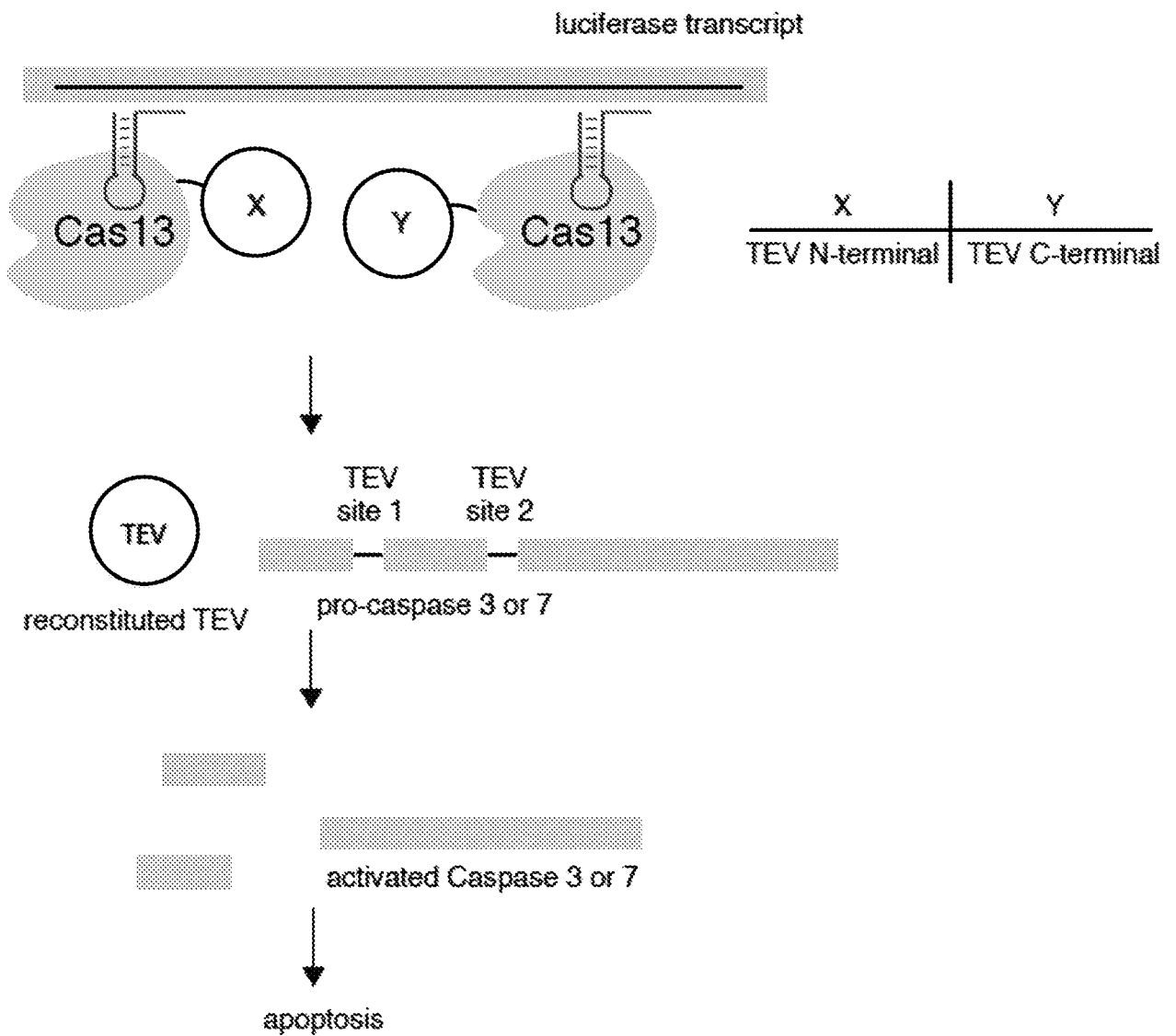
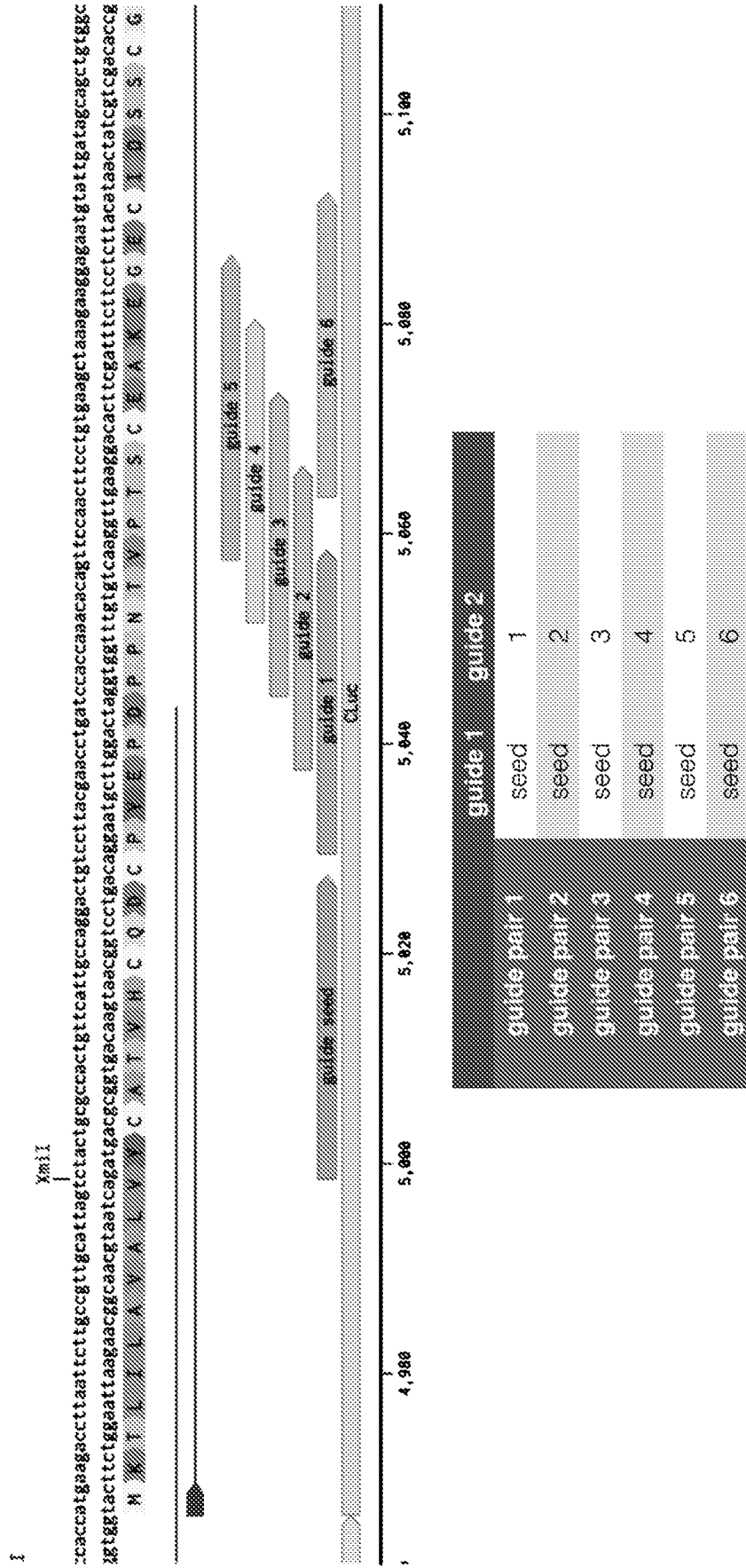


FIG. 1B



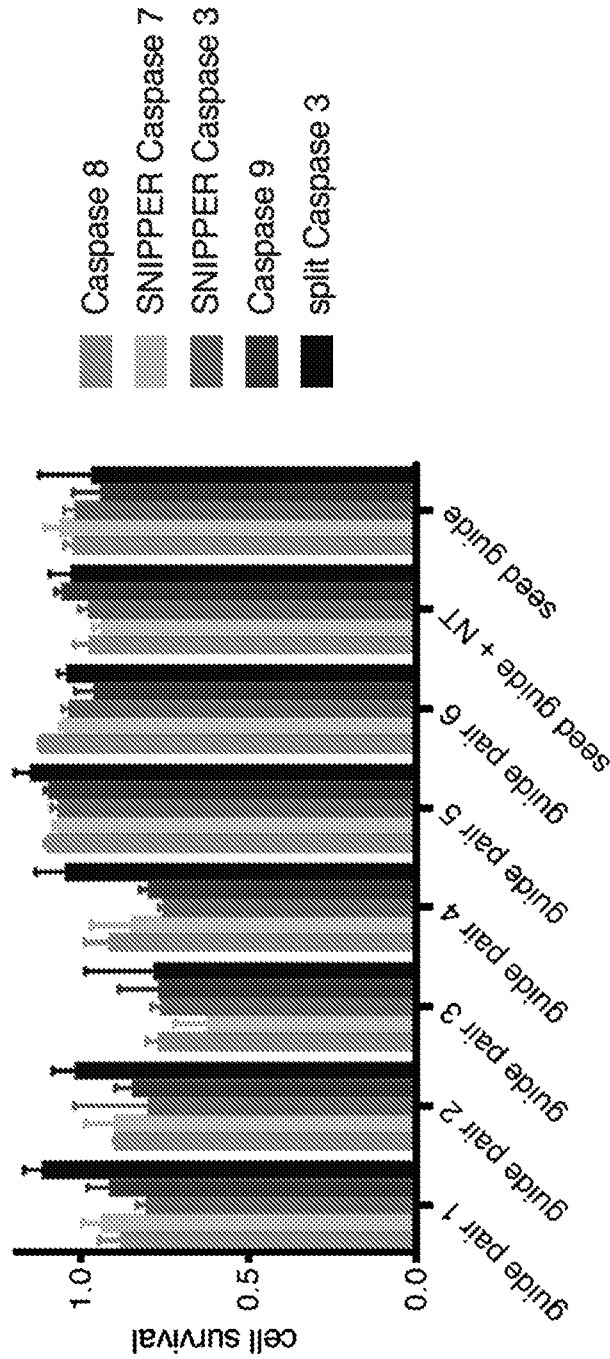


FIG. 3

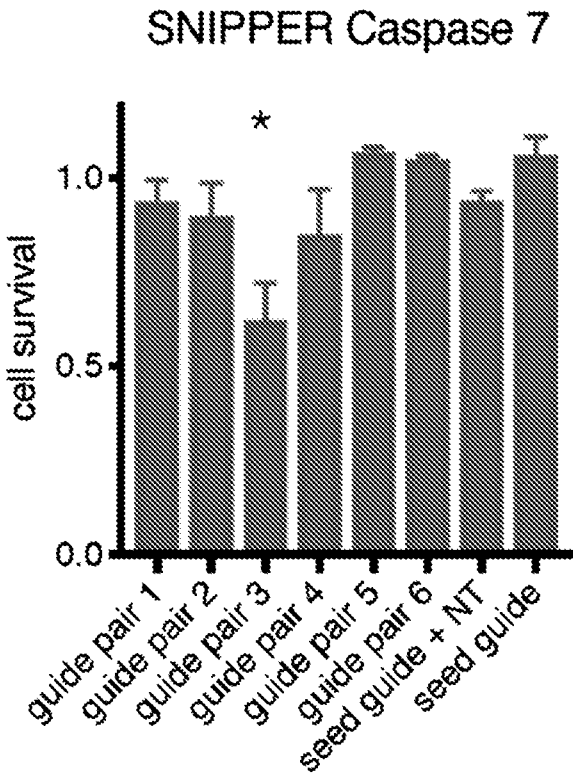


FIG. 4A

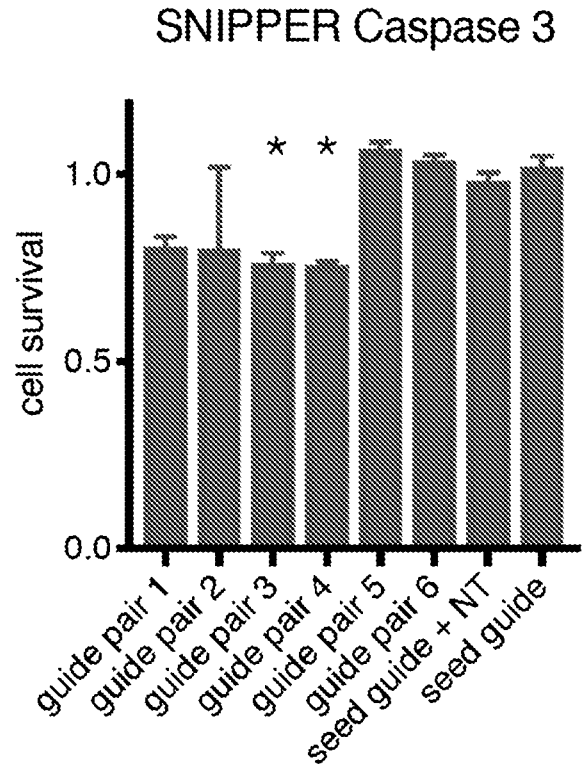


FIG. 4B

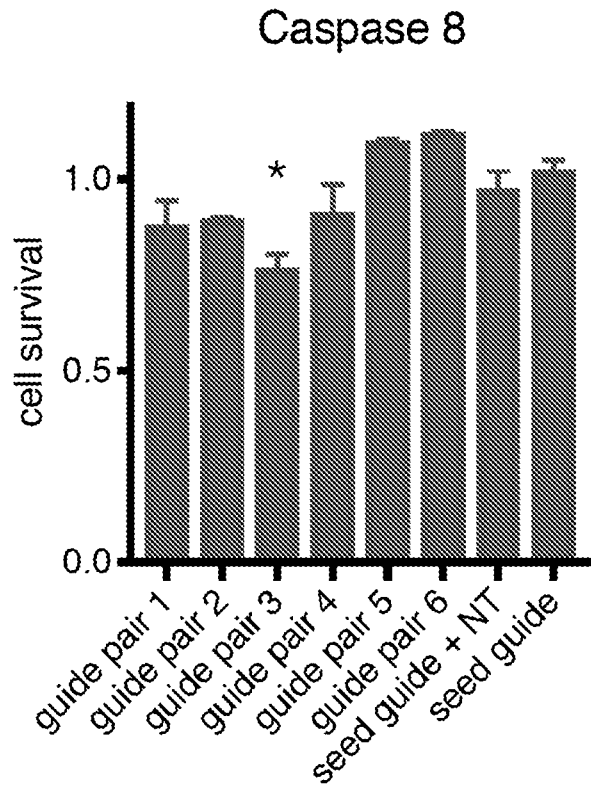


FIG. 4C

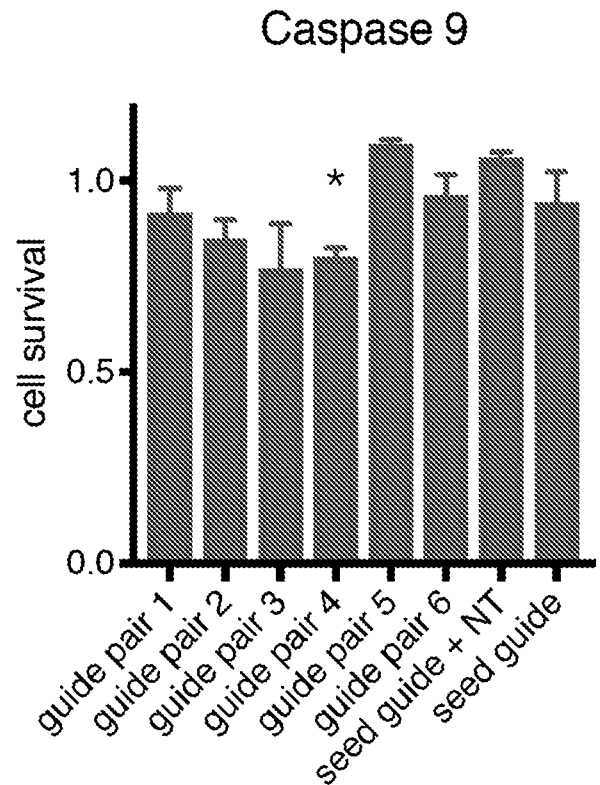


FIG. 4D

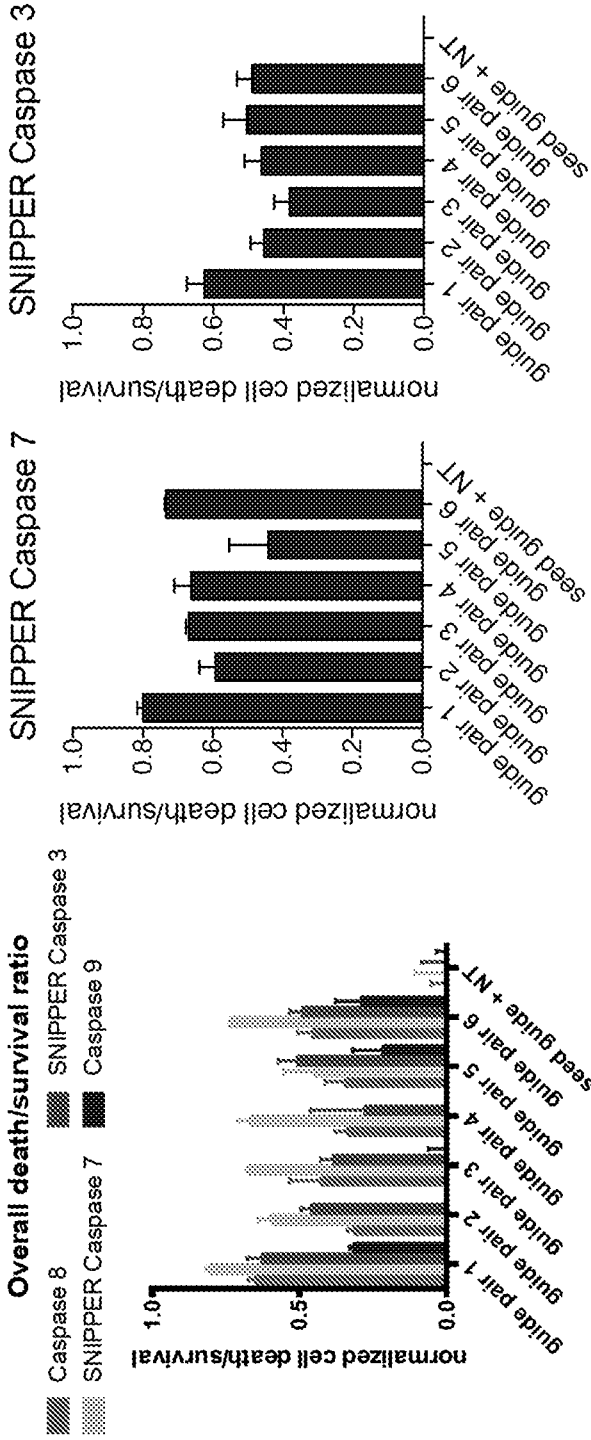


FIG. 5A

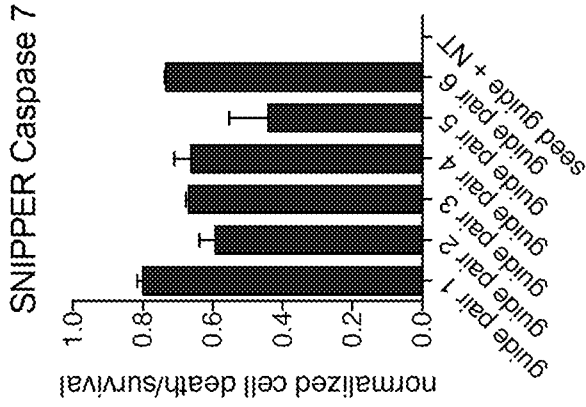


FIG. 5B

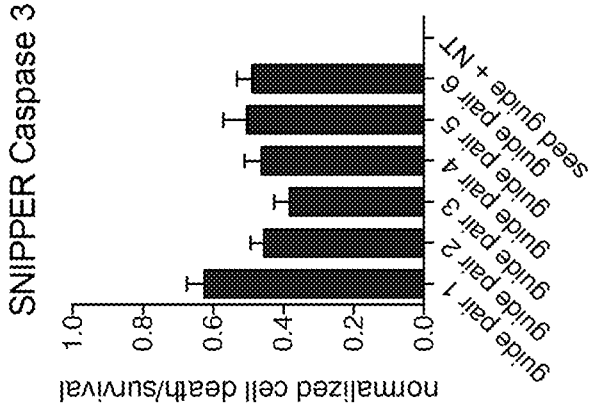


FIG. 5C

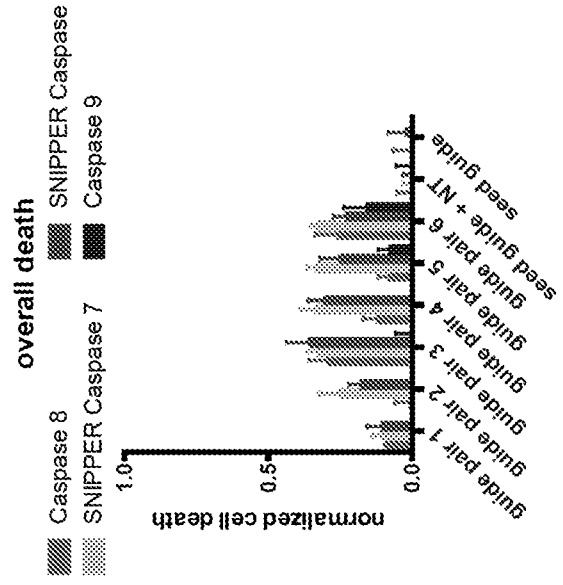


FIG. 5D



FIG. 5E

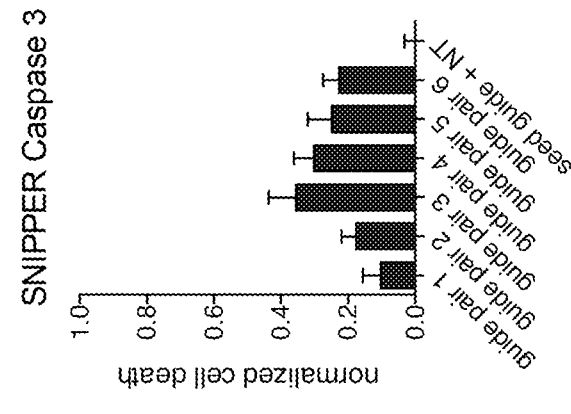


FIG. 5F

6/39

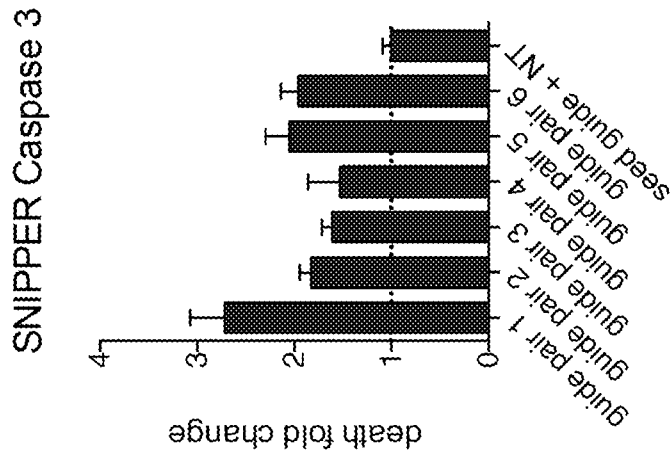


FIG. 5I

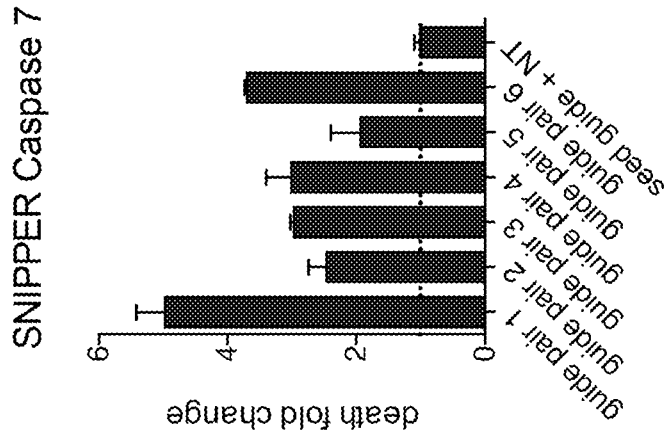


FIG. 5H

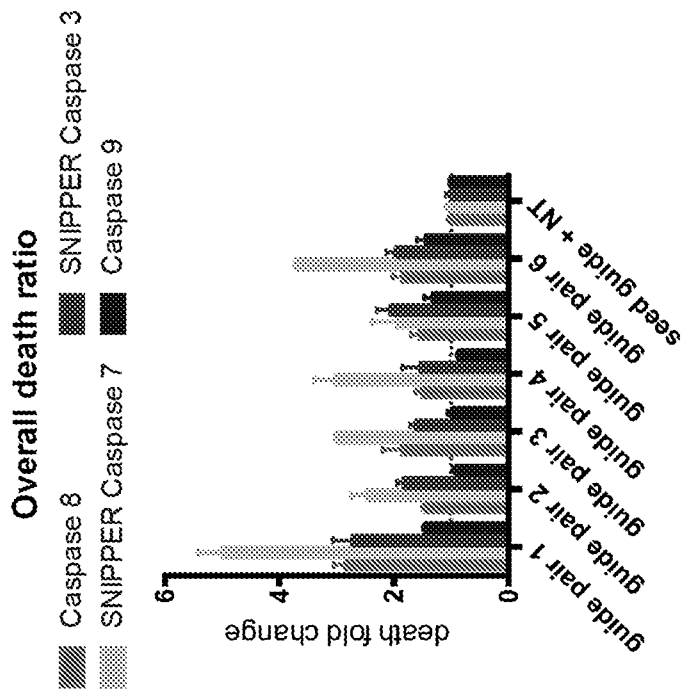


FIG. 5G

7/39

	1	10	20	30	40	50	60
15-RcSBC2C2	-----	-----	-----	-----	-----	-----	-----
16-RcRC2C2	-----	-----	-----	-----	-----	-----	-----
01-Lb(MA)C2C2	-----	-----	-----	-----	-----	-----	-----
02-Lb(NK-1)C2C2	-----	-----	-----	-----	-----	-----	-----
03-CaC2C2	-----	-----	-----	-----	-----	-----	-----
04-Lb(NK-2)C2C2	-----	-----	-----	-----	-----	-----	-----
05-Cg1C2C2	-----	-----	-----	-----	-----	-----	-----
06-Cg2C2C2	-----	-----	-----	-----	-----	-----	-----
07-PpC2C2	-----	-----	-----	-----	-----	-----	-----
08-LsC2C2	MWISIKTLIHHLGVLFFCDYMYNRREKKIIEVKI	-----	-----	-----	-----	-----	-----
09-LiwC2C2	-----	-----	-----	-----	-----	-----	-----
10-LibC2C2	-----	-----	-----	-----	-----	-----	-----
13-LspC2c2	-----	-----	-----	-----	-----	-----	-----
14-LsC2C2	-----	-----	-----	-----	-----	-----	-----
11-Lew1C2C2	-----	-----	-----	-----	-----	-----	-----
12-Lew2C2C2	-----	-----	-----	-----	-----	-----	-----
	61	70	80	90	100	110	120
15-RcSBC2C2	-----	-----	-----	-----	-----	-----	-----
16-RcRC2C2	-----	-----	-----	-----	-----	-----	-----
01-Lb(MA)C2C2	-----	-----	-----	-----	-----	-----	-----
02-Lb(NK-1)C2C2	-----	-----	-----	-----	-----	-----	-----
03-CaC2C2	-----	-----	-----	-----	-----	-----	-----
04-Lb(NK-2)C2C2	-----	-----	-----	-----	-----	-----	-----
05-Cg1C2C2	-----	-----	-----	-----	-----	-----	-----
06-Cg2C2C2	-----	-----	-----	-----	-----	-----	-----
07-PpC2C2	-----	-----	-----	-----	-----	-----	-----
08-LsC2C2	-----	-----	-----	-----	-----	-----	-----
09-LiwC2C2	-----	-----	-----	-----	-----	-----	-----
10-LibC2C2	-----	-----	-----	-----	-----	-----	-----
13-LspC2c2	LNINENNNKEKIDNNKFIGEFVNY	-----	-----	-----	-----	-----	-----
14-LsC2C2	LNINENNNKEKIDNNKFIGEFVNY	-----	-----	-----	-----	-----	-----
11-Lew1C2C2	VKFRNNKNINEIKERLKKLKE	-----	-----	-----	-----	-----	-----
12-Lew2C2C2	--VKSTSEENRTSERLSELLS	-----	-----	-----	-----	-----	-----
	121	130	140	150	160	170	180
15-RcSBC2C2	-----	-----	-----	-----	-----	-----	-----
16-RcRC2C2	-----	-----	-----	-----	-----	-----	-----
01-Lb(MA)C2C2	-----	-----	-----	-----	-----	-----	-----
02-Lb(NK-1)C2C2	NRRNGLVLSDMLSVDRAFEKAFEKYAE	-----	-----	-----	-----	-----	-----
03-CaC2C2	-----	-----	-----	-----	-----	-----	-----
04-Lb(NK-2)C2C2	-----	-----	-----	-----	-----	-----	-----
05-Cg1C2C2	-----	-----	-----	-----	-----	-----	-----
06-Cg2C2C2	-----	-----	-----	-----	-----	-----	-----
07-PpC2C2	-----	-----	-----	-----	-----	-----	-----
08-LsC2C2	-----	-----	-----	-----	-----	-----	-----
09-LiwC2C2	-----	-----	-----	-----	-----	-----	-----
10-LibC2C2	-----	-----	-----	-----	-----	-----	-----
13-LspC2c2	-----	-----	-----	-----	-----	-----	-----
14-LsC2C2	-----	-----	-----	-----	-----	-----	-----
11-Lew1C2C2	-----	-----	-----	-----	-----	-----	-----
12-Lew2C2C2	-----	-----	-----	-----	-----	-----	-----

FIG. 6A

	181	190	200	210	220	230	240
15-RcSBC2C2	DLGEAFWKL	VSEAGLAQ	----DSDYDQFKRRLHPYGD	KFPADSGAKL	KFEAD	-----	
16-RcRC2C2	DLGEAFWKL	VSEAGLAQ	----DSDYDQFKRRLHPYGD	KFPADSGAKL	KFEAD	-----	
01-Lb(MA)C2C2	QQVIKELKDK	FGGARIY	-----DDDIEEAL	TETLKKSF	RKENVRNS	IKVLI	ENAAGIRS
02-Lb(NK-1)C2C2	ISETNF	INRIKNNIDNK	----VSEDI	IDRIIAKYL	KKSLCRERV	KRGLK	KLMLMNAFDLPY
03-CaC2C2	EIVSNLGM	VESLKER	----DPQKVI	SKLLLYL	RKGLKPG	PTDGLR	MILEASCGLRKL
04-Lb(NK-2)C2C2	DTVLRN	VEKINAFDRISTE	--EREQI	IDDLLEI	QLRKGLR	KGKAGL	REVLLIGAGVIVRT
05-Cg1C2C2	KDLALFI	KRKF	-----NDDIVE	IKNLDL	ISLFY	NALQK	VPGEHFTDES
06-Cg2C2C2	AGLKRFF	KSKYHEEIKY	----TSEKN	QSVPTK	NLIPLFF	NAVDRI	QEDKFDEKNWSYFC
07-PpC2C2	EEILPFL	NHKFQEPVKYW	---KNGKEE	SFNLTL	LIVEAV	QAQDKR	-----
08-LsC2C2	LNISN	FLNHRFKKSLY	YFPENSPDK	SEEYRIE	INLSQ	LLED	SLKKOQGTFC---WESFS
09-LiwC2C2	ESLTKL	FKPKFQNMISY	----PPSKGAE	-HVQFCL	TDIAVPA	IRDLDE	---IKPDWGIF
10-LibC2C2	ESLTKL	FKPKFQNMISY	----PPSKGAE	-HVQFCL	TDIAVPA	IRDLDE	---IKPDWGIF
13-LspC2c2	EEVVL	YIEVYGKSEK	LK----ALEIT	KKKI	IIDEA	IRQGIT	KDDKKIEIKRQEN
14-LsC2C2	EEVVL	YIEAYGKSEK	LK----ALGIT	KKKI	IIDEA	IRQGIT	KDDKKIEIKRQEN
11-Lew1C2C2	NQEID	YDVKSK	-----KNQOE	IFDVL	KEILEL	KIKETE	KEEIIITFDSE
12-Lew2C2C2	KDSVL	YLKNRKEK	NAVQ----DKNY	SEEDISE	YDLKN	--KNSF	SVLKKILLNEDV

	241	250	260	270	280	290	300
15-RcSBC2C2	---PPEPQA	---FHGRWYG	AMSKRG	NDAKELAA	ALYEH	LHVDEKR	---I--DG-----
16-RcRC2C2	---PPEPQA	---FHGRWYG	AMSKRG	NDAKELAA	ALYEH	LHVDEKR	---I--DG-----
01-Lb(MA)C2C2	SLSKDEE	ELIQEYFVK	QLVVEEY	TTKLQKN	VVKS	IKNQNM	VIQPDSDS
02-Lb(NK-1)C2C2	SDPDID	VQR---DFIDY	VLEDFYH	VRAKSQ	VRS	IKNMN	MPVQPE---
03-CaC2C2	G--DEKEL	K---VFLQ	TLED	DFEKKT	FKKNL	LIRSIEN	QNMVAVQPS
04-Lb(NK-2)C2C2	D--KKQEI	A---DFLE	ILDED	FNKTNQ	AKNI	KL	SIENQGLV
05-Cg1C2C2	Q--EMMPYR	---EYKN	KFIE	-----R	KIILL	LANS	IEQNKG
06-Cg2C2C2	K--EMSPYL	---DYK	KSYLN	-----R	KEIL	LANS	IQNRGFS
07-PpC2C2	---KLQPY	---DWKT	WYIQ	-----T	KSDL	LKKS	SIENNR
08-LsC2C2	K--DMELY	I---NWA	ENYIS	-----S	KT	LKKS	SIRNNR
09-LiwC2C2	E--KLKPY	T---DWA	ESYIH	-----Y	QTTI	QKS	IEQNKI
10-LibC2C2	E--KLKPY	T---DWA	ESYIH	-----Y	QTTI	QKS	IEQNKI
13-LspC2c2	---EEEIE	I---DIR	EYTN	---KTL	NDCS	IILRI	IENDELE
14-LsC2C2	---EEEIE	I---DIR	EYTN	---KTL	NDCS	IILRI	IENDELE
11-Lew1C2C2	---KLK	---VFG	EDFVK	---KEA	KIKAI	EKSL	KINKANY
12-Lew2C2C2	---NSEE	LE---IFR	KDVEA	---KLN	KIN	SLKYS	FEEKANY

	301	310	320	330	340	350	360
15-RcSBC2C2	-----Q	PKRNP	KTDKF	APGLVV	-----A	RALGI	---ESSV
16-RcRC2C2	-----Q	PKRNP	KTDKF	APGLVV	-----A	RALGI	---ESSV
01-Lb(MA)C2C2	QSSAV	SSDTLV	NCKEK	DVLF	KAF	FLTDY	AV---L
02-Lb(NK-1)C2C2	SKGGT	ESGNK	-RSAE	KEAF	KKFL	SDYAS	---L
03-CaC2C2	TQGRF	NSQ---K	NEEK	SAIER	MMSMY	AD---L	NEDHRE
04-Lb(NK-2)C2C2	VSGAQ	KGKSS	KAQE	KEALS	AFLLD	YAD---L	DKNVRF
05-Cg1C2C2	-----E	TFSKR	KRVL	HQWA	IEVQE	-----R	GDFS
06-Cg2C2C2	-----L	SKRK	QLFQ	WAMK	FQES	PLIQ	NNFAVE
07-PpC2C2	-----L	SKR	KALL	AWET	EFTA	-----S	GIDL
08-LsC2C2	-----E	SRGQ	LMDRY	MKDIL	N---K	NKPF	DI---Q
09-LiwC2C2	-----D	SPRKL	VLQ	YVTA	FLN---G	EPLGL	-----D
10-LibC2C2	-----D	SPRKL	VLQ	YVTA	FLN---G	EPLGL	-----D
13-LspC2c2	LYKIE	IEKIE	-NETE	KVFEN	RYEEH	LREK	LLKDN
14-LsC2C2	LYKIE	IEKIE	-NETE	KVFEN	RYEEH	LREK	LLKDD
11-Lew1C2C2	VKG	-----E	NKRS	RIY	EYK	-----S	ENLKK
12-Lew2C2C2	G	-----K	SKRN	IYD	YRES	AKR---N	DYINN

FIG. 6B

9/39

	361	370	380	390	400	410	420
15-RcSBC2C2	GEEEIQTYFVVDVAASVKEVAKAAVSAQAQAFDPPR	---	QVSGRSLSPKVG	-----			
16-RcRC2C2	GEEEIQTYFVVDVAASVKEVAKAAVSAQAQAFDPPR	---	QVSGRSLSPKVG	-----			
01-Lb(MA)C2C2	NLVNLYFYGSESIRD	-----	YSYTKK	---	SVWKEHDEQKANKTLFIDEICH		
02-Lb(NK-1)C2C2	RLVVLVYFYGSDDSKL	-----	SDVNEKF	---	DVWEDHAARRVDNREFIKLPLE		
03-CaC2C2	RLNVLYFNVDTEKTEEPTLPGE	-----	VDTNPVF	---	EVWHDHEKKGKENDRQF	ATFAK	
04-Lb(NK-2)C2C2	RLINLYFYVKNDDVMSLTEIPAE	-----	VNLEKDF	---	DIWRDHEQRKEENGDF	VGCPD	
05-Cg1C2C2	NFKMKCRVQDELNDLE	-----	KSMKKGK	---	NPEKEKEAYKKQKNF	-----	
06-Cg2C2C2	NVDELCTAITEKL	-----	MNFDKDK	---	SNKTRNF	-----	
07-PpC2C2	MTDVLCKMLQDV	-----	KPLTDDK	---	GKINTNAYHRGLK	-----	
08-LsC2C2	QLEKLTSAK	-----	ATFKEAK	---	KNDK	-----	
09-LiwC2C2	KLADLAESFKV	-----	VDLNEDK	-----			
10-LibC2C2	KLADLAESFKL	-----	VDLNEDK	-----			
13-LspC2c2	GFVKFYLVNSGDKKSKEN	-----	KKMFVEK	---	ILNTNVDLTVEDIVDFIVKELK		
14-LsC2C2	GFVKFYLVNGGDKKSKN	-----	KKMLVEK	---	ILNINVDLTVEDIADFVIKELE		
11-Lew1C2C2	NIKELYSKIEEILKTHL	-----	KSIVREFYQNEIIGESEFSSKKN	GDG	SILYNQ		
12-Lew2C2C2	DIEKLFFLIENSKK	-----	HEKYKIR	---	EYYHKIIGRKN	DKENFAKIYE	

	421	430	440	450	460	470	480
15-RcSBC2C2	-----	FALAEHLERVTGSKRCSFD	---	PAAGPSVLAL	-----	HDE	
16-RcRC2C2	-----	FALAEHLERVTGSKRCSFD	---	PAAGPSVLAL	-----	HDE	
01-Lb(MA)C2C2	ITKIGKNGKEQKVL DYE	----	ENRSRCKQININYRSALN	---	YAKNNTSGIFE	---	NED
02-Lb(NK-1)C2C2	NKLANGKTDKAERIRKN	---	TVKELYRNQINIGCYRQAVK	---	AVEEDNNGRYFD	---	DKM
03-CaC2C2	ILTEDRETRKKEKLAVKEALNDLKS	AIRDH	NIMAYRCSIK	---	VTEQDKDGLFFE	---	DQR
04-Lb(NK-2)C2C2	ILLADRDRVKKSNKQVKIAERQLRES	SIREK	NIKRYRFSIK	---	TIEKDDGTYFFA	---	NKQ
05-Cg1C2C2	-----	KIKTIWKDYPYKTHIGLIE	---	KIKENEELNQF	-----	NIE	
06-Cg2C2C2	-----	EIKKLWKQHPHNKDKALIKL	FNQEGNEALNQF	-----	NIE		
07-PpC2C2	-----	KALQNHQPAIFGTREVPNE	---	ANRADNQLSIY	-----	HLE	
08-LsC2C2	-----	EINYKLLKSTLQNHQRIIE	---	ELKENSELNQF	-----	NIE	
09-LiwC2C2	-----	SANYKIKACLQHQHORNILD	---	ELKEDPELNQY	-----	GIE	
10-LibC2C2	-----	SANYKIKACLQHQHORNILD	---	ELKEDPELNQY	-----	GIE	
13-LspC2c2	FWNITKRIEKVKKFNN	----	EFLENRRNRTYIKSYVLLD	---	KHEKFKIERENK	---	KDK
14-LsC2C2	FWNITKRIEKVKKVNN	----	EFLEKRRNRTYIKSYVLLD	---	KHEKFKIERENK	---	KDK
11-Lew1C2C2	IKDSIKKEENFIEFIENTGNLEL	KELT	KSQIFKYFLENE	---	ELNDENIKFAFCYFVEIE		
12-Lew2C2C2	EIQNVNNIKELIEKIP	----	DMSELKKSQVIFYKYLDKE	---	ELNDKNIKYAFCHFVEIE		

	481	490	500	510	520	530	540
15-RcSBC2C2	VKPTYKRLCA	-----	RGKNAARAFPADKTELLALMRHTHEN	---	RVRNQVMRMGRVSEYR		
16-RcRC2C2	VKPTYKRLCA	-----	RGKNAARAFPADKTELLALMRHTHEN	---	RVRNQVMRMGRVSEYR		
01-Lb(MA)C2C2	SNHFWIHLIENEVERLYNGIENGEEFKFETGYISEKVWKAVIN	---	HLSIKYIALGKAVYN				
02-Lb(NK-1)C2C2	LNMFFIHRIEYGVKEKIYANLKQVTEFKARTGYLSEKIWKDLIN	---	YISIKYIAMGKAVYN				
03-CaC2C2	INRFWIHHIESAVERILASINPEKLYKLRIGYLGEKVWKDILLN	---	YLSIKYIAVGKAVFH				
04-Lb(NK-2)C2C2	ISVFWIHRIENEVERILGSINDKKLYRLRLGYLGEKVWKDILLN	---	FLSIKYIAVGKAVFN				
05-Cg1C2C2	IGKYFEHYFP	---	IKK	---	ERCTEDEPYLNSSETIATTVNYQLKN	---	ALISYLMQIGKYKQF
06-Cg2C2C2	LGKYFEHYFP	---	KTGKESAESYLNQPTI	IKTVGYQLRN	---	AFVQYLLQVGLHQY	
07-PpC2C2	VVKYLEHYFP	IKTSKR	---	RNTADDIAHYLKAQTLKTTIEKQLVN	---	AIRANI IQQGKTNNH	
08-LsC2C2	IRKHLETYFP	---	IKKTNRKVGDIRN	---	LEIGEIQKIVNHRLKN	---	KIVQRILQEGKLASY
09-LiwC2C2	VVKYIQRYFP	---	IKRAPNRSKHARADFLKKELIESTVEQOFKN	---	AVYHYVLEQGMKMEAY		
10-LibC2C2	VVKYIQRYFP	---	IKRAPNRSKHARADFLKKELIESTVEQOFKN	---	AVYHYVLEQGMKMEAY		
13-LspC2c2	IVKFFVENIK	---	NNSIKEKIEKILAEFKINELIKKLEKELKKGNC	DT	EIFGIFPKKHYK		
14-LsC2C2	IVKFFVENIK	---	NNSIKEKIEKILAEFKIDELIKKLEKELKKGNC	DT	EIFGIFPKKHYK		
11-Lew1C2C2	VNNLLKENVY	---	KIKRFNESNKKRIENIFEYGKLLKLLIVYKLEN	---	KLNNYVRNCGKYNH		
12-Lew2C2C2	MSQLLKNYVY	---	KRLSNISNDKIKRIFEYQNLKLLIENKLLN	---	KLDTYVRNCGKYNYY		

FIG. 6C

10/39

	541	550	560	570	580	590	600
15-RcSBC2C2	GQQAGDLAQSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMK--AWIDPMGKIVNTEKN						
16-RcRC2C2	GQQAGDLAQSHYWTSAGQTEIKESEIFVRLWVGAFALAGRSMK--AWIDPMGKIVNTEKN						
01-Lb(MA)C2C2	-YAMKELSSPG-----DIEPGKIDDSYI---NGITSFYDYEIIKAEESLQORDISMNVVFAT						
02-Lb(NK-1)C2C2	-YAMDELNASD---KKEIELGKISEEYL---SGISSFDYELIKAEEMLQRETAVYVAFAA						
03-CaC2C2	-FAMEDLGKGTG-----QDIELGKLSNSVS---GGLTSFDYEQIRADETLQRQLSVEVAFAA						
04-Lb(NK-2)C2C2	-FAMDDLQEKD-----RDIEPGKISENAV---NGLTSFDYEQIKADEMLQREVAVNVAFAA						
05-Cg1C2C2	-GLENQVLDSK-----KLQEIGIYEGFQTKFMDACVFATSSLK--NIIIEPMRSGDILGKR						
06-Cg2C2C2	---NKGVLDSQ-----TLQEIGMYEGFQTKFMDACVFASSSLR--NIIQATTNEDILTRE						
07-PpC2C2	--ELKAD'TTSN-----DLIRIKTNEAFVLNLTGTCAFAANNIR--NMVDNEQ'TNDILGKG						
08-LsC2C2	--EIESTVNSN-----SLQKIKIEEAFALKFINACLFASNNLR--NMVYPVCKKDILMIG						
09-LiwC2C2	-----ELTDPK---TKDLQDIRSGEAFSFKFINACAFASNNLK--MILNPECEKDILGKG						
10-LibC2C2	-----ELTDPK---TKDLQDIRSGEAFSFKFINACAFASNNLK--MILNPECEKDILGKG						
13-LspC2c2	NFDSKKFSNKS---DEEKELYKIIYRYLKGRIEKILVNEQKVRL--KKMEKIEIEKILNES						
14-LsC2C2	NFDSKKFSKKS---DEEKELYKIIYRYLKGRIEKILVNEQKVRL--KKMEKIEIEKILNES						
11-Lew1C2C2	-MENGDIA'TSD-----INMRNRQTEAFLRSIIGVSSFGYFSLR--NILGVNDDDFYETEE						
12-Lew2C2C2	-LQVGEIATSD-----FIARNRQNEAFLRNIIGVSSVAYFSLR--NILETENENDITGRM						

	601	610	620	630	640	650	660
15-RcSBC2C2	DRDLTAAVNIRQVIS-----NKEMVAEAMARR-----GIYFG						
16-RcRC2C2	DRDLTAAVNIRQVIS-----NKEMVAEAMARR-----GIYFG						
01-Lb(MA)C2C2	NYLACATVDTDK-----DFLLFSKEDIRSCTKKD--GNLCKNIMQFWG						
02-Lb(NK-1)C2C2	RHLSSQTVELDSENSDFLLLPKPGTMDKNDKNKLASNNILNFKDK--ETLRD'TILQYFG						
03-CaC2C2	NNLFRAVVGQ'TGKKIE---QSKSEENEE'DFLWKAEKIAESIKKEGEGNTLKSILQFFG						
04-Lb(NK-2)C2C2	NNLARV'TVDIPQNGE-----KEDILLWNKSDIKKYKKNK--KKGILKSILQFFG						
05-Cg1C2C2	E-----FKEAIATSSFVN-----YHHFFPYFP						
06-Cg2C2C2	K-----FKKELEKNVELK-----HDLFFK						
07-PpC2C2	D-----FIKSLKDN'TNS-----QLYSFFF						
08-LsC2C2	E-----FKNSFKEIKHKK---FIR-QWSQFFS						
09-LiwC2C2	D-----FKKNLPNSTTQS---DVVKKMI'PFFS						
10-LibC2C2	N-----FKKNLPNSTTTRS---DVVKKMI'PFFS						
13-LspC2c2	ILSEKILKRVKQY'TLEHIMYLGKLRHNDIVKMTVNTDDFSRLHAKE---ELDLELITFFA						
14-LsC2C2	ILSEKILKRVKQY'TLEHIMYLGKLRHNDIDMTTVNTDDFSRLHAKE---ELDLELITFFA						
11-Lew1C2C2	DLTKKERNLEKAKE-----DITIKNTFDEVVVKSFQKKGIYNIKENLKM'FYG						
12-Lew2C2C2	RGKTVKNNKGEEKY-----VSGEVDKIY'NENKQNEVKENLKM'FYS						

	661	670	680	690	700	710	720
15-RcSBC2C2	ETPELDRL-----GAEGNEGFV-FALLRYLRGCRNQT'FH-L---GARAGFLKE						
16-RcRC2C2	ETPELDRL-----GAEGNEGFV-FALLRYLRGCRNQT'FH-L---GARAGFLKE						
01-Lb(MA)C2C2	GYS'TWKNFCEE-----YLKDDKDALELL-YSLKSMLYSMRNS'SFH-F-----STENV						
02-Lb(NK-1)C2C2	GHSLW'TDFPFDK----YLAGGKDDVDFL-TDLKDVIIYSMRNDS'FH-Y-----ATENH						
03-CaC2C2	GASSWDLNH'FCAAYGNESALGYETKFA-DDLRKAIYSLRNET'FH-F-----TTLNK						
04-Lb(NK-2)C2C2	GASTWNM'KMFEIAY--HDQPGDYEENYL-YDIIQIIYS'LNRKSFH-F-----KTYDH						
05-Cg1C2C2	FELKGMKDR'ESLIPFGEQTEAKQMQNI-WALRGSVQQIRNEIFHS'FDKNQKFNLPQLDK						
06-Cg2C2C2	TEIVEERDEN'PA----KKIAMTPNELDL-WAIRGAVQVRV'RNQIFH-QQINKRHEPNQLKV						
07-PpC2C2	GEG'LSTNK-----AEKETQL-WGIRGAVQQIR'NNVNH-Y---KKDALKTVFNI						
08-LsC2C2	QEITVDDI-----ELAS-WGLRGAIAPIRNEI'IH-L---KKHSWK'KFFNN						
09-LiwC2C2	DEIQNVNF-----DEAI-WAIRGSIQQIRNEV'YH-C---KKHSWKSILKI						
10-LibC2C2	DELQNVNF-----DEAI-WAIRGSIQQIRNEV'YH-C---KKHSWKSILKI						
13-LspC2c2	STNMELNKI'FN-----GKEKVTDFFGFNLNGQKITL'KEKVPK-FKLNILKKNLFINN						
14-LsC2C2	STNMELNKI'FSR----ENINNDENIDFF-GGDREK'NYVLDKKILN-SKI'KIRDLDFIDN						
11-Lew1C2C2	DSFDNADK-----DELKQFF-VNMLNAITSIR'HRVVH-Y----NMNTN'SENI						
12-Lew2C2C2	YDFNMDNK-----NEIEDFF-ANIDEA'ISSIRHGIVH-F-----NLELE'GK						

FIG. 6D

11/39

	721	730	740	750	760	770	780
15-RcSBC2C2	IRK-----	ELEKTRWGKAKEAEHVVL	TDKTVAAIRAIIDNDAKAL	---	GARLLA-DLSGA		
16-RcRC2C2	IRK-----	ELEKTRWGKAKEAEHVVL	TDKTVAAIRAIIDNDAKAL	---	GARLLA-DLSGA		
01-Lb(MA)C2C2	DNG-----		SWDTE	LIGKLFEE	CNRA--	ARIEKE-KFYNN	
02-Lb(NK-1)C2C2	NNG-----		KWNKELISAMFEHETERM	---	TVVMKD-KFYNS		
03-CaC2C2	GSF-----		DWNAKLIGDMFSHEAATG	---	IAVERT-RFYSN		
04-Lb(NK-2)C2C2	GDK-----		NWNRELIGKMIHDAERV	---	ISVERE-KFHSN		
05-Cg1C2C2	SNF-----	EFDASENSTGKSQ-----		SYIETDYKFLFEAEKNQL	---	EQFFIE-RIKSS	
06-Cg2C2C2	GSF-----		ENGDLGNVSYQKTIYQKLFDAEIKDI	---	EIYFAE-KIKSS		
07-PpC2C2	SNF-----	ENPTITDPKQQT-----		NYADTIYKARFINELEKI	---	PEAFAQ-QLKTG	
08-LsC2C2	PTF-----	KVKKSKIINGKTKDVTSEFLYKETF	KDYFYSELDSV	---	PELIIN-KMESS		
09-LiwC2C2	KGF-----	EFEPNNM-----		KYTDSMQKLMKDIAKI	---	PDFIEE-KLKSS	
10-LibC2C2	KGF-----	EFEPNNM-----		KYADSDMQKLMKDIAKI	---	PEFIEE-KLKSS	
13-LspC2c2	ENNIDEKLSHFYSFQKEGYLLRNKILHNSY	IGNIQETKNL	KGEYENV	---	EKLKELKVSDE		
14-LsC2C2	KNNITN--	NFIRKFTKIGTNERNRILHAISKERDLQGT	---	QDDYNKV	---	INLIQNLKISDE	
11-Lew1C2C2	FNF-----		SGIEVSKLLKSIFEKETDKRELKLIKIFR	---	QLNSA		
12-Lew2C2C2	DIF-----	AFKN-----		IAPSEISKMMFQNEINEKKLKLIKIFK	---	QLNSA	

	781	790	800	810	820	830	840
15-RcSBC2C2	FVAHYASKEHFSTLYSEIVKAVKDA-----		PEVSSGLPRLKLLLKRADGVRGVVHGLRD				
16-RcRC2C2	FVAHYASKEHFSTLYSEIVKAVKDA-----		PEVSSGLPRLKLLLKRADGVRGVVHGLRD				
01-Lb(MA)C2C2	NLHMFYSSSLEKVLERLYSSH-----		HERASQVPSFNRFVFR	---	KNFPSSLSE		
02-Lb(NK-1)C2C2	NLPMFYKNDLKKLLIDLYKDN-----		VERASQVPSFNKVFVR	---	KNFPALVRD		
03-CaC2C2	NLPMFYRESDLKRIMDHLYNTY-----		HPRASQVPSFNRFVFR	---	KNFRLFLSN		
04-Lb(NK-2)C2C2	NLPMFYKDADLKKILDLLYSY-----		AGRASQVPAFNTVLVR	---	KNFPEFLRK		
05-Cg1C2C2	GALEYYPKLSLEKLF	FAKKEMKFSLG-----		SQVVAFAPS	YKLVK-----	KGHSYQAT	
06-Cg2C2C2	GALEQYSMKDLEKLF	SNKELTSLG-----		GQVVAFAPS	YKLYK-----	QGYFYQNEK	
07-PpC2C2	GAVSYTYIENLKSLLTTFQFSLC-----		RSTIPFAPGFKKVFN	---	GGINYQNAK		
08-LsC2C2	KILDYSSDQLNQVFTIPNFELSLL-----		TSAPVFAPSFKRVYL	---	KGFDYQNOQ		
09-LiwC2C2	GIIRFYSHDKLQSIWE-MKQGFSL-----		TTNAPFVPSFKRVYA	---	KGHDYQTSK		
10-LibC2C2	GVVRFYRHDELQSIWE-MKQGFSL-----		TTNAPFVPSFKRVYA	---	KGHDYQTSK		
13-LspC2c2	EISKSLSDVIFEGKVDIINKINDIKISEENNNDIKYLP	SFKVLP-----		EILNLYRNN			
14-LsC2C2	EVSKALNLDVVFKDKKNIITKINDIKISEENNNDIKYLP	SFKVLP-----		EILNLYRNN			
11-Lew1C2C2	GVPDYWENRKIDKYLENIEFKFV-----		NKNIPFVPSFTKLYN	---	RIDNLRKGN		
12-Lew2C2C2	NVFNYEKDVI IKYLKNTKFNFV-----		NKNIPFVPSFTKLYN	---	KIEDLRNT-		

	841	850	860	870	880	890	900
15-RcSBC2C2	TRKHAFATKLPPPPAPRELD	DDPATKARYIALLRLYDGP	-----				FRAYASGI
16-RcRC2C2	TRKHAFATKLPPPPAPRELD	DDPATKARYIALLRLYDGP	-----				FRAYASGI
01-Lb(MA)C2C2	Q-----	RITPKFTDSKDEQI	WQSAVYYLCKEIIYND	-----			FLQSK--E
02-Lb(NK-1)C2C2	KDNLGIELDLKADADKGENELKFYNALYYMFKEIIYNA	-----					FLNDK--N
03-CaC2C2	T-----	LNTNTSFDTE-VYQ	KWESGVYYLFKEIIYNS	-----			FLPSG--D
04-Lb(NK-2)C2C2	D-----	MGYKVHFNNPEVENQ	WHSVYYLYKEIIYNL	-----			FLRDK--E
05-Cg1C2C2	EGTANY-LGLSYNRYE-LKEES	FQAQYLLKLIYQYV	-----				FLPNFSQG
06-Cg2C2C2	T-----	IELEQFTDYD-FSND	VFKANYLLIKLIYHYV	-----			FLPQFSQA
07-PpC2C2	QDESFYELMLEQYLRKENFAEESYNARYFMLKLIYNNL	-----					FLPGFT-T
08-LsC2C2	EAQPDYNLKLNIYNEKA-FNSEAFQAQYSLFKM	VYYQV-----					FLPQFT-T
09-LiwC2C2	NRY--	YDLGLTTFDILE-YGEED	FRARYFLT	TKLVYYQO-----			FMPWFT-A
10-LibC2C2	NRY--	YNLDTTTFDILE-YGEED	FRARYFLT	TKLVYYQO-----			FMPWFT-A
13-LspC2c2	K-----	LFDIE-SEKIILNAV	KYVNKILYEKI	---	TSNEENEFLKTL	PKLV-K	
14-LsC2C2	P-----	KNEPFDTIE-TEKIVL	NALIYVNKELYKLLILEDDLEENESKNIFLQELK	---	K		
11-Lew1C2C2	A-----	LNLGYINIPK--RKEARDS	QIYLLKNIYYGE	-----			FVEKFV-N
12-Lew2C2C2	-----	LKFFWSVP-KDKEEKDA	QIYLLKNIYYGE	-----			FLNKFV-K

FIG. 6E

12/39

	901	910	920	930	940	950	960
15-RcSBC2C2	TGTALAGPAARA	-----	KEAATALAQS	VNVTKA	-----	YSDVMEGR	TSRRLR
16-RcRC2C2	TGTALAGPAARA	-----	KEAATALAQS	VNVTKA	-----	YSDVMEGR	SRRRLR
01-Lb(MA)C2C2	AYKLFREGVKNL	-----	DKNDINNQK	-----	AADSFKQAV	VVYYG	
02-Lb(NK-1)C2C2	VRERFITKATKVADNYDRNKERNLKDRIKSAGSDEK	KKLREQLQNYIA	ENDFGQRI	KNIV			
03-CaC2C2	AHHLFFEGLRRI	-----	RKEADNLP	IVGKEAKKR	-----	AVQDFGRR	CDELK
04-Lb(NK-2)C2C2	VKNLFYTSLKNI	-----	RSEVSDKKQK	L	-----	ASDDFASR	CEEIE
05-Cg1C2C2	NSPAFRETVKAI	-----	LRINKDEAR	KMKKN	-----	KKFLRKYA	FEQVR
06-Cg2C2C2	NNKLFKDTVHYV	-----	IQQNKELN	TTEKDKN	-----	NKKIRKYA	FEQVK
07-PpC2C2	DRKAFADSVGFV	-----	QMKNKQAE	KVN	-----	PRKKEAYA	FEAVR
08-LsC2C2	NNDLFFKSSVDFI	-----	LTLNK	-----	-----	ERKGYAKA	FQDIR
09-LiwC2C2	DNNAFRDAANFV	-----	LRLNK	-----	-----	NRQQDAKA	FINIR
10-LibC2C2	DNNAFRDAANFV	-----	LRLNK	-----	-----	NRQQDAKA	FINIR
13-LspC2c2	KSNKKNENKNLL	---	SIEEYKNAQ	VSSSKGD	---	KKAIKKYQ	NKVTNAYLEYLENTFTEIID
14-LsC2C2	TLGNIDEIDENI	---	IENYKNAQ	ISASKGN	---	NKAIKKYQ	KKVIECYIGYLRKNYEELFD
11-Lew1C2C2	NNDNFEEKIFREI	-----	IEINKKDG	T	-----	NTKTKFYK	LEKFE
12-Lew2C2C2	NSKVFFKITNEV	-----	IKINKQR	-----	-----	NQKTGHYK	YQKFE

	961	970	980	990	1000	1010	1020
15-RcSBC2C2	PPNDGETLREYLSALTGETATEFRVQIGYESDSENARK	---	QAEFIENYRR	DMLAFMFEDY			
16-RcRC2C2	PPNDGETLREYLSALTGETATEFRVQIGYESDSENARK	---	QAEFIENYRR	DMLAFMFEDY			
01-Lb(MA)C2C2	KAIGNATLSQVCQAIMTEYNRQNNNDGLKKSAYAEKQNSN	KYKHYPLFLKQV	LQSAFWEY				
02-Lb(NK-1)C2C2	QVNPDYTLAQICQLIMTEYNQONNGCMQKKSAA	RDKINKDSYQHYK	MLLLVNL	RKAFLF			
03-CaC2C2	----	NLSLSAICQMIMTEYNEQONNGNRKVK	STREDKRKPDIFQHYK	MLLLRTLQEA	FATY		
04-Lb(NK-2)C2C2	----	DRSLPEICQIMTEYNAQNFGRKVK	SQRVIEKNKDIRHYK	MLLIKTLAGAF	SLY		
05-Cg1C2C2	EMEFKETPDQYMSYLQSEMREEKVRKA	EKNKDGFEKN	---	ITMNF	EKLLMQIFVKG	FDVF	
06-Cg2C2C2	LMK-NESPEKYMQLQREMQEERTIKEAKKTNEEK	---	PNYNF	EKLLIQIFIKG	FDTF		
07-PpC2C2	PMTAADSADYMAVQSELMQEQNKKEEKVAEE	-----	TRINF	EKFVQLQVFIKGF	DSF		
08-LsC2C2	KMNKDEKPSYMSYIQSQLMLYQKQEEKE	-----	KINH	EKFINQVFIKGF	NSF		
09-LiwC2C2	EVEEGEMPRDYMGYVQGOIAIHEDSTED	-----	TPNH	EKFISQVFIKGF	DSH		
10-LibC2C2	EVEEGEMPRDYMGYVQGOIAIHEDSIED	-----	TPNH	EKFISQVFIKGF	DRH		
13-LspC2c2	FSKFNLNYDEIKTKIEERKDNKSKII	IDSISTNINITN	---	DIEYI	ISIFALLNSNTY	INK	
14-LsC2C2	FSDFKMNIQEIKKQIKDINDNKTYERITVK	TSDKTIVINDDFEYI	ISIFALLNS	NAVIN	K		
11-Lew1C2C2	TLK-ANAPIEYLEKLQSLHQINYNREKVEE	-----	DKDI	VDFVQKIFL	KGFINY		
12-Lew2C2C2	NIE-KTVPVEYLAIQSREMINNQDKE	-----	EKN	TYIDFIQ	IFLKG	IDY	

	1021	1030	1040	1050	1060	1070	1080
15-RcSBC2C2	I--RAKGF	DWILKIEPG	-----	ATAMTRA	-----	-----	-----
16-RcRC2C2	I--RAKGF	DWILKIEPG	-----	ATAMTRA	-----	-----	-----
01-Lb(MA)C2C2	L	DENKEIYGFISAQIHK	-----	SNVEIKA	-----	-----	EDFIAN
02-Lb(NK-1)C2C2	I--K-ENYAF	VLK--PY	-----	KHDLCDK	-----	-----	ADFVPD
03-CaC2C2	I--RREEF	KFIFDLPKT	-----	LYVMKPV	-----	-----	EEFLPN
04-Lb(NK-2)C2C2	L--KQERFA	FIGKATPI	-----	PYETTDV	-----	-----	KNFLPE
05-Cg1C2C2	L--TTFAG	KELLSSEEK	-----	VIKETEI	-----	-----	SLSKKINEREKT
06-Cg2C2C2	L----	RNFDLNLPAAE	-----	LVGTVKE	-----	-----	KAEGLRKRKER
07-PpC2C2	L--RAKE	FDVQMPQPQ	-----	LTATASN	-----	-----	QQKADKLNQLEAS
08-LsC2C2	I--EKNR	LTYICHPTKN	-----	TVPENDN	-----	-----	-----
09-LiwC2C2	M--RSADL	KFIKNPRNQ	-----	GLEQSEI	-----	-----	EE
10-LibC2C2	M--RSANL	KFIKNPRNQ	-----	GLEQSEI	-----	-----	EE
13-LspC2c2	I--RNRFF	FATSVWLEKQNGTKEYDYENI	ISILDEVLLINLL	RENNTDILDL	KNAITDAK		
14-LsC2C2	I--RNRFF	FATSVWLNTS	-----	EYQNI	IDILDEIMQLN	TLRNECITENWN	NLNLEEFIQK
11-Lew1C2C2	L--QGS	DLLKSLNLLNL	-----	KKDEAIA	-----	-----	NKKSFYDEKLKL
12-Lew2C2C2	L--NKNNL	KYIESNNNN	-----	DNNDIF	SKIKIKKDNKEKYDKIL	KNYEKHNRNKEIPH	

FIG. 6F

13/39

	1081	1090	1100	1110	1120	1130	1140
15-RcSBC2C2	-----PVLPE-----			PIDTRGQYEHWQ--	AALYLVM--HFVPASDVS	SN--LLH	
16-RcRC2C2	-----PVLPE-----			PIDTRGQYEHWQ--	AALYLVM--HFVPASDVS	SN--LLH	
01-Lb(MA)C2C2	YSSQQ--YKKLVD-----			KVKKTPELQKW----	YTLG--RLINPRQANQ--	FLG	
02-Lb(NK-1)C2C2	FAKYVKPYAGLIS-----			RVAGSSELOKW----	YIVS--RFLSPAQANH--	MLG	
03-CaC2C2	WKSGM--FDSLVE-----			RVKQSPDLQRW----	YVLC--KFLNGRLLNQ--	LSG	
04-Lb(NK-2)C2C2	WKSGM--YASFVE-----			EIKNNLDLQEW----	YIVG--RFLNGRMLNQ--	LAG	
05-Cg1C2C2	LKASIQVEHQLV-----			ATNSAISYW-----	LFC--KLLDSRHLNE--	LRN	
06-Cg2C2C2	IAKILNVDEQIK-----			TGDEEIAFW-----	IFA--KLLDARHLSE--	LRN	
07-PpC2C2	ITADCKLTPQYA-----			KADDATHIAFY-----	VFC--KLLDAAHLSN--	LRN	
08-LsC2C2	-----IEIPFH-----			TDMDDSNIAFW-----	LMC--KLLDAKQLSE--	LRN	
09-LiwC2C2	MSFDIKVEPSFLK-----			NKDDYIAFW-----	TFC--KMLDARHLSE--	LRN	
10-LibC2C2	MSFDIKVEPSFLK-----			NKDDYIAFW-----	IFC--KMLDARHLSE--	LRN	
13-LspC2c2	IVENDETYIKNYI-----			FESNEEKLKRR-----	LFCEELVDKEDIRKIFED		
14-LsC2C2	MKEIEKDFDDFKI-----			QTKKEIFMNYIEDIKNNILT--	EFKDDINGCDVLEK		
11-Lew1C2C2	WQNGSNLSKMPEEIIYDYIKKIKINKINYSDRMSIF-----				YLLL--KLIDHKELTN--	LRG	
12-Lew2C2C2	EINEFVREIKLGK-----			ILKYTENLNMF-----	YLIL--KLLNHKELTN--	LKG	

	1141	1150	1160	1170	1180	1190	1200
15-RcSBC2C2	QLRKWEA--LQKYEVLVDGDA-----				TDQADARREALDLVKRFRDVLVL		
16-RcRC2C2	QLRKWEA--LQKYEVLVDGDA-----				TDQADARREALDLVKRFRDVLVL		
01-Lb(MA)C2C2	SIRNYVQFVKDIQRRAKENGNI-----			RNYEVLSDSI IKILEMCTKLN	GTTSNDIHD		
02-Lb(NK-1)C2C2	FLHSYKQYVWDIYRRASETGTEINHSIAEDKIAGVDITD			DAVIDLSVKLCGTISSEISD			
03-CaC2C2	VIRSYIQFAGDIQRRAKANHNL-----			YMDNTQRVEYYSNVLEVVD	FCIKGTSRFSNVFSD		
04-Lb(NK-2)C2C2	SLRSYIQYAEDIERRAAENRNKL-----			FSKPDEKIEACKAVRVL	DLCKISTRISAEFTD		
05-Cg1C2C2	EMIKFKQ--SRIKFNHTQHAELI-----				QNLPIVELTILSNDYDEK	NDSSQ	
06-Cg2C2C2	EMIKFKQ--SSVKKGLIKNGDLI-----				EQMQPILELCILSNDESME--	KE	
07-PpC2C2	ELIKFRESVNEFKF-----				HLLLEIIEICLLSADVVP	TDYRD	
08-LsC2C2	EMIKFSCSLQSTEEISTF-----				TKAREVIGLALLNGEKGC	NDWKE	
09-LiwC2C2	EMIKYDGH-----				TGEQEIIGLALLGVDSRE	NDWKQ	
10-LibC2C2	EMIKYDGH-----				TGEQEIIGLALLGVDSRE	NDWKQ	
13-LspC2c2	ENFKFKSFIKKNEIGNFKINFGI-----			LSNLECNSEVEAKKIIGK	NSKKLESFIQN		
14-LsC2C2	KLEKIVIFDDETKFEIDKKSNI-----				QDEQR--KLSNINKKDLK	KVVDQ	
11-Lew1C2C2	NLEKYVSMNKNKIY-----				SEELNIVNLVSLDNNKVR	ANFNL	
12-Lew2C2C2	SLEKYQSANKEETF-----				SDELELINLLNLDNNRV	TEDFEL	

	1201	1210	1220	1230	1240	1250	1260
15-RcSBC2C2	FLKTGEA-----		RFEGRAAPFDLKPFRALFANPAT	FDRLFM--ATPTTARPAED			
16-RcRC2C2	FLKTGEA-----		RFEGRAAPFDLKPFRALFANPAT	FDRLFM--ATPTTARPAED			
01-Lb(MA)C2C2	YFRDEDE-----		YAEYISQFVNFQDV-----		HSGAALNAFCNSE--SEGKKN		
02-Lb(NK-1)C2C2	YFKDDEV-----		YAEYISSYLDFFEYDG-----		GNYKDSLNRFCNSDAVNDQKV		
03-CaC2C2	YFRDEDA-----		YADYLDNYLQFKDEKI--	AEVSSFAALKTFCNEE---	EVKA		
04-Lb(NK-2)C2C2	YFDSEDD-----		YADYLEKYLKYQDDAIKELSGSS	YAALDHFCNKD---	DLKF		
05-Cg1C2C2	NVDVSA-----		YFEDKSLYET-----		APYVQTD-----		
06-Cg2C2C2	SFDKIEV-----		FLEKVELAKN-----		EPYMQED-----		
07-PpC2C2	LYSSEAD-----		CLARLRPFIEQGADITNWS-----		DLFVQSD-----		
08-LsC2C2	LFDDKEA-----		WKKNMSLYVSEELLQS-----		LPYTQED-----		
09-LiwC2C2	FFSSERE-----		YEKIMKGYVGEELYQR-----		EPYRQSD-----		
10-LibC2C2	FFSSERE-----		YEKIMKGYVVEELYQR-----		EPYRQSD-----		
13-LspC2c2	IIDEYKSNIRTLF-----		SSEFLEKYKKEIDNLVEDTESEN	KNKFEKIYYPK---	EHKN		
14-LsC2C2	YIKDKDQEIKSKILCRIIFNSDFL	KKYKKEIDNLIEDMESE	ENENKFOEIIYYPK---	ERKN			
11-Lew1C2C2	-----		KPEDIGKFLKTETSIRNINQL	MNFS--EIFAD-----			
12-Lew2C2C2	EANEIGK-----		FLDFNENKIKDRKEL-----		KKFDTNK-----		

FIG. 6G

14/39

	1261	1270	1280	1290	1300	1310	1320
15-RcSBC2C2							
16-RcRC2C2	DPEGDGASEPELRVARTLRGLRQIAR--YNHMAVLSDLFAKH-----KVRDEE						
01-Lb(MA)C2C2	GIYYD-GIN-----PIVNRNWVLCCKL-YGSPDLISKIIS-----RVNENM						
02-Lb(NK-1)C2C2	ALYYD-GEH-----PKLNRNIILSKL-YGERRFLEKITD-----RVSRS						
03-CaC2C2	GIYMD-GEN-----PVMQRNIVMAKL-FGPDEVLKNVVP-----KVTR						
04-Lb(NK-2)C2C2	DIYVNAGQK-----PILQRNIVMAKL-FGPDNILSEVME-----KVTESA						
05-Cg1C2C2	-----DRT-----RVSFRPILKLEK-YHTKSLIEALLKDN-----PQFRVAATD						
06-Cg2C2C2	-----KLT-----PVKFRFMKQLEK-YQTRNFIEENLVIE-----PEFKVSEKI						
07-PpC2C2	-----KHS-----PVIHANIELSVMK-YGTPTKLEQIINKD-----TQFKTTEAN						
08-LsC2C2	-----GQT-----PVINRSIDLKVK-YGTETVILEKLFSS-----DDYKVS						
09-LiwC2C2	-----GKT-----PILFRGVEQARK-YGTETVIQRLFDAS-----PEFKVSKCN						
10-LibC2C2	-----GKT-----PILFRGVEQARK-YGTETVIQRLFDAN-----PEFKVSKCN						
13-LspC2c2	ELYIY-KKN-----LFLNIGNPNPFDKIYGLISKD IKNVDTKILFD---DDIKKNKISEID						
14-LsC2C2	ELYIY-KKN-----LFLNIGNPNPFDKIYGLISNDIKMADAKFLFNIDGKNIRKNKISEID						
11-Lew1C2C2	-----GEN-----VIKHSRFSYNIKK-YGILDLEKIVDKA-----DLKITKEE						
12-Lew2C2C2	-IYFD-GEN-----IIKHRAFYNIKK-YGMLNLLLEKIADKA-----KYKISLKE						

	1321	1330	1340	1350	1360	1370	1380
15-RcSBC2C2							
16-RcRC2C2	--VARLAEIEDETQEQSIVAAQELRTDLHDK-----VMKCHPKTISPEER						
01-Lb(MA)C2C2	--VARLAEIEDETQEQSIVAAQELRTDLHDK-----VMKCHPKTISPEER						
02-Lb(NK-1)C2C2	--IHDFHKQEDLIRE---YQIKGICSNKKEQQ-----						
03-CaC2C2	--IVEYYKLLKETSQ---YQTKGIFDSEDEQK-----						
04-Lb(NK-2)C2C2	--IEEYYQLEKQIAP---YRQNGYCKSEEDQK-----						
05-Cg1C2C2	--IREYYDYLLKQVSG---YRVRGKCSSTEKEQE-----						
06-Cg2C2C2	--IQEWMHKREEIGEL---VEKRKNLHTEWAEG-----QOTLGAEKR						
07-PpC2C2	--VLNWHEEKEKIALD---VDKRTKLHEEWASKAREIEEYNEKIKKNKSKKLDKPAEFAKF						
08-LsC2C2	--FTAWNTAQKSI EQ---IKQREDHHEQWVKA-----KNADDKERQERKREKSNFAQK						
09-LiwC2C2	--IAKLHEY---DVTEK---IAQQESLHKQWIEK-----PGLARSAWTKK						
10-LibC2C2	--ITEWERQKETI EET---IERRKELHNEWEKN-----PKKPQNNAFFKE						
13-LspC2c2	--LAEWERQKETI EET---IKRRKELHNEWAKN-----PKKPQNNAFFKE						
14-LsC2C2	AILKNLNDKLNYSND---YKAKYVNKLKENDD-----FFAKNIQENYSSF						
11-Lew1C2C2	AILKNLNDKLNYSKE---YKEYIKKLEKENDD-----FFAKNIQNKYKS-						
12-Lew2C2C2	--IKKYENLQNELKRNDFYKIQERLHRNYNQK-----PFLIKNNEKDFND-						
	--LKEYSNKKNEIEKN---YTMQQNLHRKYARP-----KKDEKFNDEDYKE-						

	1381	1390	1400	1410	1420	1430	1440
15-RcSBC2C2							
16-RcRC2C2	-----QSYAAAIAKTIEEHRFLVGRVYLGDLRLHRLMMDVIGRLIDYAGAYERDTGTFL						
01-Lb(MA)C2C2	-----QSYAAAIAKTIEEHRFLVGRVYLGDLRLHRLMMDVIGRLIDYAGAYERDTGTFL						
02-Lb(NK-1)C2C2	-----DLRTFQVLKNRVELRDIVEYSEI INELYGQLIKWCYLRRERDLMYFQ						
03-CaC2C2	-----NIKKFQEMKNIVEFRDLMDYSEIADELQGQLINWIYLRERDLMNQ						
04-Lb(NK-2)C2C2	-----KLLRFQRIKNRVEFQTI TEFSEI INELGQLISWSFLRRERDLLYFQ						
05-Cg1C2C2	-----DLLKFQRLKNAVEFRDVTYAEVINELGQLISWSYLRRERDLLYFQ						
06-Cg2C2C2	-----EEYRDYCKKIDRFNWKANKVTLTYLSQLHYLITDLLGRMVGFSA LFERDLVYFS						
07-PpC2C2	-----AEYKICEAIEFNRLDVKVRLTYLKNLHYLMIDLMGRMVGF SVL FERDFVYMG						
08-LsC2C2	FIEKHGDDYLDICDYINTYNWLDNKMHFVHLNRLHGLTIE LLGRMAGFVALFDRDFQFFD						
09-LiwC2C2	-----YQNVINDISNYQWAKTKVELTQVRHLHQLTIDL LSRLAGYMSIADRDFQFSS						
10-LibC2C2	-----YKECCDAIDAYNWHKNTPTLVYVNE LHLLIEILGRYVGYVAIADRDFQMA						
13-LspC2c2	-----GFEKDYKNVSEYK KIRDLVEFNLYLNKIESYLIDINWKLAIQMARFERDMHYIV						
14-LsC2C2	-----FEKDYNRVSEYK KIRDLVEFNLYLNKIESYLIDINWKLAIQMARFERDMHYIV						
11-Lew1C2C2	-----YKKAIEINI QNTQLKNKIEFNLDNLLQSL LFRILHRLAGYTS LWERDLQFKL						
12-Lew2C2C2	-----YEKAIGNIQKYTHLKNKVEFNELNLLQGL LKILHRLVGYTS IWERDLRFR						

FIG. 6H

15/39

	1441	1450	1460	1470	1480	1490	1500
15-RcSBC2C2	IN-----						
16-RcRC2C2	IN-----						
01-Lb(MA)C2C2	LGFHYLCLNNDSSKEADYIKI---NVDDRNISGAILYQIAAMYINGLPLVYVKDDMYVAL						
02-Lb(NK-1)C2C2	LGYHYACLNNDSNKQATYVTLDDYQGGKRNKINGAILYQICAMYINGLPLVYVDKDSSEWT						
03-CaC2C2	LGFHYLCLHNDTEKPAEYKEIS--REDGTVIRNAILHQVAAMYVGGPLVYTLADKKLAAF						
04-Lb(NK-2)C2C2	LGFHYMCLKNKSFKPAEYVDIR--RNNGTIIHNAILYQIVSMYINGLDFYSCDKEGKTLK						
05-Cg1C2C2	RS--FSELGGETYHISDYKNLSGVLRLNAEVKPIKIKNIKVIDNEENPYKG-----						
06-Cg2C2C2	RS--YSALKKQSIYLNDDYDT----FANIRDWEVNEKHLFGTSSSDLTFQ-----						
07-PpC2C2	EQ-----QIADDFKLHGFVNLHSIDKKLNEVPKKIKIYDIRNKIIQI-----						
08-LsC2C2	NY--ILERENSEYRVTPSWIL----LSENKKNKYNDYELYNLKNASIKVSSK-----						
09-LiwC2C2	NQ--YFKHSGITERVEYW-----						
10-LibC2C2	NQ--YFKHSGITERVEYW-----						
13-LspC2c2	NG----LRELGIKLSGYN-----TGISRAYPKRNGSDGFYTTTAYYKFF-----						
14-LsC2C2	NG----LRELGIKLSGYN-----TGISRAYPKRNGSDGFYTTTAYYKFF-----						
11-Lew1C2C2	KG--EYPENKYIDEIFNFD-----NSKNKIYNEKNERGGSVVSKYGYFLVEK-----						
12-Lew2C2C2	KG--EFPENHYIEEIFNFDNSKNVVKYKSGQIVEKYINFYKELYKDNVEKRSI-----						

	1501	1510	1520	1530	1540	1550	1560
15-RcSBC2C2	-----ASKQLGAGAD-----WAVTIAGAANTDA--						
16-RcRC2C2	-----ASKQLGAGAD-----WAVTIAGAANTDA--						
01-Lb(MA)C2C2	KSGKKASDELNSNE--QTSKKINYFLKYGNIL-G-----DKKDQLYLAGLELFENVA						
02-Lb(NK-1)C2C2	VSDGKES-----TGAKIGEFYR-----YAKSFENTSDCYASGLEIFENIS						
03-CaC2C2	EKGEADCKLSISKDTAGAGKKIKDFFRYSKYVLIKDRMLTDQNKYTIYLAGLELFENTD						
04-Lb(NK-2)C2C2	PIETGK-----GVGSKIGQFIKYSQY-----LYNDPSYKLEIYNAGLEVFENID						
05-Cg1C2C2	-----NEPEVKPFLD-----RLHAYLENVIGIKA--						
06-Cg2C2C2	-----ETAEFKNLKK-----PMENQLKALLGVTN--						
07-PpC2C2	-----NGNKINESVR-----ANLIQFISSKRNYNNAFLHVSNDE						
08-LsC2C2	-----NDPQLKVDLK-----QLRLTLEYLELFDN--						
09-LiwC2C2	-----GDNRLKSIKK-----LDTFLKKEGLFVS--						
10-LibC2C2	-----GDNRLKSIKK-----LDTFLKKEGLFVS--						
13-LspC2c2	-----DEESYKKFEK-----ICYGFGIDLSENSE						
14-LsC2C2	-----DEESYKKFEK-----ICYGFGIDLSENSE						
11-Lew1C2C2	-----DGEIQQRNAR-----DKKKNKIKKEGLE-----						
12-Lew2C2C2	-----YSDKKVKKLQ-----EKKDLY-----						

	1561	1570	1580	1590	1600	1610	1620
15-RcSBC2C2	----RTQ-TRKDLAHFNVLDR--DGTPDLTALVNRARE--MMAYDRKRKNAVPRSILDML						
16-RcRC2C2	----RTQ-TRKDLAHFNVLDR--DGTPDLTALVNRARE--MMAYDRKRKNAVPRSILDML						
01-Lb(MA)C2C2	EHENIII-FRNEIDHFHYFYDR----DRSMLDLYSEVDFRFFTYDMKLRKNVNVNMLYNIL						
02-Lb(NK-1)C2C2	EHDNITE-LRNYIEHFRYYSF----DRSFLGIYSEVDFRFFTYDLKYRKNVPTILYNIL						
03-CaC2C2	EHDNITD-VRKYVDHFKYAATS--DENAMSILDLYSEIHDRFFTYDMKYQKNVANMLLENIL						
04-Lb(NK-2)C2C2	EHDNITD-LRKYVDHFKYAYG---NKMSLLDLYSEFFDRFFTYDMKYQKNVNVNLENIL						
05-Cg1C2C2	----VHGKIRNQTAAHLSVLQLE----LSMIESMNNLRD--LMAYDRKLKNAVTKSMIKIL						
06-Cg2C2C2	----HSFEIRNNTIAHLHLRNDGKGEVSVLLSCMNDLRK--LMSYDRKLKNAVTKAIKIL						
07-PpC2C2	IKEKQMYDIRNHIAHFNYLTKD--AADFSLIDLINELRE--LLHYDRKLKNAVSKAFIDLF						
08-LsC2C2	----RLKEKRNNISHFNLYNGQ---LGNSILELFDWARD--VLSYDRKLKNAVSKSLKEIL						
09-LiwC2C2	----EKN-ARNHIAHLNLYLSLK---SECTLLYLSERLRE--IFKYDRKLKNAVSKSLIDIL						
10-LibC2C2	----EKN-ARNHIAHLNLYLSLK---SECTLLYLSERLRE--IFKYDRKLKNAVSKSLIDIL						
13-LspC2c2	INKPENESIRNYISHFYIVRNP--FADYSIAEQIDRVSN--LLSYSTRYNNSTYASVFEVF						
14-LsC2C2	INKPENESIRNYISHFYIVRNP--FADYSIAEQIDRVSN--LLSYSTRYNNSTYASVFEVF						
11-Lew1C2C2	-----IRNYIAHFNYIPDA----TKSILEILEELRN--LLKYDRKLKNAVMSKIKDIF						
12-Lew2C2C2	-----IRNYIAHFNYIPHA----EISLLEVLLENLRK--LLSYDRKLKNAIMKSIDVIDL						

FIG. 61

16/39

	1621	1630	1640	1650	1660	1670	1680
15-RcSBC2C2	ARLGLTLKWQ	--MKDHL	LQDATIT	QAAIKHL	DKVRLTV	GGPAAV	TEARFSQDY
16-RcRC2C2	ARLGLTLKWQ	--MKDHL	LQDATIT	QAAIKHL	DKVRLTV	GGPAAV	TEARFSQDY
01-Lb(MA)C2C2	LDHNIVSS	VFVFET	GEEKVGR	GDSEVI	----	KPSAKIRL	RANNGVSSD
02-Lb(NK-1)C2C2	LQHFVNV	RFEFV	SGKKMIG	DKKDRK	IAKEKE	CARITIRE	KNGVYSEQ
03-CaC2C2	LRHFVLI	RPEFFT	GSKKVGE	GKKITC	----	KARAQIE	IAE-NGMR
04-Lb(NK-2)C2C2	LRHFVLI	FYPKFG	SGKKDVG	IRDCKKE	----	RAQIEISE	-QSLTSED
05-Cg1C2C2	DKHGMIL	KLK----	IDENHKN	-----	-----	-----	-----
06-Cg2C2C2	DKHGMIL	KLK----	NNDHTKP	-----	-----	-----	-----
07-PpC2C2	DKHGMIL	KLK----	LNADHK	-----	-----	-----	-----
08-LsC2C2	SSHGMEV	TFK----	PLYQTN	NHH-----	-----	-----	-----
09-LiwC2C2	DRHGMSV	VFA----	NLKENKH	R-----	-----	-----	-----
10-LibC2C2	DRHGMSV	VFA----	NLKENKH	R-----	-----	-----	-----
13-LspC2c2	KKDVNLD	YDELK	KKKFR	LIGNND	I-----	-----	-----
14-LsC2C2	KKDVNLD	YDELK	KKKFK	LIGNND	I-----	-----	-----
11-Lew1C2C2	KEYGLI	IEFK----	ISHVNN	SEK-----	-----	-----	-----
12-Lew2C2C2	KEYGFV	VATFK----	IGADKK	-----	-----	-----	-----

	1681	1690	1700	1710	1720	1730	1740
15-RcSBC2C2	FNGSVQNP	KPRRRD	DGD	AHWKPP	KPATAQ	SQPDQ	KPPNKAPS
16-RcRC2C2	FNGSVQNP	KPRRRD	DGD	AHWKPP	KPATAQ	SQPDQ	KPPNKAPS
01-Lb(MA)C2C2	KIATLPA	KNEEFL	LNVARL	IYYPD	MEAVSEN	MVREGV	VVKVEKS
02-Lb(NK-1)C2C2	GTVYVD	DARDK	RYLQSI	IRLLFY	PEKVN	MDEMIE	VEVKEK
03-CaC2C2	STCMIA	AARDQ	KYLNTP	VARLLY	PHEAK	-KSIVD	TREKKN
04-Lb(NK-2)C2C2	EAKKFP	PARDER	YLTQTI	AKLLY	YPNEI	EDMNR	FMKKGET
05-Cg1C2C2	EIESLIP	KEI IHL	-----	-----	-----	KDKAIK	-----
06-Cg2C2C2	EIESLIP	KEI IHL	-----	-----	-----	EKSNS	HFP----
07-PpC2C2	KVESLE	PKKIYHL	-----	-----	-----	GSSAK	DKPEY
08-LsC2C2	KIDKLQ	PKKIHHL	-----	-----	-----	GEKST	VS-----
09-LiwC2C2	VIKSLE	PKKLRHL	-----	-----	-----	GEKKI	DNGYIE
10-LibC2C2	VIKSLE	PKKLRHL	-----	-----	-----	GGKKI	DGGYIE
13-LspC2c2	---	LMKPK	KVSVL	-----	-----	ELESYN	-----
14-LsC2C2	---	LMKPK	KVSVL	-----	-----	ELESYN	-----
11-Lew1C2C2	EVLNV	DSEK	IKHL	-----	-----	KNNGLV	-----
12-Lew2C2C2	EIQTL	ESEK	IVHL	-----	-----	KNLK	KKKLM

	1741	1750	1760	1770	1780	1790	1800
15-RcSBC2C2	-EVEYEG	VVVKV	IDT	GS	LGFLA	VEGVAG	NIGLHIS
16-RcRC2C2	-EVEYEG	VVVKV	IDT	GS	LGFLA	VEGVAG	NIGLHIS
01-Lb(MA)C2C2	QSKYNN	KSKNRM	NYSMGS	IFEKMD	LKFD	-----	-----
02-Lb(NK-1)C2C2	-KEYDKY	KEKKE	KNFLS	GMGGN	INWDE	INAQL	KN-----
03-CaC2C2	-KEKDNG	PREFND	TGFSNT	PPFAG	DFFRNS	-----	-----
04-Lb(NK-2)C2C2	-NSSNEV	LSS	TMGYL	FKNIKL	-----	-----	-----
05-Cg1C2C2	-EEYCQ	LVLALL	TNP	GNQLN	-----	-----	-----
06-Cg2C2C2	-QEYCD	LVKKML	VFTN	-----	-----	-----	-----
07-PpC2C2	-MAYCN	MCRSL	LLEMKK	-----	-----	-----	-----
08-LsC2C2	-NEYCQ	LVRTL	LLTMK	-----	-----	-----	-----
09-LiwC2C2	-EEYCG	IVKR	LLEI	-----	-----	-----	-----
10-LibC2C2	-EEYCG	IVKR	LLEM	-----	-----	-----	-----
13-LspC2c2	-NLIIE	LLTK	IEN	TNDTL	-----	-----	-----
14-LsC2C2	-NLIIE	LLTK	IEN	TNDTL	-----	-----	-----
11-Lew1C2C2	-EDLCE	LKMM	LLEY	KKS	-----	-----	-----
12-Lew2C2C2	-EELCE	LKV	VMF	EYKALE	-----	-----	-----

FIG. 6J

17/39

	1801	1810	1820	1830	1840	1850	1860
15-RcSBC2C2	KSNTSKLNAADLVRID						
16-RcRC2C2	KSNTSKLNAADLVRID						
01-Lb(MA)C2C2	-----						
02-Lb(NK-1)C2C2	-----						
03-CaC2C2	-----						
04-Lb(NK-2)C2C2	-----						
05-Cg1C2C2	-----						
06-Cg2C2C2	-----						
07-PpC2C2	-----						
08-LsC2C2	-----						
09-LiwC2C2	-----						
10-LibC2C2	-----						
13-LspC2c2	-----						
14-LsC2C2	-----						
11-Lew1C2C2	-----						
12-Lew2C2C2	-----						

FIG. 6K

	586	603	1276	1290
c2c2_Leptotrichia shahii	IRKFTKIGIN ERNRILNA		SIRNYISHFYIVRNP	
c2c2-5 Lachnospiraceae bacterium MA2020	LYSLKSMLYS MRNSSFHF		IFRNEIDNFHYFYDR	
c2c2-6 Lachnospiraceae	LTDLKDVIYS MRNDSFHY		ELRNYIENFRYSSF	
c2c2-7 [Clistidium] aminophilum DSM 10710	ADDLRKAIYS LRNETFHF		DVRKYVDHFKYATS	
c2c2-8 Carnobacterium gallinarum DSM 4847	IWAIRGAVQQ IRNEIFHS		KIRNQTAHL SVLQLE	
c2c2-9 Carnobacterium gallinarum DSM 4847	LWATRGAVQR VRNQIFHQ		EIRNVIAMHIVLRND	
c2c2-10 Paludibacter propionicigenes WB4	LNGIRGAVQQ IRNNVWHY		DIRNVIAMHNYLTKD	
c2c2-11 Listeria weihenstephanensis FSL R9-0317	IWAIRGSIQQ IRNEVYHK		NARNVIAMHNYLSLK	
c2c2-12 Listeriaceae bacterium FSL MB-0636	IWAIRGSIQQ IRNEVYHK		NARNVIAMHNYLSLK	
c2c2-13 Leptotrichia wadei F0279	FANIDEAISS IRNGIVHF		YIRNYIAHFNYIPHA	
c2c2-14 Rhodobacter capsulatus SB 1003	VFALLRYLRG CRNQTFHL		QTRKDLAHFNVLDR	
c2c2-15 Rhodobacter capsulatus R121	VFALLRYLRG CRNQTFHL		QTRKDLAHFNVLDR	
c2c2-16 Rhodobacter capsulatus DE442	VFALLRYLRG CRNQTFHL		QTRKDLAHFNVLDR	
c2-3 L wadei (Lw2)	FANIDEAISS IRNGIVHF		YIRNYIAHFNYIPHA	
c2-4 Listeria seeligen	SWGLRGAIAP IRNETIHL		EKRNIISHFNYLNGQ	
		↑↑ ↑	↑↑ ↑	

FIG. 6L

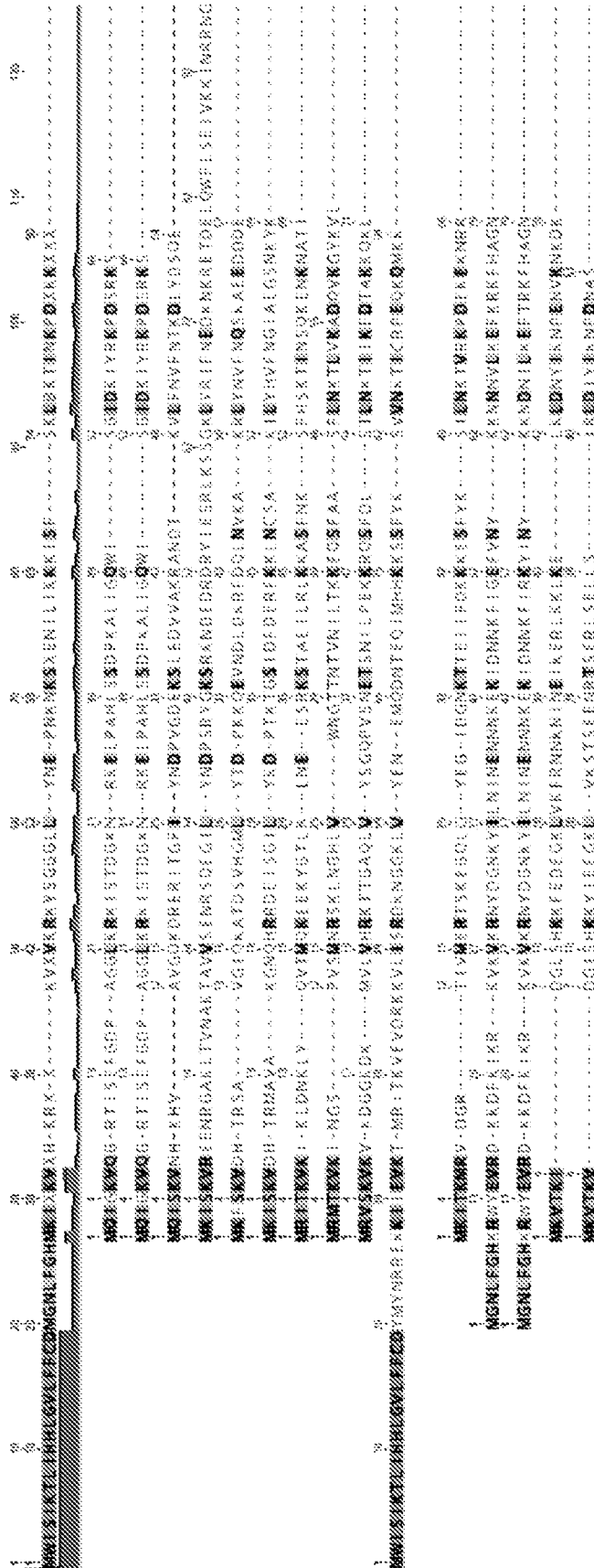


FIG. 7A

```

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

```

FIG. 7B

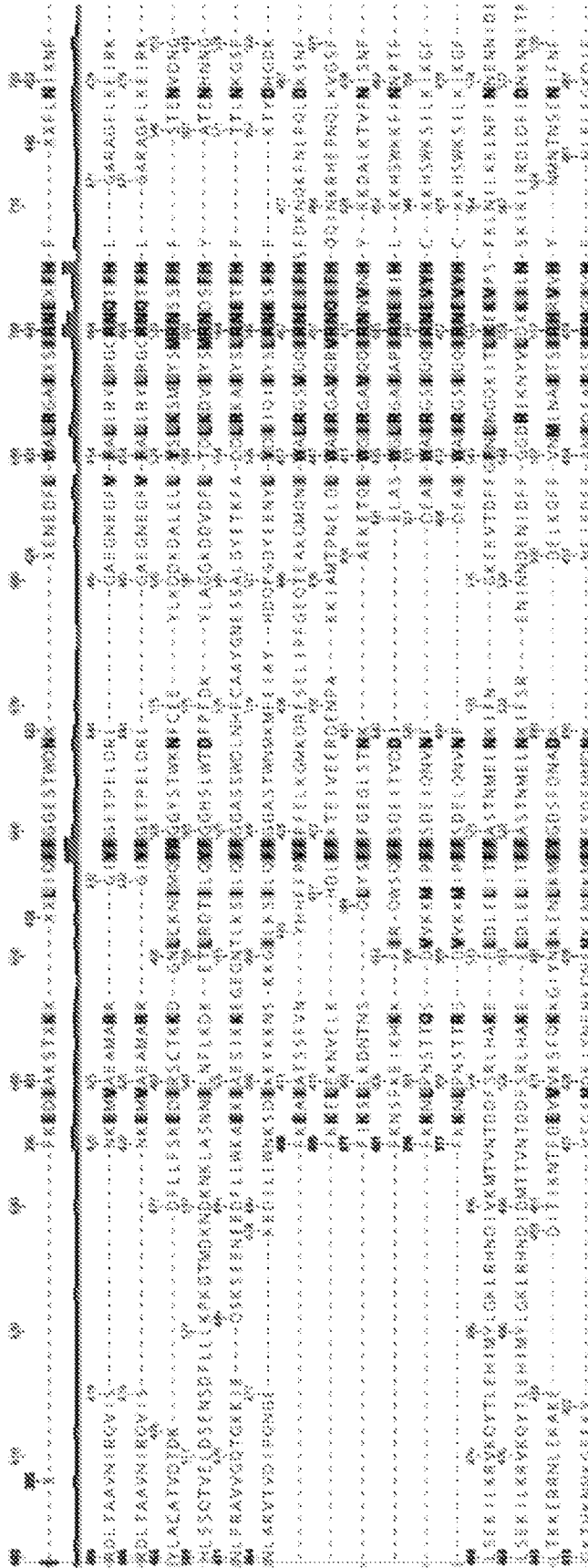


FIG. 7F

```

68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
...
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
...
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
...
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250
...
251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
...
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350
...
351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
...
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
...
451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
...
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550
...
551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
...
601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650
...
651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700
...
701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750
...
751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800
...
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850
...
851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900
...
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950
...
951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
...

```

FIG. 7G

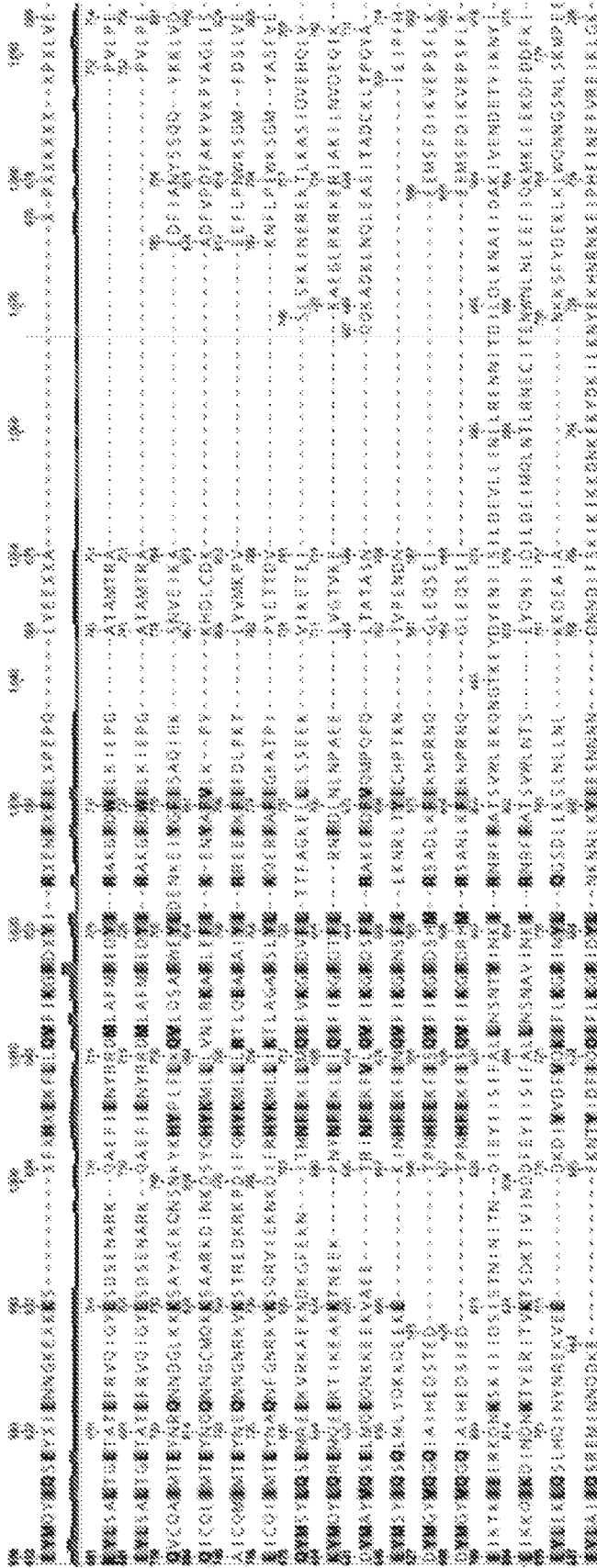


FIG. 71

1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257
 1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286
 1287
 1288
 1289
 1290
 1291
 1292
 1293
 1294
 1295
 1296
 1297
 1298
 1299
 1300
 1301
 1302
 1303
 1304
 1305
 1306
 1307
 1308
 1309
 1310
 1311
 1312
 1313
 1314
 1315
 1316
 1317
 1318
 1319
 1320
 1321
 1322
 1323
 1324
 1325
 1326
 1327
 1328
 1329
 1330
 1331
 1332
 1333
 1334
 1335
 1336
 1337
 1338
 1339
 1340
 1341
 1342
 1343
 1344
 1345
 1346
 1347
 1348
 1349
 1350
 1351
 1352
 1353
 1354
 1355
 1356
 1357
 1358
 1359
 1360
 1361
 1362
 1363
 1364
 1365
 1366
 1367
 1368
 1369
 1370
 1371
 1372
 1373
 1374
 1375
 1376
 1377
 1378
 1379
 1380
 1381
 1382
 1383
 1384
 1385
 1386
 1387
 1388
 1389
 1390
 1391
 1392
 1393
 1394
 1395
 1396
 1397
 1398
 1399
 1400
 1401
 1402
 1403
 1404
 1405
 1406
 1407
 1408
 1409
 1410
 1411
 1412
 1413
 1414
 1415
 1416
 1417
 1418
 1419
 1420
 1421
 1422
 1423
 1424
 1425
 1426
 1427
 1428
 1429
 1430
 1431
 1432
 1433
 1434
 1435
 1436
 1437
 1438
 1439
 1440
 1441
 1442
 1443
 1444
 1445
 1446
 1447
 1448
 1449
 1450
 1451
 1452
 1453
 1454
 1455
 1456
 1457
 1458
 1459
 1460
 1461
 1462
 1463
 1464
 1465
 1466
 1467
 1468
 1469
 1470
 1471
 1472
 1473
 1474
 1475
 1476
 1477
 1478
 1479
 1480
 1481
 1482
 1483
 1484
 1485
 1486
 1487
 1488
 1489
 1490
 1491
 1492
 1493
 1494
 1495
 1496
 1497
 1498
 1499
 1500
 1501
 1502
 1503
 1504
 1505
 1506
 1507
 1508
 1509
 1510
 1511
 1512
 1513
 1514
 1515
 1516
 1517
 1518
 1519
 1520
 1521
 1522
 1523
 1524
 1525
 1526
 1527
 1528
 1529
 1530
 1531
 1532
 1533
 1534
 1535
 1536
 1537
 1538
 1539
 1540
 1541
 1542
 1543
 1544
 1545
 1546
 1547
 1548
 1549
 1550
 1551
 1552
 1553
 1554
 1555
 1556
 1557
 1558
 1559
 1560
 1561
 1562
 1563
 1564
 1565
 1566
 1567
 1568
 1569
 1570
 1571
 1572
 1573
 1574
 1575
 1576
 1577
 1578
 1579
 1580
 1581
 1582
 1583
 1584
 1585
 1586
 1587
 1588
 1589
 1590
 1591
 1592
 1593
 1594
 1595
 1596
 1597
 1598
 1599
 1600
 1601
 1602
 1603
 1604
 1605
 1606
 1607
 1608
 1609
 1610
 1611
 1612
 1613
 1614
 1615
 1616
 1617
 1618
 1619
 1620
 1621
 1622
 1623
 1624
 1625
 1626
 1627
 1628
 1629
 1630
 1631
 1632
 1633
 1634
 1635
 1636
 1637
 1638
 1639
 1640
 1641
 1642
 1643
 1644
 1645
 1646
 1647
 1648
 1649
 1650
 1651
 1652
 1653
 1654
 1655
 1656
 1657
 1658
 1659
 1660
 1661
 1662
 1663
 1664
 1665
 1666
 1667
 1668
 1669
 1670
 1671
 1672
 1673
 1674
 1675
 1676
 1677
 1678
 1679
 1680
 1681
 1682
 1683
 1684
 1685
 1686
 1687
 1688
 1689
 1690
 1691
 1692
 1693
 1694
 1695
 1696
 1697
 1698
 1699
 1700
 1701
 1702
 1703
 1704
 1705
 1706
 1707
 1708
 1709
 1710
 1711
 1712
 1713
 1714
 1715
 1716
 1717
 1718
 1719
 1720
 1721
 1722
 1723
 1724
 1725
 1726
 1727
 1728
 1729
 1730
 1731
 1732
 1733
 1734
 1735
 1736
 1737
 1738
 1739
 1740
 1741
 1742
 1743
 1744
 1745
 1746
 1747
 1748
 1749
 1750
 1751
 1752
 1753
 1754
 1755
 1756
 1757
 1758
 1759
 1760
 1761
 1762
 1763
 1764
 1765
 1766
 1767
 1768
 1769
 1770
 1771
 1772
 1773
 1774
 1775
 1776
 1777
 1778
 1779
 1780
 1781
 1782
 1783
 1784
 1785
 1786
 1787
 1788
 1789
 1790
 1791
 1792
 1793
 1794
 1795
 1796
 1797
 1798
 1799
 1800
 1801
 1802
 1803
 1804
 1805
 1806
 1807
 1808
 1809
 1810
 1811
 1812
 1813
 1814
 1815
 1816
 1817
 1818
 1819
 1820
 1821
 1822
 1823
 1824
 1825
 1826
 1827
 1828
 1829
 1830
 1831
 1832
 1833
 1834
 1835
 1836
 1837
 1838
 1839
 1840
 1841
 1842
 1843
 1844
 1845
 1846
 1847
 1848
 1849
 1850
 1851
 1852
 1853
 1854
 1855
 1856
 1857
 1858
 1859
 1860
 1861
 1862
 1863
 1864
 1865
 1866
 1867
 1868
 1869
 1870
 1871
 1872
 1873
 1874
 1875
 1876
 1877
 1878
 1879
 1880
 1881
 1882
 1883
 1884
 1885
 1886
 1887
 1888
 1889
 1890
 1891
 1892
 1893
 1894
 1895
 1896
 1897
 1898
 1899
 1900
 1901
 1902
 1903
 1904
 1905
 1906
 1907
 1908
 1909
 1910
 1911
 1912
 1913
 1914
 1915
 1916
 1917
 1918
 1919
 1920
 1921
 1922
 1923
 1924
 1925
 1926
 1927
 1928
 1929
 1930
 1931
 1932
 1933
 1934
 1935
 1936
 1937
 1938
 1939
 1940
 1941
 1942
 1943
 1944
 1945
 1946
 1947
 1948
 1949
 1950
 1951
 1952
 1953
 1954
 1955
 1956
 1957
 1958
 1959
 1960
 1961
 1962
 1963
 1964
 1965
 1966
 1967
 1968
 1969
 1970
 1971
 1972
 1973
 1974
 1975
 1976
 1977
 1978
 1979
 1980
 1981
 1982
 1983
 1984
 1985
 1986
 1987
 1988
 1989
 1990
 1991
 1992
 1993
 1994
 1995
 1996
 1997
 1998
 1999
 2000
 2001
 2002
 2003
 2004
 2005
 2006
 2007
 2008
 2009
 2010
 2011
 2012
 2013
 2014
 2015
 2016
 2017
 2018
 2019
 2020
 2021
 2022
 2023
 2024
 2025
 2026
 2027
 2028
 2029
 2030
 2031
 2032
 2033
 2034
 2035
 2036
 2037
 2038
 2039
 2040
 2041
 2042
 2043
 2044
 2045
 2046
 2047
 2048
 2049
 2050
 2051
 2052
 2053
 2054
 2055
 2056
 2057
 2058
 2059
 2060
 2061
 2062
 2063
 2064
 2065
 2066
 2067
 2068
 2069
 2070
 2071
 2072
 2073
 2074
 2075
 2076
 2077
 2078
 2079
 2080
 2081
 2082
 2083
 2084
 2085
 2086
 2087
 2088
 2089
 2090
 2091
 2092
 2093
 2094
 2095
 2096
 2097
 2098
 2099
 2100
 2101
 2102
 2103
 2104
 2105
 2106
 2107
 2108
 2109
 2110
 2111
 2112
 2113
 2114
 2115
 2116
 2117
 2118
 2119
 2120
 2121
 2122
 2123
 2124
 2125
 2126
 2127
 2128
 2129
 2130
 2131
 2132
 2133
 2134
 2135
 2136
 2137
 2138
 2139
 2140
 2141
 2142
 2143
 2144
 2145
 2146
 2147
 2148
 2149
 2150
 2151
 2152
 2153
 2154
 2155
 2156
 2157
 2158
 2159
 2160
 2161
 2162
 2163
 2164
 2165
 2166
 2167
 2168
 2169
 2170
 2171
 2172
 2173
 2174
 2175
 2176
 2177
 2178
 2179
 2180
 2181
 2182
 2183
 2184
 2185
 2186
 2187
 2188
 2189
 2190
 2191
 2192
 2193
 2194
 2195
 2196
 2197
 2198
 2199
 2200
 2201
 2202
 2203
 2204
 2205
 2206
 2207
 2208
 2209
 2210
 2211
 2212
 2213
 2214
 2215
 2216
 2217
 2218
 2219
 2220
 2221
 2222
 2223
 2224
 2225
 2226
 2227
 2228
 2229
 2230
 2231
 2232
 2233
 2234
 2235
 2236
 2237
 2238
 2239
 2240
 2241
 2242
 2243
 2244
 2245
 2246
 2247
 2248
 2249
 2250
 2251
 2252
 2253
 2254
 2255
 2256
 2257
 2258
 2259
 2260
 2261
 2262
 2263
 2264
 2265
 2266
 2267
 2268
 2269
 2270
 2271
 2272
 2273
 2274
 2275
 2276
 2277
 2278
 2279
 2280
 2281
 2282
 2283
 2284
 2285
 2286
 2287
 2288
 2289
 2290
 2291
 2292
 2293
 2294
 2295
 2296
 2297
 2298
 2299
 2300
 2301
 2302
 2303
 2304
 2305
 2306
 2307
 2308
 2309
 2310
 2311
 2312
 2313
 2314
 2315
 2316
 2317
 2318
 2319
 2320
 2321
 2322
 2323
 2324
 2325
 2326
 2327
 2328
 2329
 2330
 2331
 2332
 2333
 2334
 2335
 2336
 2337
 2338
 2339
 2340
 2341
 2342
 2343
 2344
 2345
 2346
 2347
 2348
 2349
 2350
 2351
 2352
 2353
 2354
 2355
 2356
 2357
 2358
 2359
 2360
 2361
 2362
 2363
 2364
 2365
 2366
 2367
 2368
 2369
 2370
 2371
 2372
 2373
 2374
 2375
 2376
 2377
 2378
 2379
 2380
 2381
 2382
 2383
 2384
 2385
 2386
 2387
 2388
 2389
 2390
 2391
 2392
 2393
 2394
 2395
 2396
 2397
 2398
 2399
 2400
 2401
 2402
 2403
 2404
 2405
 2406
 2407
 2408
 2409
 2410
 2411
 2412
 2413
 2414
 2415
 2416
 2417
 2418
 2419
 2420
 2421
 2422
 2423
 2424
 2425
 2426
 2427
 2428
 2429
 2430
 2431
 2432
 2433
 2434
 2435
 2436
 2437
 2438
 2439
 2440
 2441
 2442
 2443
 2444
 2445
 2446
 2447
 2448
 2449
 2450
 2451
 2452
 2453
 2454
 2455
 2456
 2457
 2458
 2459
 2460
 2461
 2462
 2463
 2464
 2465
 2466
 2467
 2468
 2469
 2470
 2471
 2472
 2473
 2474
 2475
 2476
 2477
 2478
 2479
 2480
 2481
 2482
 2483
 2484
 2485
 2486
 2487
 2488
 2489
 2490
 2491
 2492
 2493
 2494
 2495
 2496
 2497
 2498
 2499
 2500
 2501
 2502
 2503
 2504
 2505
 2506
 2507
 2508
 2509
 2510
 2511
 2512
 2513
 2514
 2515
 2516
 2517
 2518
 2519
 2520
 2521
 2522
 2523
 2524
 2525
 2526
 2527
 2528
 2529
 2530
 2531
 2532
 2533
 2534
 2535
 2536
 2537
 2538
 2539
 2540
 2541
 2542
 2543
 2544
 2545
 2546
 2547
 2548
 2549
 2550
 2551
 2552
 2553
 2554
 25

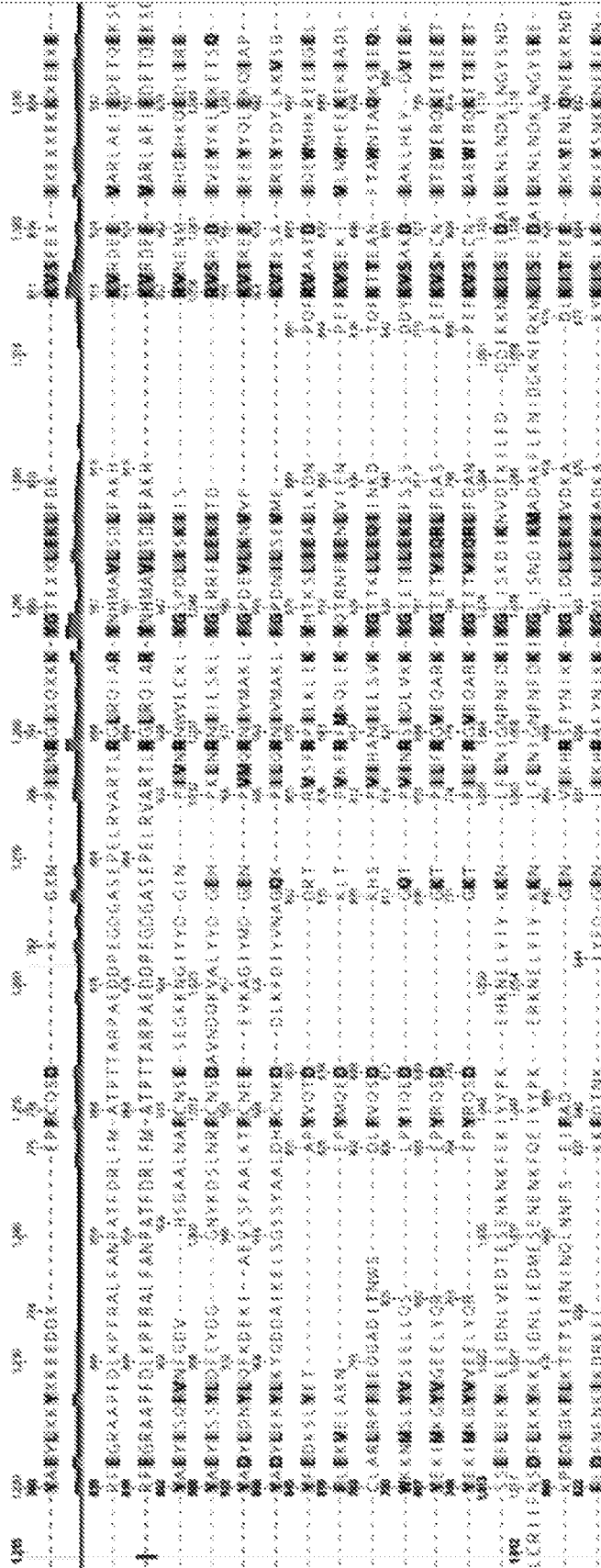


FIG. 7K

1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000

FIG. 7N

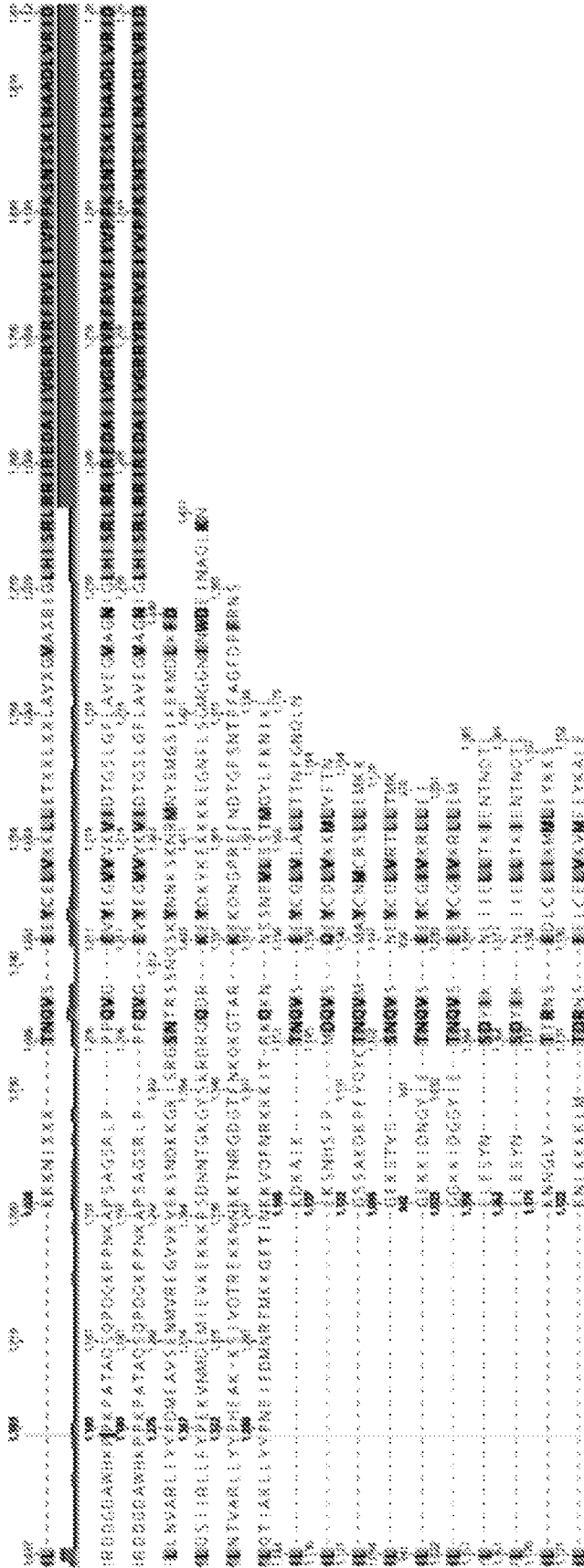


FIG. 70

33/39

>LibFSL and Lw2 Alignment of 2 sequences: 10-LibC2C2, 12-Lew2C2C2

Identities = 782/1752 (44%),
Positives = 995/1752 (56%), Gaps = 1301/1752 (74%)

```

10-LibC2C2      1 -----MKITKMRVDG-----
RTIVME 16
-----MK+TK+-----
- I +
12-Lew2C2C2    1 -----MKVTKVDG-----
ISHK 12

10-LibC2C2     17 RTSKEGQLGYEG-----IDGNKTTEIIFDKK-KESFYKSILNKTVRK-----
-- 57
+ +EG+L ----- + N+T+E + + - I N -----
--
12-Lew2C2C2    13 KYIEEGKLVKST-----SEENRTSERLSELL-SIRLDIYIKNPDNA-----
-- 52

10-LibC2C2     58 PDEKEKNRRKQAINKAINKEITE-----LMLAVLHQEV-----
-- 90
+ E+ + RR + + K + ----- + L + +-----
--
12-Lew2C2C2    53 SEEENRIRR-----ENLKKFFSN-----KVLHLKDSVL-----
-- 80

10-LibC2C2     91 -----PSQKLHNLKSLNTES---LTKLFPKPKFQNMISYPPSKGAE-----
HVQFCL 133
----- L N K N --- + Y -----
-
12-Lew2C2C2    81 -----YLKNRKEKNAVQ---DKNYSEEDISE---YDLKNKNS-----
FSVLKK 117

10-LibC2C2    134 TDIAPPAIR---DLDEIKPD---WGIFFEKLPYTDWAESYIHY--KQTTIQKS----
-- 179
+ ---D++ + +--- IF + ++ + S + + + ----
--
12-Lew2C2C2   118 ILLNE-----DVNSEELE-----IFRKDVEAKLNKINSLKYSFEENKANYQK----
-- 159

10-LibC2C2    180 --IEQNK-----IQSPD-----SPRKLVLQKY--VTAFLNGEPLGLDLV---
-- 214
--I +N ----- + -----
- R ++ Y + +
12-Lew2C2C2   160 --INENN-----VEKVGGS-----
KRNIIDYDYRESAKRNDYINNVQEAQFDKLYK 203

10-LibC2C2    215 -----AKKYKLADLAESFKLVLDLNEDK-----
-- 236
----- +KYK+ + +++ -----
--
12-Lew2C2C2   204 KEDIEKLFFLIENSKK-----HEKYKIREYYHKIIGRKNDKENFAKIIYEEI-
-- 250

10-LibC2C2    237 -----SANYKIKACLQQH-----QRNILD--ELKEDPELNQY----
GIEV 270
-----

```

FIG. 8A

```

-      IK  +++          + LD EL +      +      IE+
12-Lew2C2C2    251 -----
QNVNNIKELIEKIPDMSELKKSQVFYKYYLDKEELNDKNIKYAFCHFVEIEM  302

10-LibC2C2    271 KKYI---QRYFPIKRAPNR-
SKHARADFLKKELIESTVEQQFKNAVYHYVLEQGKMEAY-  325
          + +--- + + KR N -
S          + + ++ +E + N + YV GK Y-
12-Lew2C2C2    303 SPLL----KNYVYKRLSNI-
SNDKIKRIFEQNLKLIENKLLNKLDTYVRNCGKYNYY-  356

10-LibC2C2    326 --ELTDPKTKD-----LQDIRSGEAFSFKFINACAFASNNL-
KM  361
          - ++ + T D----- + R EAF I + A +L-
+
12-Lew2C2C2    357 -LQVGEIATSD-----FIARNRQNEAFLRNIIGVSSVAYFSL-
RN  394

```

FIG. 8A (continued)

35/39

10-LibC2C2 362 ILNP-----ECE-----KDILG---KGNFKKNLPNS-----
 -- 384
 IL -----E E----- DI G---
 + K N
 12-Lew2C2C2 395 ILET-----ENE-----NDITG---
 RMRGKTVKNNKGEEKYVSGEVDKI 430

10-LibC2C2 385 ---TTRSDVVKMIPFFS-----DELQNVNFD-----EAIWA-
 IRGSIQQIRNEV 425
 +++V + + F+S----- + N +--
 - E +A I +I IR+ +
 12-Lew2C2C2 431 YNENKQNEVKENLKMFFYS-----YDFNMDNKN---
 EIEDFFANIDEAIISSIRHGI 477

10-LibC2C2 426 YH-C--KKHSWSILKIKGFEPNMM-----KYADSDMQKLMKDIA--
 KIP 468
 H- -- + F F+ -----
 - +K+ +I K+
 12-Lew2C2C2 478 VH-F-----NLELEGKDIFAFKN-----
 IAPSEISKMFQNEINEKCLK 515

10-LibC2C2 469 EFIEEKLKSSGVVRFYRHDELQSIWEM-KQGFSLTTNAPFVPSFKRVYAKG---
 HDYQT 524
 I ++L S+ V +Y D + + - F+ + N PFVPSF ++Y K --
 - D +
 12-Lew2C2C2 516 LKIFKQLNSANVFNYEKDVIKYLK--NTKFNFNKNI PFVPSFTKLYNKI---
 EDLRN 570

10-LibC2C2 525 SKN--RYYNLDLTT---FDILE-YGEEDFRARYFLTKLVYYQQFMPWFT-
 ADNNAFRDAA 577
 + -- L ---
 F + - +E+ A+ +L K +YY +F+ F - ++ F
 12-Lew2C2C2 571 T-----LKF---FWSVP-KDKEEKDAQIYLLKNIYYGEFLNKFV-
 KNSKVFFKIT 615

10-LibC2C2 578 NFV-----LRLNKNRQQDAK-----
 AFINIREVEEGEMPRDYM 610
 N V-----++NK R Q -----
 - + +E +P +Y+
 12-Lew2C2C2 616 NEV-----IKINKQRNQKTG-----HYKYQKFENIE-
 KTVPVEYL 649

10-LibC2C2 611 GYVQGQIAIH-----EDSIEDTPN-HFEKFISQVFIKGFDRHM--
 RSANL 652
 +Q + I+ ----- ++ N- + FI Q+F+KGF ++-
 - NL
 12-Lew2C2C2 650 AIIQSREMINN-----QDKEEKN-TYIDFIQQIFLKGFIIDYL--
 NKNNL 690

10-LibC2C2 653 KFIKNPRNQGLEQSEI-----
 EEMSFDIKVEPS 680
 K+I++ N E +IK+

Lew2C2C2 691 KYIESNNNDNDNDFSKIKIKKDNKEKYDKILKNYEKHNRNKEIPHEINEFVREIKLGKI
 750

FIG. 8B

```

10-LibC2C2      681 FLKNKDD-YIAFWIFCKMLDARHLSELRNEMIKYDGH-----
                -- 717
                ++ - F++ K+L+ + L+ L+ + KY -----
                --
12-Lew2C2C2     751 LKYTEN--LNMFYLIKLLNHNKELTNLKGSLEKYQSANKEETF-----
                -- 791

10-LibC2C2      718 -----TGEQEIIIGLALLGVDSRENDWKQFFSSEREYEKIMKGYVVEE-----
                -- 759
                -----+ E E+I L L + D++ + ++ + + + -----
                --
12-Lew2C2C2     792 -----SDELELINLLNLDNNRVTEDFELEANEIGKFLDFNENKIKDRKE-----
                -- 835

10-LibC2C2      760 --LYQ-REPY-----RQSD-GKTPILF-----
                RGVEQARKYGTETVIQRLFDANPEF 802
                --L + ----- D-G+ I -----
                R +KYG +++++ D ++
12-Lew2C2C2     836 --LKKFDTNK-----IYFD-GENIIKH-----RAFYNIKKYGMLNLLEKIADKA-
                KY 878

```

FIG. 8B (continued)

37/39

10-LibC2C2 803 KVSKCNLAEWERQKETIEET--IKRRKELHNEWAK-----
 NPKKPQN 842
 K+S L E+ +K IE+ -- ++ LH ++A+-----
 - + N
 12-Lew2C2C2 879 KISLKEKKEYSNKKNEIEKN--YTMQQNLHRKYAR-----
 PKKDEKFN 919

10-
 LibC2C2 843 NAFFKEYKECCDAIDAYNWHKNKTTLAYVNELHLLIEILGRYVGYVAIADRDFQCMANQ
 902
 + +KEY++ I Y KNK +N L LL++IL R VGY +I +RD +

12-
 Lew2C2C2 920 DEDYKEYEKAIGNIQKYTHLKNKVEFNELNLLQGLLLKILHRLVGYTSIWERDLRFRLKG
 979

10-LibC2C2 903 --YFKHSGITERVEYW-----
 -- 916
 -- F + E + + -----
 --

12-Lew2C2C2 980 --EFPENHYIEEIFNFDNSKNVKYKSGQIVEKYINFYKELYKDNVEKRSIY-----
 -- 1028

10-LibC2C2 917 -----GDNRLKSIKK-----
 LDTFLKKEGLFVSE- 940
 ----- D ++K +K+-----
 - +

12-Lew2C2C2 1029 -----SDKKVKKLLKQ-----
 EKKD 1042

10-LibC2C2 941 -----KNARNHIAHLNYLSLK---SECTLLYLSERLRE-
 IFKYDRKLNKNAVSKSLIDILD 991
 ----- RN+IAH NY+ --- +LL + E LR+-
 + YDRKLNKNA+ KS++DIL

12-Lew2C2C2 1043 -----LYIRNYIAHFNYIPHA---E-ISLLEVLENLRK-
 LLSYDRKLNKNAIMKSIVDILK 1092

10-LibC2C2 992 RHGMSVVFA-----NLKENKHLRVIKS-----
 -- 1013
 +G F ----- ++ I++-----
 --

12-Lew2C2C2 1093 EYGFVATFK-----IGADKKIEIQT-----
 -- 1112

10-LibC2C2 1014 -----LEPKKLRHLGGKKIDGGYIE--
 TNQVSEEY 1041
 -----LE +K+ HL K --

T++ SEE
 12-Lew2C2C2 1113 -----LESEKIVHLKNLKKKKLM-----
 TDRNSEEL 1138

10-LibC2C2 1042 CGIVKRLLEM-----
 -- 1051
 C +VK + E -----
 --

12-Lew2C2C2 1139 CELVKVMFEYKA-----LE-----
 -- 1152

FIG. 8C

10-LibC2C2 -----

12-Lew2C2C2 -----

FIG. 8C (continued)

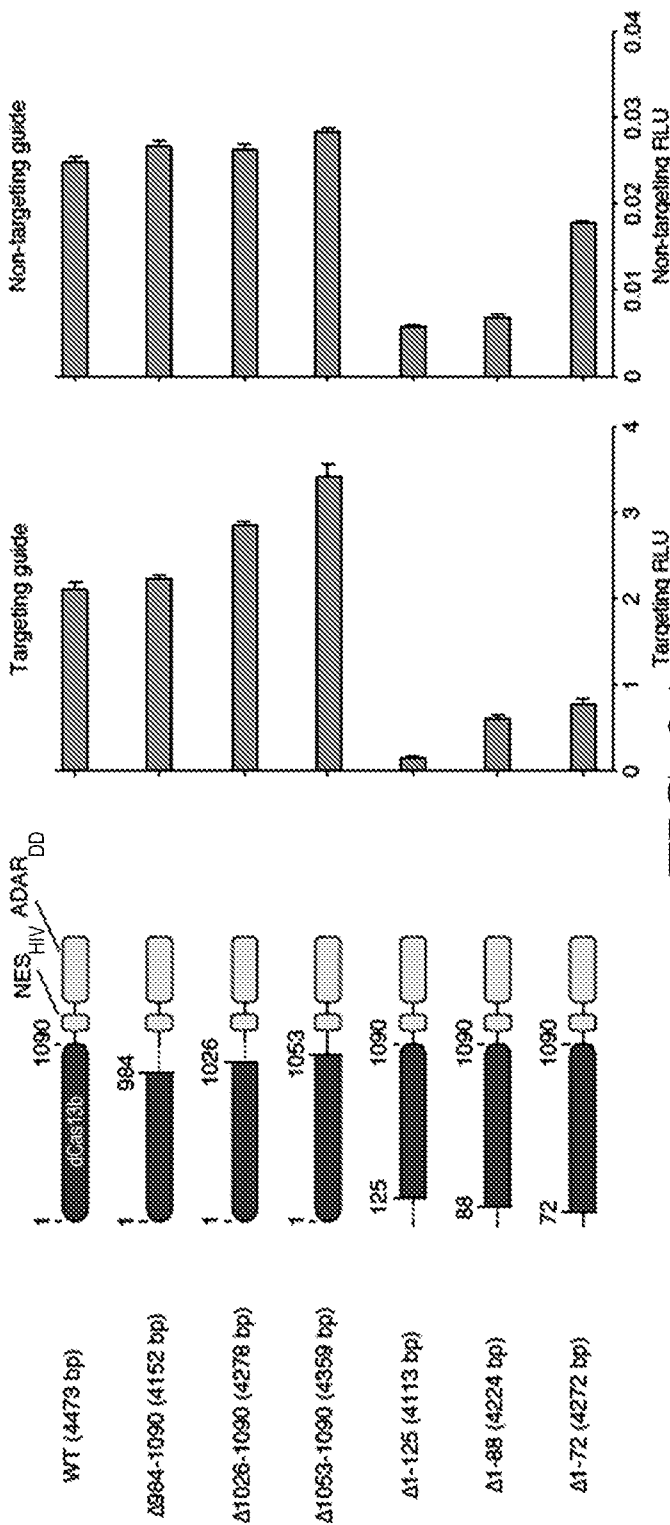


FIG. 9A

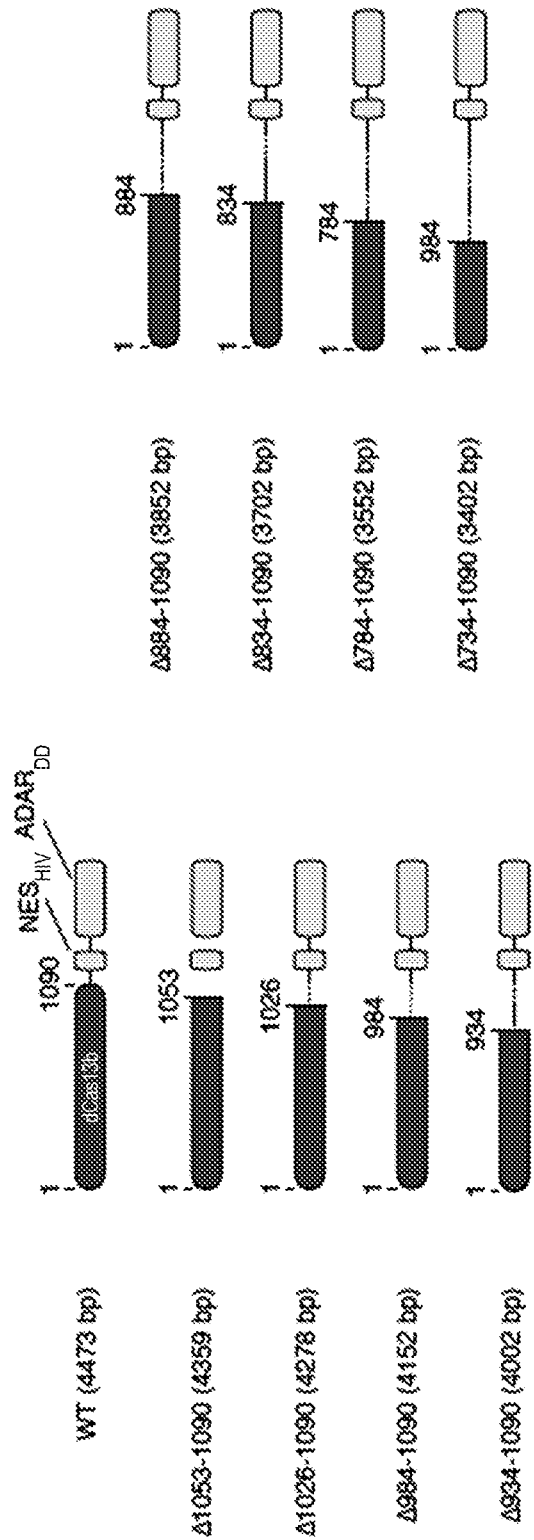


FIG. 9B

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2018/042529

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. forming part of the international application as filed:
 - in the form of an Annex C/ST.25 text file.
 - on paper or in the form of an image file.
 - b. furnished together with the international application under PCT Rule 3ter. 1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. furnished subsequent to the international filing date for the purposes of international search only:
 - in the form of an Annex C/ST.25 text file (Rule 3ter. 1(a)).
 - on paper or in the form of an image file (Rule 3ter. 1(b) and Administrative Instructions, Section 713).
2. In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US201 8/042529

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.: 5-18
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 18/042529

A. CLASSIFICATION OF SUBJECT MATTER**IPC(8) - C12N 9/22; C12N 15/09; C12N 15/11; C12N 15/85 (201 8.01)****CPC - C12N 9/22; C12N 15/09; C12N 15/1 1; C12N 15/85; C12N 231 0/20; C12N 231 0/351 9; C12Q 2561/1 07; G01 N 251 0/00 (201 8.08)****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

USPC - 435/325; 435/6.13; 435/69.1; 435/69.7; 536/23.1 (keyword delimited)

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.
Y	US 2009/0029370 A1 (BROUDE et al) 29 January 2009 (29.01.2009) entire document	1-4, 26-30
Y	US 2016/0208243 A1 (THE BROAD INSTITUTE, INC. et al) 21 July 2016 (21.07.2016) entire document	1-4, 19-30
Y	WEHR et al. "Monitoring regulated protein-protein interactions using split TEV," Nature Methods, 29 October 2006 (29.10.2006), Vol. 3, No. 12, Pgs. 985-993. entire document	2, 19, 23-25
Y	US 2003/0124629 A1 (TSE et al) 03 July 2003 (03.07.2003) entire document	20-24
Y	OBERST et al. "Inducible dimerization and inducible cleavage reveal a requirement for both processes in caspase-8 activation," J Biol Chem, 22 March 2010 (22.03.2010), Vol. 285, No. 22, Pgs. 16632-16642. entire document	24

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

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 17 October 2018	Date of mailing of the international search report 05 NOV 2018
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774