

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE,

(57) Abstract: Engineered or non-naturally occurring systems and compositions comprising novel Type V Cas polypeptides and orthologs thereof are disclosed herein. Also provided are methods of use for the novel Type V Cas polypeptide systems and compositions for reprogrammable targeting of nucleic acid and polynucleotide components.

SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,
GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished
upon receipt of that report (Rule 48.2(g))*
- *with sequence listing part of description (Rule 5.2(a))*

SMALL NOVEL CRISPR-CAS SYSTEMS AND METHODS OF USE THEREOF**CROSS-REFERENCE TO RELATED APPLICATIONS**

[0001] Reference is made to U.S. Provisional Application No. 63/287,884, filed December 9, 2021; the contents of which are incorporated by reference in their entireties herein.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No. HL141201 awarded by The National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] The contents of the electronic sequence listing ("BROD-5485WP_ST26.xml"; Size is 60,565 bytes and it was created on November 30, 2022) is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to systems, methods and compositions used for targeted gene modification and nucleic acid editing utilizing systems comprising Cas12b polypeptides. In particular, the present disclosure provides DNA or RNA-targeting compositions comprising novel DNA or RNA-targeting nucleases and at least one targeting nucleic acid component.

BACKGROUND

[0005] The ongoing exploration of CRISPR-Cas systems of bacterial adaptive immunity to bacteriophage (viral) infection has shown extreme diversity of protein composition and genomic architecture of these systems. Among the systems identified, it has been shown that size sometimes is a constraint to full deployment of these systems and that smaller, robust systems might confer some advantages. Thus, there exists a pressing need for alternative and robust systems of smaller size for targeting nucleic acids or polynucleotides.

[0006] Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present invention.

SUMMARY

[0007] In one aspect, the present disclosure provides a non-naturally or engineered composition comprising, (a) a Cas protein that comprises a RuvC-I, -II, -III domain but does not comprise a HNH domain and is less than 850 amino acids in size and (b) a guide sequence capable of forming a complex with the Cas protein and directing the complex to bind to a target sequence. In an embodiment, the guide comprises a scaffold sequence between about 170 nt and about 210 nt in length. In some embodiments the Cas protein is a Type V protein. In an embodiment, the Type V protein is a Cas12b protein. In some embodiments, the Cas12b protein is derived from *Phycisphaerae* bacterium ST-NAGAB-D1 or *Planctomycetes* bacterium RBG 134610. In an embodiment, the scaffold sequence is derived from *Phycisphaerae* bacterium ST-NAGAB-D1 or *Planctomycetes* bacterium RBG 134610.

[0008] In an embodiment, the complex binds a PAM sequence comprising YANTTN, where Y is C or T, and N is any nucleotide.

[0009] In an embodiment, the *Phycisphaerae* bacterium ST-NAGAB-D1 complex is stable and active between about 37°C to about 60°C.

[0010] In an embodiment, the present disclosure provides a vector system comprising one or more polynucleotide sequences encoding the Cas and scaffold sequence as disclosed above.

[0011] In an embodiment, the present disclosure provides a delivery system comprising any of the compositions disclosed above. In some embodiments, the delivery system comprises a ribonucleoprotein complex, one or more particles, one or more vesicles, or one or more liposomes, nanoparticles, exosomes, microvesicles, nucleic acid nanoassemblies, a gene gun, an implantable device.

[0012] In an embodiment, the present disclosure provides a host cell or progeny thereof comprising the composition as disclosed above.

[0013] In an embodiment, the present disclosure provides an *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising the compositions disclosed above.

[0014] In another aspect, the present disclosure provides a non-naturally occurring or engineered composition comprising (a) a Cas protein, wherein the Cas protein is catalytically inactive, (b) a nucleotide deaminase associated with or otherwise capable of forming a complex with the Cas protein, and (c) a guide sequence capable of forming a complex with the Cas protein and directing the complex to bind to a target sequence. In an embodiment, the Cas

protein is a Cas12b protein. In an embodiment, the nucleotide deaminase is an adenosine deaminase or a cytidine deaminase.

[0015] In an embodiment, the present disclosure provides a composition comprising one or more polynucleotides encoding one or more components of the composition.

[0016] In an embodiment, the present disclosure provides one or more vectors encoding one or more of the polynucleotides of the composition.

[0017] In an embodiment, the present disclosure provides a cell or progeny thereof genetically engineered to express one or more components of the composition disclosed above.

[0018] In another aspect, the present disclosure provides a method of editing nucleic acids in the target polynucleotides comprising delivering the composition, the one or more polynucleotides, or one or more vectors as disclosed above to a cell or population of cells comprising the target polynucleotides.

[0019] In an embodiment, the present disclosure provides a method wherein the target polynucleotides are target sequences within genomic DNA. In an embodiment, the target polynucleotide is edited at one or more bases to introduce a G→A or C→T mutation.

[0020] In another aspect, the present disclosure provides an isolated cell or progeny thereof comprising one or more base edits made using the method disclosed above.

[0021] In another aspect, the present disclosure provides an engineered, non-naturally occurring composition comprising (a) a catalytically dead Cas12b polypeptide, (b) a reverse transcriptase associated with or otherwise capable of forming a complex with the Cas12b polypeptide, and (c) a scaffold sequence between about 170 nt and about 210 nt in length capable of forming a complex with the Cas12b protein and directing site-specific binding of the complex to a target sequence of a target polynucleotide, the scaffold sequence further comprising a donor template encoding a donor sequence for insertion into the target polynucleotide.

[0022] In an embodiment, the present disclosure provides a composition comprising one or more polynucleotides encoding one or more components of the composition.

[0023] In another aspect, the present disclosure provides one or more vectors encoding the one or more polynucleotides.

[0024] In another aspect, the present disclosure provides a method of modifying target polynucleotides comprising, delivering the composition, the one or more polynucleotides, or the one or more vectors described above to a cell, or population of cells, comprising the target

polynucleotides, wherein the complex directs the reverse transcriptase to the target sequence and the reverse transcriptase facilitates insertion of a donor sequence encoded by the donor template from the scaffold sequence into the target polynucleotide.

[0025] In an embodiment, the present disclosure provides a method where insertion of the donor sequence (a) introduces one or more base edits, (b) corrects or introduces a premature stop codon, (c) disrupts a splice site, (d) inserts or restores a splice site, (e) inserts a gene or gene fragment at one or both alleles of the target polynucleotide, or (f) a combination thereof.

[0026] In another aspect, the present disclosure provides an isolated cell or progeny thereof comprising the modifications made using the methods disclosed above.

[0027] In another aspect, the present disclosure provides a method for programmable and targeted gene editing of a target sequence comprising delivery of the compositions disclosed above.

[0028] 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 example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] An 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 may be utilized, and the accompanying drawings of which:

[0030] **FIG. 1** – Illustrates the phylogenetic relationships of Cas12b orthologs. The genomic architecture of Cas12b from *Planctomycetes bacterium* RBG 13 46 10 and *Phycisphaera bacterium* ST-NAGAB-D1 are shown. The rectangle surrounds the accession numbers of the two Cas12b orthologs for *Planctomycetes bacterium* RBG 13 46 10 and *Phycisphaera bacterium* ST-NAGAB-D1.

[0031] **FIG. 2** – Illustrates the ancestral Cas12b loci for *Planctomycetes bacterium* RBG 13 46 10 and *Phycisphaera bacterium* ST-NAGAB-D1. The Cas12b amino acid lengths are 743 amino acids for *Planctomycetes bacterium* RBG 13 46 10 and 747 amino acids *Phycisphaera bacterium* ST-NAGAB-D1. The small RNA-seq data of the tracrRNA region and array for *Phycisphaera bacterium* ST-NAGAB-D1 Cas12b is shown.

[0032] **FIG. 3A-3B** – Illustrates the (3A) PAM depletion vs. rank for *Phycisphaera bacterium* ST-NAGAB-D1 Cas12b and a Weblogo at >0.9 depletion (43 PAMs) and >0.8 depletion (195 PAMs) and (3B) PAM depletion vs. rank for *Planctomycetes bacterium* RBG 13 46 10 Cas12b and a Weblogo graph at >0.9 depletion and >0.8 depletion. PAM preference is YANTTN for both species.

[0033] **FIG. 4A-4B** – shows that the (4A) *Phycisphaera bacterium* ST-NAGAB-D1 and (4B) *Planctomycetes bacterium* RBG 13 46 10 Cas12b systems are functional in an *E. coli* expression system.

[0034] **FIG. 5** – shows *in vitro* activity in different buffer systems of the RNP *Phycisphaera bacterium* ST-NAGAB-D1 complex with pulldown of the RNA complex. The expected products are at 320 bp and 250 bp.

[0035] **FIG. 6** – shows that the *Phycisphaera bacterium*/RNA binary complex is more temperature stable than the apo-protein (Cas12b only) based on fluorescence absorption at 580 nm.

[0036] **FIG. 7** – illustrates RNA-seq and mapping of *Phycisphaera bacterium* tracrRNA. A long tracrRNA is required for activity in *E. coli*.

[0037] **FIG. 8** – shows *Phycisphaera bacterium in vitro* loading and activity of Cas12b (C2c1~90 kDa; at left) and that cleavage activity (preference at 48°C) is dependent on Cas12b and a sgRNA consisting of a 194 nt sgRNA without spacer (middle). Sequencing revealed a TTTA PAM, a 25 nt protospacer and a potential 13 nt staggered overhang (at right).

[0038] **FIG. 9** – illustrates that *Phycisphaera bacterium* Cas12b lacks activity in HEK293T cells targeting the DNMT-1 and VEGFA loci.

[0039] **FIG. 10** – illustrates the *Planctomycetes bacterium* RBG 13 46 10 tracrRNA region with accompanying direct repeat and Fn Spacer regions (top). The *Phycisphaera bacterium* sgRNA is aligned with the *Planctomycetes bacterium* sgRNA (bottom) for comparison.

[0040] **FIG. 11A-11B** – illustrates *Planctomycetes bacterium in vitro* loading and activity. (11A) shows very weak *in vitro* activity for *Planctomycetes bacterium* Cas12b and also requires a 189 nt tracrRNA in addition to crRNA. (11B) shows no RNP pulldown at 48°C and slight pulldown activity at 37°C.

[0041] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0042] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.); PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.); Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0043] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0044] 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.

[0045] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0046] 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.

[0047] 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.

[0048] 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.

[0049] 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.

[0050] 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.

OVERVIEW

[0051] In one aspect, embodiments disclosed herein provide alternative, CRISPR-Cas systems comprising smaller Cas polypeptides and a nucleic acid component that functions as nucleic acid-guided re-programmable nuclease and use in methods of modifying target polynucleotides. In an embodiment, the Cas polypeptide may comprise a split RuvC nuclease domain, but does not comprise a HNH nuclease domain, and is less than 850 amino acids in size. The nucleic acid component is capable of forming a complex with the Cas polypeptide and directing sequence-specific binding of the complex to a target sequence on a target polynucleotide. In an embodiment, the nucleic acid component may be a single molecule comprising a guide component and a scaffold component. In another embodiment, the guide component and scaffold component may be on separate molecules which are capable of forming a complex with one another and the Cas polypeptide. The guide component comprises a guide sequence that directs sequence-specific binding to the target sequence. In an embodiment, certain example Cas polypeptides disclosed herein and having the above domain and size characteristics may be classified as Type V CRISPR-Cas system. In another embodiment, example Cas polypeptides disclosed herein and having the above domain and size characteristics may be classified as Cas12b CRISPR-Cas systems.

[0052] The Cas polypeptides disclosed herein, are considerably smaller than other known Cas polypeptides, e.g., Cas9, Cas12a and Cas12b. As such, they do not suffer from the delivery size limitations of other larger single-effector, RNA-guided nucleases, such as Type II and other, larger Type V CRISPR-Cas systems. Due to their smaller size, these Cas polypeptides may be easily combined with other functional domains, such as nucleobase deaminases, reverse transcriptases, transposases, ligases, topoisomerases, and serine and threonine recombinases (integrases) and still be packaged into conventional delivery systems, like certain adenoviral- and lentiviral-based viral vectors. Thus, among other improvements, the CRISPR-Cas systems

disclosed herein allows more flexible and effective strategies to manipulate and modify target polynucleotides.

[0053] In another aspect, embodiments disclosed herein are directed to plasmids and vectors encoding components of the CRISPR-Cas systems, as well as delivery systems for delivery of the CRISPR-Cas systems, or components thereof, to cells including eukaryotic cells.

[0054] In another aspect, embodiments disclosed herein are directed to isolated cells, and progeny thereof, modified to express said CRISPR-Cas systems and or engineered and modified using said CRISPR-Cas systems.

[0055] In another aspect, embodiments disclosed herein are direct to detection composition comprising said CRISPR-Cas systems, along with a detection construct, for use in detecting target polynucleotides in a sample. Such systems and methods may further include optional amplification reagents, such as isothermal amplification reagents, which may be used in combination with collateral activity of the CRISPR-Cas systems to detect target polynucleotides with high sensitivity.

CRISPR-CAS SYSTEMS AND COMPOSITIONS

[0056] Embodiments disclosed herein provide CRISPR-Cas systems comprising a Cas polypeptide and a guide molecule capable of forming a complex with the Cas polypeptide and directing sequence-specific binding of the complex to a target polynucleotide. The guide may comprise a reprogrammable element, a guide sequence, which may be designed to bind a target sequence in the target polynucleotide. Thus, the complex may be re-programmed to bind to different target polynucleotides by changing the composition of the guide sequence portion of the guide sequence. Embodiments disclosed herein further include compositions comprising such CRISPR-Cas systems.

Cas Polypeptides

[0057] The Cas polypeptides comprise a split RuvC endonuclease system less than 850 amino acids in size, but do not include a HNH or other endonuclease domain. In one example embodiment, the Cas polypeptides may have activity between about 37⁰C to about 60⁰C. In one example embodiment, the Cas genes encoding the Cas polypeptides may be located on the genome next to other Cas genes (e.g., cas1, cas2, cas4) or not be located near other Cas genes. In an embodiment, the Cas genes encoding the Cas polypeptides may be located in the genome next to or near a CRISPR array region or not be located next to or near a CRISPR array region.

In an example embodiment, the Cas polypeptide is a Type V polypeptide. In an embodiment, the Type V polypeptide is a Cas12b polypeptide.

RuvC Domain

[0058] The RuvC domain may comprise multiple subdomains, e.g., RuvC-I, RuvC-II and RuvC-III. The subdomains may be separated by interval sequences on the amino acid sequence of the protein.

[0059] Examples of RuvC domains include any polypeptides having a structural similarity and/or sequence similarity to a RuvC domain described in the art. For example, the RuvC domain may share a structural similarity and/or sequence similarity to a RuvC of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC domains.

[0001] In some examples, the RuvC domain comprise RuvC-I polypeptide, RuvC-II polypeptide, and RuvC-III polypeptide. Examples of the RuvC-I domain also include any polypeptides having a structural similarity and/or sequence similarity to a RuvC-I domain described in the art. For example, the RuvC-I domain may share a structural similarity and/or sequence similarity to a RuvC-I of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-I domain. The RuvC-II domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-II domain described in the art. For example, the RuvC-II domain may share a structural similarity and/or sequence similarity to a RuvC-II of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-II domains. The RuvC-III domain also include any polypeptides a structural similarity and/or sequence similarity to a RuvC-III domain described in the art. For example, the RuvC-III domains may share a structural similarity and/or sequence similarity to a RuvC-III of Cas9. In some examples, the RuvC domain may have an amino acid sequence that share at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 99%, or 100% sequence identity with RuvC-III domains.

[0060] For example, and as described in the art (e.g., Crystal structure of Cas9 in complex with guide RNA and target DNA, Nishimasu et al. Cell, 2014) the RuvC domain of Cas9 consists of a six-stranded mixed β -sheet (β 1, β 2, β 5, β 11, β 14 and β 17) flanked by α -helices (α 33, α 34 and α 39– α 45) and two additional two-stranded antiparallel β -sheets (β 3/ β 4 and β 15/ β 16). It has been described that the RuvC domain of Cas9 shares structural similarity with the retroviral integrase superfamily members characterized by an RNase H fold, such as *Escherichia coli* RuvC (PDB code 1HJR, 14% identity, root-mean-square deviation (rmsd) of 3.6 Å for 126 equivalent C α atoms) and *Thermus thermophilus* RuvC (PDB code 4LD0, 12% identity, rmsd of 3.4 Å for 131 equivalent C α atoms). *E. coli* RuvC is a 3-layer alpha-beta sandwich containing a 5-stranded beta-sheet sandwiched between 5 alpha-helices. RuvC nucleases have four catalytic residues (e.g., Asp7, Glu70, His143 and Asp146 in *T. thermophilus* RuvC), and cleave Holliday junctions (or structurally analogous cruciform junctions) through a two-metal mechanism. Asp10 (Ala), Glu762, His983 and Asp986 of the Cas9 RuvC domain are located at positions similar to those of the catalytic residues of *T. thermophilus* RuvC.

[0061] In an example embodiment, split Ruv-C domain of the Cas12b proteins do not have an HNH domain located between the Ruv-C II and Ruv-C III subdomains. For example, the Cas12b protein domain architecture is comprised of RuvC-I-II-III domains, a bridge domain (B), and a 3' terminal carboxyl (C) domain as shown in FIG. 1. The bridge domain is located between the RuvC-I and RuvC-II domains and the HNH domain is located between the RuvC-II and RuvC-III domains.

Bridge Helix

[0062] The nucleic-acid guided Cas12b polypeptide nuclease comprises a bridge helix (BH) domain. The bridge helix domain refers to a helix and arginine rich polypeptide. The bridge helix domain may be located next to anyone of the amino acid domains in the nucleic-acid guided nuclease. In one embodiment, the bridge helix domain is next to a RuvC domain, e.g., next to RuvC-I, RuvC-II, or RuvC-III subdomain. In one example, the bridge helix domain is between a RuvC-1 and RuvC2 subdomains.

[0063] The bridge helix domain may be from 10 to 100, from 20 to 60, from 30 to 50, e.g., 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46 or 47, 48, 49, or 50 amino acids in length. Examples of bridge helix includes the polypeptide of amino acids 60-93 of the sequence of *S. pyogenes* Cas9.

Example Cas Proteins and Orthologs

[0064] In a certain example embodiment, provided herein is the Cas12b polypeptide of *Phycisphaera bacterium* ST-NAGAB-D1, Accession No. AQT69685-1. In an embodiment, the Cas12b polypeptide of *Phycisphaera bacterium* ST-NAGAB-D1, Accession No. AQT69685-1 is 747 amino acids in length.

[0065] In a certain example embodiment, provided herein is the Cas12b polypeptide of *Planctomycetes bacterium* RBG-13-46-10, Accession No. OHB62175. In an embodiment, the Cas12b polypeptide of *Planctomycetes bacterium* RBG-13-46-10, Accession No. OHB62175 is 743 amino acids.

[0066] In another example embodiment, provided herein are Cas12b polypeptide orthologs as shown in Fig. 1, with their corresponding Accession Nos. provided therein.

Protein Modifications

[0067] The Cas12b polypeptide nucleases may comprise one or more modifications. As used herein, the term “modified” with regard to a Cas12b polypeptide nuclease generally refers to a Cas12b polypeptide nuclease having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type counterpart from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[0068] The modified proteins, e.g., modified Cas12b polypeptide nuclease may be catalytically inactive (also referred as dead). As used herein, a catalytically inactive or dead nuclease may have reduced or no nuclease activity compared to a wildtype counterpart nuclease. In some cases, a catalytically inactive or dead nuclease may have nickase activity. In some cases, a catalytically inactive or dead nuclease may not have nickase activity. Such a catalytically inactive or dead nuclease may not make either double-strand or single-strand break on a target polynucleotide, but may still bind or otherwise form complex with the target polynucleotide.

[0069] In an embodiment, the Cas12b comprises one or more mutation in the RuvC-II of the polypeptide. In an aspect, the mutation of a catalytic RuvC-II residue abolishes the nucleolytic activity on the non-target DNA strand. In an embodiment, mutation at the RuvC domain abolishes all nucleolytic activity, providing a dead Cas12b polypeptide (dCas12b).

[0070] In one embodiment, the modifications of the Cas12b polypeptide may or may not cause an altered functionality. By means of example, modifications which do not result in an altered functionality include for instance codon optimization for expression into a particular host, or providing the nuclease with a particular marker (e.g., for visualization). Modifications which may result in altered functionality may also include mutations, including point mutations, insertions, deletions, truncations (including split nucleases), etc., as well as chimeric nucleases (e.g., comprising domains from different orthologues or homologues) or fusion proteins. A chimeric enzyme can comprise a first fragment and a second fragment, and the fragments can be of Cas12b polypeptide nuclease orthologs of organisms of a genus or of a species, e.g., the fragments are from Cas12b polypeptide nuclease orthologs of different species. Fusion proteins may without limitation include, for instance, fusions with heterologous domains or functional domains (e.g., localization signals, catalytic domains, etc.). In an embodiment, various different modifications may be combined (e.g., a mutated nuclease which is catalytically inactive and which further is fused to a functional domain, such as for instance to induce DNA methylation or another nucleic acid modification, such as including without limitation, a break (e.g. by a different nuclease (domain)), a mutation, a deletion, an insertion, a replacement, a ligation, a digestion, a break or a recombination). As used herein, “altered functionality” includes without limitation an altered specificity (e.g., altered target recognition, increased (e.g., “enhanced” Cas12b polypeptide nuclease) or decreased specificity, or altered PAM recognition), altered activity (e.g., increased or decreased catalytic activity, including catalytically inactive nucleases or nickases), and/or altered stability (e.g., fusions with destabilization domains). Examples of all these modifications are known in the art. It will be understood that a “modified” nuclease as referred to herein, and in particular a “modified” Cas12b polypeptide nuclease or system or complex preferably still has the capacity to interact with or bind to the polynucleic acid (e.g., in complex with the guide sequence). Such modified Cas12b polypeptide nuclease can be combined with the deaminase protein or active domain thereof as described herein.

[0071] In one embodiment, unmodified Cas12b polypeptide nucleases may have cleavage activity. In one embodiment, the Cas12b polypeptide nucleases may direct cleavage of one or both DNA 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 one embodiment, the Cas12b polypeptide nucleases may direct cleavage of one

or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs or nucleotides from the first or last nucleotide of a target sequence. In one embodiment, the cleavage may be staggered, i.e., generating sticky ends. In one embodiment, the cleavage is a staggered cut with a 5' overhang. In one embodiment, the cleavage is a staggered cut with a 5' overhang of 1 to 15 nucleotides, preferably of 4 or 9 nucleotides.

[0072] In one embodiment, the cleavage site is distant from the Proto-Adjacent Motif (PAM), e.g., the cleavage occurs after the *n*th nucleotide on the non-target strand and after the nucleotide on the targeted strand. In one embodiment, the cleavage site occurs after an identified nucleotide (counted from the PAM) on the non-target strand and after the further identified nucleotide (counted from the PAM) on the targeted strand. In one embodiment, a vector encodes a nucleic acid-targeting effector protein that may be mutated with respect to a corresponding wild-type enzyme such that the mutated nucleic acid-targeting effector protein lacks the ability to cleave one or both DNA and RNA strands of a target polynucleotide containing a target sequence. As a further example, two or more catalytic domains of a Cas12b polypeptide nuclease (e.g., RuvC I, RuvC II, and RuvC III domain) may be mutated to produce a mutated Cas12b polypeptide nuclease substantially lacking all DNA cleavage activity. As described herein, corresponding catalytic domains of a Cas12b polypeptide nuclease may also be mutated to produce a mutated Cas12b polypeptide nuclease lacking all DNA cleavage activity or having substantially reduced DNA cleavage activity. In one embodiment, a Cas12b polypeptide nuclease may be considered to substantially lack all polynucleotide cleavage activity when the polynucleotide cleavage activity of the mutated enzyme is no more than 25%, no more than 10%, no more than 5%, no more than 1%, no more than 0.1%, no more than 0.01% 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. A Cas12b polypeptide nuclease 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 I, II, III, IV, V, or VI CRISPR systems.

[0073] PAM identification and specificity may be identified, for example, using the methods disclosed in the Examples section below.

[0074] In an embodiment, the RuvC nuclease domain of the Cas12b polypeptide is catalytically inactive, or modified to be catalytically inactive, or when the protein is a nickase.

In an embodiment, the nuclease domain is catalytically inactive, in which case the Cas12b is referred to as dCas12b.

[0075] In an embodiment, the Cas12b polypeptide nuclease may comprise one or more modifications resulting in enhanced activity and/or specificity, such as including mutating residues that stabilize the targeted or non-targeted strand. In an embodiment, the altered or modified activity of the engineered Cas12b polypeptide nuclease comprises increased targeting efficiency or decreased off-target binding. In an embodiment, the altered activity of the engineered Cas12b polypeptide nuclease comprises modified cleavage activity. In an embodiment, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In an embodiment, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In an embodiment, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In an embodiment, the altered or modified activity of the modified nuclease comprises altered helicase kinetics. In an embodiment, the modified nuclease comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA, or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered Cas12b polypeptide nuclease comprises a modification that alters formation of the Cas12b polypeptide nuclease and related complex. In an embodiment, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in an embodiment, there is increased specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In an embodiment, the mutations result in decreased off-target effects (e.g., cleavage or binding properties, activity, or kinetics), such as in case for Cas12b polypeptide nuclease for instance resulting in a lower tolerance for mismatches between target and the guide sequences. Other mutations may lead to increased off-target effects (e.g., cleavage or binding properties, activity, or kinetics). Other mutations may lead to increased or decreased on-target effects (e.g., cleavage or binding properties, activity, or kinetics). In an embodiment, the mutations result in altered (e.g., increased or decreased) helicase activity, association or formation of the functional nuclease complex. In an embodiment, the mutations result in an altered PAM recognition, i.e., a different PAM may be (in addition or in the alternative) be recognized, compared to the unmodified Cas12b polypeptide nuclease. Examples mutations include positively charged residues and/or

(evolutionary) conserved residues, such as conserved positively charged residues, in order to enhance specificity. In an embodiment, such residues may be mutated to uncharged residues, such as alanine.

Nuclear Localization Sequences

[0076] In one embodiment, the nucleic acid-guided nuclease is fused to one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In one embodiment, the Nucleic acid-guided nuclease comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g. zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus).

[0077] In one embodiment, the Cas12b polypeptide nuclease is fused to one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In one embodiment, the Cas12b polypeptide nuclease comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g. zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus).

[0078] When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy and/or in combination with one or more other NLSs present in one or more copies. In a preferred embodiment of the invention, the Nucleic acid-guided nuclease comprises at most 6 NLSs. In a preferred embodiment of the invention, the Cas12b polypeptide nuclease comprises at most 6 NLSs.

[0079] In one embodiment, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. 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 NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (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 PQPKKKPL (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 REKKKFLKRR (SEQ ID NO: 14) of the mouse Mx1 protein; the sequence KRKGDEV DGVDEVAKKKSKK (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.

[0080] In general, the one or more NLSs are of sufficient strength to drive accumulation of the nucleic acid-guided nuclease in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the nucleic acid-guided nuclease, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-guided nuclease, 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). 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 complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by complex formation and/or nucleic acid-guided nuclease activity), as compared to a control not exposed to the nucleic acid-guided nuclease or complex, or exposed to a nucleic acid-guided nuclease lacking the one or more NLSs. In an embodiment of the herein described nucleic acid-guided nuclease protein complexes and systems the codon optimized nucleic acid-guided nuclease proteins comprise an NLS attached to the C-terminal of the protein. In an embodiment, other localization tags may be fused to the nucleic acid-guided nuclease, such as without limitation for localizing the nucleic acid-guided nuclease 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.

[0081] In general, the one or more NLSs are of sufficient strength to drive accumulation of the Cas12b polypeptide nuclease in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the Cas12b polypeptide nuclease, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the Cas12b polypeptide nuclease, 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). 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 complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by complex formation and/or Cas12b polypeptide nuclease activity), as compared to a control not exposed to the Cas12b polypeptide nuclease or complex, or exposed to a Cas12b polypeptide nuclease lacking the one or more NLSs. In an embodiment of the herein described Cas12b polypeptide nuclease protein complexes and systems the codon optimized Cas12b polypeptide nuclease proteins comprise an NLS attached to the C-terminal of the protein. In an embodiment, other localization tags may be fused to the Cas12b polypeptide nuclease, such as without limitation for localizing the Cas12b polypeptide nuclease 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.

[0082] In an embodiment of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the Cas12b polypeptide nuclease. 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 Cas12b polypeptide nuclease can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. 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.

Linkers

[0083] In an embodiment of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the nucleic acid-guided nuclease or the Cas12b polypeptide. 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 nucleic acid-guided nuclease or Cas12b polypeptide nuclease can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. 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.

[0084] In some preferred embodiments, the functional domain is linked to a nucleic acid-guided nuclease (e.g., an active or a dead nucleic acid-guided nuclease) 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 nucleic acid-guided nuclease to such promoters or enhancers.

[0085] In some preferred embodiments, the functional domain is linked to a Cas12b polypeptide nuclease (e.g., an active or a dead Cas12b polypeptide nuclease) 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 Cas12b polypeptide nuclease to such promoters or enhancers.

[0086] The term “associated with” is used here in relation to the association of the functional domain to the Cas12b polypeptide nuclease protein, nucleic acid-guided nuclease, 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, between the Cas12b polypeptide nuclease protein and a functional domain, or between the nucleic acid guided

nuclease 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 one embodiment, the Cas12b polypeptide nuclease protein, nucleic acid-guided nuclease, or adaptor protein is associated with a functional domain by binding thereto. In other embodiments, the Cas12b polypeptide nuclease, nucleic acid-guided nuclease, or adaptor protein is associated with a functional domain because the two are fused together, optionally via an intermediate linker.

[0087] The term “linker” as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in an embodiment, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker.

[0088] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond).

[0089] In one embodiment, the linker is used to separate the Cas12b polypeptide nuclease and the nucleotide deaminase by a distance sufficient to ensure that each protein retains its required functional property. In one embodiment, the linker is used to separate the nucleic acid-guided nuclease and the nucleotide deaminase by a distance sufficient to ensure that each protein retains its required functional property.

[0090] Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In an embodiment, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric.

Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in one embodiment, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS (SEQ ID NO: 17) or GSG can be used. GGS, GSG, GGGS (SEQ ID NO: 17) or GGGGS (SEQ ID NO: 18) linkers can be used in repeats of 3 (such as (GGS)₃, (SEQ ID NO: 19) (GGGGS)₃ (SEQ ID NO: 20) or 5, 6, 7, 9 or even 12 or more, to provide suitable lengths. In some cases, the linker may be (GGGGS)₃₋₁₅. For example, in some cases, the linker may be (GGGGS)₃₋₁₁, e.g., GGGGS (SEQ ID NO: 18), (GGGGS)₂ (SEQ ID NO: 21), (GGGGS)₃ (SEQ ID NO: 20), (GGGGS)₄ (SEQ ID NO: 22), (GGGGS)₅ (SEQ ID NO: 23), (GGGGS)₆ (SEQ ID NO: 24), (GGGGS)₇ (SEQ ID NO: 25), (GGGGS)₈ (SEQ ID NO: 26), (GGGGS)₉ (SEQ ID NO: 27), (GGGGS)₁₀ (SEQ ID NO: 38), or (GGGGS)₁₁ (SEQ ID NO: 29).

[0091] In one embodiment, linkers such as (GGGGS)₃ (SEQ ID NO: 20) are preferably used herein. (GGGGS)₆ (SEQ ID NO: 24), (GGGGS)₉ (SEQ ID NO: 27) or (GGGGS)₁₂ (SEQ ID NO: 30) may preferably be used as alternatives. Other preferred alternatives are (GGGGS)₁ (SEQ ID NO: 18), (GGGGS)₂ (SEQ ID NO: 21), (GGGGS)₄ (SEQ ID NO: 22), (GGGGS)₅ (SEQ ID NO: 23), (GGGGS)₇ (SEQ ID NO: 25), (GGGGS)₈ (SEQ ID NO: 26), (GGGGS)₁₀ (SEQ ID NO: 38), or (GGGGS)₁₁ (SEQ ID NO: 29). In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTTR (SEQ ID NO: 31) is used as a linker. In yet an additional embodiment, the linker is an XTEN linker. In one embodiment, the Cas12b polypeptide nuclease or the nucleic acid-guided nuclease is linked to the deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTTR (SEQ ID NO: 31) linker. In further one embodiment, Cas12b polypeptide nuclease is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTTR (SEQ ID NO: 31) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID NO: 32)).

Table 2. Examples of linkers used in the invention are shown.

GGS	GGTGGTAGT
-----	-----------

GGs _x 3 (9) (SEQ ID NO: 19)	GGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO: 34)
GGs _x 7 (21) (SEQ ID NO: 33)	gggtggaggaggctctggtggaggcggtagcggaggcgagggtcgGGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO: 35)
XTEN	TCGGGATCTGAGACGCCTGGGACCTCGGAATCGGCTACGCCCCGAAAGT (SEQ ID NO: 36)
Z-EGFR_Short	Gtgataacaaattaacaaagaaatgtgggctggggaagaaattcgtaacctgccgaacctgaacggctggcagatgaccgcgtttattcgagcctggtgatgatccgagccagagcggaacctgctggcggaagcgaaaaactgaacgatgctgcaggcgccgaaaaccggcggtggttctggt (SEQ ID NO: 37)
GSAT	Ggtggttctgccggtggctccggttctggctccagcggtggcagctctggtgcgtccggcacgggtactgcgggtggcactggcagcggttccggtactggctctggc (SEQ ID NO: 38)

[0092] Linkers may be used between the guide sequences and the functional domain (activator or repressor), or between the Cas12b polypeptide nuclease and the functional domain. In an embodiment, linkers may be used between the guide molecules and the functional domain (e.g. activator or repressor), or between the Cas12b polypeptide and the functional domain. The linkers may be used to engineer appropriate amounts of “mechanical flexibility.”

[0093] In an embodiment, the one or more functional domains are controllable, e.g., inducible.

Nucleic Acid Component

[0094] The compositions herein may further comprise one or more nucleic acid guide components capable of forming a complex with the Cas polypeptide and directing site-specific binding of the complex to a target polynucleotide. The term “nucleic acid component” may be used interchangeably with the term “guide molecule” or “guide RNA.” The nucleic acid component may comprise a guide sequence and a scaffold sequence. When the spacer sequence and the scaffold are fused to form a single molecule, that single molecule may also be referred to as a “single-guide RNA or “sgRNA” for short. The guide sequence and/or the scaffold are modified such that they are not equivalent to the Cas12b’s naturally occurring crRNA. A guide sequence may form a complex with a nucleic acid-guided nuclease and direct the complex to bind with a target sequence. In some examples, the guide sequence may comprise a first and second nucleic acid molecules, the first and second nucleic acid molecules capable of forming a duplex, the duplex capable of forming a complex with the nucleic acid-guided nuclease, wherein the second nucleic acid molecule is a recombinant molecule comprising a heterologous guide sequence capable of directing site-specific binding of the complex to a target sequence

of a target polynucleotide. In some examples, the single guide sequence capable of forming a complex with the nucleic acid-guided nuclease and directing site-specific binding of the complex to a target sequence of a target polynucleotide.

[0095] As used herein, a heterologous guide sequence is a guide sequence that is not derived from the same species as the nucleic acid-guided nuclease. For example, a heterologous guide sequence of a nucleic acid-guided nuclease derived from species A is a polynucleotide derived from a species different from species A, or an artificial polynucleotide.

[0096] As used herein, the term “guide,” “guide sequence,” has the meaning as used herein elsewhere and comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In one embodiment, the degree of complementarity of the guide sequence to a given target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In certain example embodiments, the guide sequence comprises a guide sequence that may be designed to have at least one mismatch with the target sequence, such that a RNA duplex formed between the guide sequence and the target sequence. Accordingly, the degree of complementarity is less than 99%. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less. In one embodiment, the guide sequence is designed to have a stretch of two or more adjacent mismatching nucleotides, such that the degree of complementarity over the entire guide sequence is further reduced. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less, more particularly, about 92% or less, more particularly about 88% or less, more particularly about 84% or less, more particularly about 80% or less, more particularly about 76% or less, more particularly about 72% or less, depending on whether the stretch of two or more mismatching nucleotides encompasses 2, 3, 4, 5, 6 or 7 nucleotides, etc. In one embodiment, aside from the stretch of one or more mismatching nucleotides, 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-guided nuclease-guide system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0097] A guide sequence, and hence a nucleic acid-targeting guide, 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 one embodiment, 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 cytoplasmatic 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.

[0098] In an embodiment, the guide sequence or spacer length of the guide sequence is from 15 to 50 nt. In an embodiment, the spacer length of the guide sequence RNA at least 15

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

[0099] In one embodiment, the sequence of the guide sequence (direct repeat and/or spacer) is selected to reduce the degree secondary structure within the guide molecule. In one embodiment, 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 of a 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 44790.11.2022; Broad Reference BI-2013/004A); incorporated herein by reference.

[0100] In a particular embodiment, the guide sequence comprises a guide sequence is linked to a direct repeat (used interchangeably with “tracr mate”) sequence, wherein the direct repeat sequence comprises one or more stem loops or optimized secondary structures. In one embodiment, the direct repeat has a minimum length of 16 nts and a single stem loop. In further embodiments the direct repeat has a length longer than 16 nts, preferably more than 17 nts, and has more than one stem loops or optimized secondary structures. In one embodiment, the guide sequence comprises or consists of the spacer linked to all or part of the natural direct repeat sequence. In one embodiment, certain aspects of the direct repeat (or scaffold) can be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects

of the direct repeat or scaffold architecture are maintained. Preferred locations for engineered sequence modifications, including but not limited to insertions, deletions, and substitutions include sequence termini and regions of the direct repeat or scaffold sequence that are exposed when complexed with nucleic acid-guided nuclease and/or target, for example the tetraloop and/or loop2.

[0101] In one embodiment, a loop in the nucleic acid component 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.

[0102] In one embodiment, the nucleic acid component forms a stem loop with a separate non-covalently linked sequence, which can be DNA or RNA. In one embodiment, the sequences forming the nucleic acid component sequence 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 one embodiment, these sequences 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 this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the direct repeat sequence. 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.

[0103] In one embodiment, these stem-loop forming sequences can be chemically synthesized. In one embodiment, 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: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0104] The repeat:anti repeat duplex will be apparent from the secondary structure of the nucleic acid component. It may be typically a first complimentary stretch after (in 5' to 3' direction) the poly U tract and before the tetraloop; and a second complimentary stretch after (in 5' to 3' direction) the tetraloop and before the poly A tract. The first complimentary stretch (the "repeat") is complimentary to the second complimentary stretch (the "anti-repeat"). As such, they Watson-Crick base pair to form a duplex of dsRNA when folded back on one another. As such, the anti-repeat sequence is the complimentary sequence of the repeat and in terms to A-U or C-G base pairing, but also in terms of the fact that the anti-repeat is in the reverse orientation due to the tetraloop.

[0105] In an embodiment of the invention, modification of scaffold sequence architecture comprises replacing bases in stem loop 2. For example, In one embodiment, "actt" ("acuu" in RNA) and "aagt" ("aagu" in RNA) bases in stemloop2 are replaced with "cgcc" and "gcgg". In one embodiment, "actt" and "aagt" bases in stemloop2 are replaced with complimentary GC-rich regions of 4 nucleotides. In one embodiment, the complimentary GC-rich regions of 4 nucleotides are "cgcc" and "gcgg" (both in 5' to 3' direction). In one embodiment, the complimentary GC-rich regions of 4 nucleotides are "gcgg" and "cgcc" (both in 5' to 3' direction). Other combination of C and G in the complimentary GC-rich regions of 4 nucleotides will be apparent including CCCC and GGGG.

[0106] In one aspect, the stem loop 2, e.g., "ACTTgtttAAGT" (SEQ ID NO: 39) can be replaced by any "XXXXgtttYYYY", e.g., where XXXX and YYYY represent any complementary sets of nucleotides that together will base pair to each other to create a stem.

[0107] In one aspect, the stem comprises at least about 4 bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-12 and Y2-12 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "gttt," will form a complete hairpin in the

overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y base pairing sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire nucleic acid component is preserved. In one aspect, the stem can be a form of X:Y base pairing that does not disrupt the secondary structure of the whole nucleic acid component in that it has a DR:tracr duplex, and 3 stem loops. In one aspect, the "gttt" tetraloop that connects ACTT and AAGT (or any alternative stem made of X:Y base pairs) can be any sequence of the same length (e.g., 4 base pair) or longer that does not interrupt the overall secondary structure of the sgRNA. In one aspect, the stem loop can be something that further lengthens stemloop2, e.g. can be MS2 aptamer. In one aspect, the stemloop3 "GGCACCGagtCGGTGC" (SEQ ID NO: 40) can likewise take on a "XXXXXXXXagtYYYYYYY" form, e.g., wherein X7 and Y7 represent any complementary sets of nucleotides that together will base pair to each other to create a stem. In one aspect, the stem comprises about 7bp comprising complementary X and Y sequences, although stems of more or fewer base pairs are also contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "agt", will form a complete hairpin in the overall secondary structure. In one aspect, any complementary X:Y base pairing sequence is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y base pairing that doesn't disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stem loops. In one aspect, the "agt" sequence of the stem loop 3 can be extended or be replaced by an aptamer, e.g., a MS2 aptamer or sequence that otherwise generally preserves the architecture of stemloop3. In one aspect for alternative Stem loops 2 and/or 3, each X and Y pair can refer to any base pair. In one aspect, non-Watson Crick base pairing is contemplated, where such pairing otherwise generally preserves the architecture of the stem loop at that position.

[0108] In one aspect, the DR:tracrRNA duplex can be replaced with the form: gYYYYag(N)NNNNxxxxNNNN(AAN)uuRRRRu (using standard IUPAC nomenclature for nucleotides), wherein (N) and (AAN) represent part of the bulge in the duplex, and "xxxx" represents a linker sequence. NNNN on the direct repeat can be anything so long as it basepairs with the corresponding NNNN portion of the tracrRNA. In one aspect, the DR:tracrRNA duplex can be connected by a linker of any length, any base composition, as long as it does not alter the overall structure.

[0109] In one embodiment, the natural hairpin or stem loop structure of the or scaffold sequence is extended or replaced by an extended stem loop. Extension of the stem can enhance the assembly of the nucleic acid component with the nucleic acid-guided nuclease. In one embodiment the stem of the stem loop is extended by at least 1, 2, 3, 4, 5 or more complementary basepairs (i.e. corresponding to the addition of 2, 4, 6, 8, 10 or more nucleotides in the scaffold sequence). In one embodiment these are located at the end of the stem, adjacent to the loop of the stem loop.

[0110] In one embodiment, the susceptibility of the sequence to RNases or to decreased expression can be reduced by slight modifications of the sequence of the sequence which do not affect its function. For instance, in one embodiment, premature termination of transcription, such as premature transcription of U6 Pol-III, can be removed by modifying a putative Pol-III terminator (4 consecutive U's) in the guide or scaffold sequence. Where such sequence modification is required in the stem loop of the guide or scaffold sequence, it is preferably ensured by a base pair flip.

[0111] In one embodiment, the nucleic acid-guided nuclease may need a tracr sequence. The "tracrRNA" or "scaffold" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In one embodiment, the degree of complementarity between the tracrRNA sequence and crRNA sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In one embodiment, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In one embodiment, the tracr sequence and guide sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop may correspond to the tracr mate sequence, and the portion of the sequence 3' of the loop then corresponds to the tracr sequence. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop may alternatively correspond to the tracr sequence, and the portion of the sequence 3' of the loop corresponds to the tracr mate sequence.

[0112] In one embodiment, the tracr and tracr mate sequences can be chemically synthesized. In one embodiment, 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: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0113] In one embodiment, the tracr and tracr mate sequences 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.

[0114] In one embodiment, the tracr and tracr mate sequences can be covalently linked using click chemistry. In one embodiment, the tracr and tracr mate sequences can be covalently linked using a triazole linker. In one embodiment, the tracr and tracr mate sequences 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 one embodiment, the tracr and tracr mate sequences are covalently linked by ligating a 5'-hexyne tracrRNA and a 3'-azide crRNA. In one embodiment, either or both of the 5'-hexyne tracrRNA and a 3'-azide crRNA 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: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[0115] In one embodiment, the tracr and tracr mate sequences 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)), polyphosphodiester, 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.

[0116] The linker (e.g., a non-nucleotide loop) can be of any length. In one embodiment, the linker has a length equivalent to about 0-16 nucleotides. In one embodiment, the linker has a length equivalent to about 0-8 nucleotides. In one embodiment, the linker has a length equivalent to about 0-4 nucleotides. In one embodiment, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in International Patent Publication No. WO 2011/008730.

[0117] In an embodiment, the tracrRNA, guide sequence, and tracr mate sequence may reside in a single guide RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation or alternatively arranged in a 3' to 5' orientation), or the tracrRNA may be a different RNA than the RNA containing the guide sequence and tracr mate sequence. In these embodiments, the tracrRNA hybridizes to the tracr mate sequence and the guide sequence directs the nucleic acid-guided nuclease-guide molecule complex to the target sequence. In some examples, the nucleic acid component comprises (in 5' to 3' direction): a guide sequence, a poly U tract, a first complementary stretch (the "repeat"), a loop (tetraloop), a second complementary stretch (the "anti-repeat" being complementary to the repeat), a stem, and further stem loops and stems and a poly A (often poly U in RNA) tail (terminator).

Guide Chemical Modifications

[0118] In an embodiment, the molecule comprises non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Preferably, these non-naturally occurring nucleic acids and non-naturally occurring nucleotides are located outside the guide sequence. 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 sequence nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide

sequence comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide sequence comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate 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, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine. Examples of ω 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 ω RNA can comprise increased stability and increased activity as compared to unmodified guide sequence, 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 Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066). In one embodiment, the 5' and/or 3' end of a guide sequence 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 an embodiment, a ω RNA comprises ribonucleotides in a region that binds to a target sequence and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to the Cas12b polypeptide nuclease. In an embodiment, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide sequence structures. In one embodiment, 3-5 nucleotides at either the 3' or the 5' end of a guide sequence is chemically modified. In one embodiment, only minor modifications are introduced in the seed region, such as 2'-F modifications. In one embodiment, 2'-F modification is introduced at the 3' end of a guide sequence. In an embodiment, three to five nucleotides at the 5' and/or the 3' end of the guide sequence 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 an embodiment, all of the phosphodiester bonds of a guide sequence are

substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In an embodiment, more than five nucleotides at the 5' and/or the 3' end of the guide sequence are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified guide sequence can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E7110-E7111). In an embodiment of the invention, a guide sequence 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 sequence by a linker, such as an alkyl chain. In an embodiment, the chemical moiety of the modified guide sequence can be used to attach the guide sequence to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide sequence can be used to identify or enrich cells generically edited by a Cas12b polypeptide nuclease and related systems (see Lee et al., eLife, 2017, 6:e25312, DOI:10.7554).

[0119] In a particular embodiment, the conserved nucleotide sequence may be modified to comprise one or more protein-binding RNA aptamers. In a particular 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 as detailed further herein.

[0120] In embodiments, the Cas12b polypeptide utilizes a guide sequence comprising a polynucleotide sequence that facilitates the interaction with the Cas12b protein, allowing for sequence specific binding and/or targeting of the guide sequence with the target polynucleotide. Chemical synthesis of the guide sequence is contemplated, using covalent linkage 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; chemical synthesis using automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., *J. Am. Chem. Soc.* (1998) 120: 11820-11821; Scaringe, *Methods Enzymol.* (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., *J. Am. Chem. Soc.* (2011) 133: 11540-11546; Hendel et al., *Nat. Biotechnol.* (2015) 33:985-989).

[0121] In certain example embodiments, the guide sequence and scaffold may be designed as two separate molecules that can hybridize or covalently joined into a single molecule.

Covalent linkage can be 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)), polyphosphodiester, 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.

[0122] The linker (e.g., a non-nucleotide loop) can be of any length. In one embodiment, the linker has a length equivalent to about 0-16 nucleotides. In one embodiment, the linker has a length equivalent to about 0-8 nucleotides. In one embodiment, the linker has a length equivalent to about 0-4 nucleotides. In one embodiment, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in International Patent Publication No. WO 2011/008730.

Escorted Guides

[0123] In one embodiment, the compositions or complexes have a nucleic acid component with a functional structure designed to improve guide sequence structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0124] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically

relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." *Nature Reviews Drug Discovery* 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." *Trends in biotechnology* 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." *J Clin Invest* 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a cue by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." *Science* 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." *Silence* 1.1 (2010): 4).

[0125] Accordingly, in one embodiment, the nucleic acid component is modified, e.g., by one or more aptamer(s) designed to improve nucleic acid component delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the nucleic acid component deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends a nucleic acid component that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O₂ concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0126] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Cryptochrome-2 activation is also highly sensitive, allowing for the use of low light intensity

stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0127] Energy sources such as electromagnetic radiation, sound energy or thermal energy may induce the nucleic acid component. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm². In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0128] The chemical or energy sensitive nucleic acid component may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide molecule and have the nucleic acid-guided nuclease system or complex function. The invention can involve applying the chemical source or energy so as to have the guide molecule function and the nucleic acid-guided nuclease system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0129] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Absciscic Acid (ABA) (see, e.g., stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0130] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0131] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel-based system inducible by energy, heat or radio-wave (see, e.g., www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide molecule and the other components of the nucleic acid-guided nuclease/ guide molecule complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the nucleic acid-guided nuclease/ guide molecule complex will be active and modulating target gene expression in cells.

[0132] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for *in vivo* applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0133] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under *in vivo* conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μ s and 500 milliseconds, preferably between 1 μ s and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0134] As used herein, ‘electric field energy’ is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under *in vivo* conditions (see WO97/49450).

[0135] As used herein, the term “electric field” includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc., as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0136] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0137] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0138] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 μ s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0139] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vivo* conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0140] Preferably, the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term “pulse” includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0141] Preferably, the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0142] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0143] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm² to about 100 W/cm². Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0144] As used herein, the term "ultrasound" refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0145] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm² (FDA recommendation), although energy densities of up to 750 mW/cm² have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm² (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm² (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[0146] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0147] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will

appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[0148] Preferably, the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm⁻². Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm⁻².

[0149] Preferably, the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0150] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0151] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm⁻² to about 10 Wcm⁻² with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm⁻², but for reduced periods of time, for example, 1000 Wcm⁻² for periods in the millisecond range or less.

[0152] Preferably, the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0153] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm⁻² or 1.25 Wcm⁻² as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0154] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues

unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[0155] In one embodiment, the nucleic acid molecule is modified by a secondary structure to increase the specificity of the nucleic acid-guided nuclease and related system and the secondary structure can protect against exonuclease activity and allow for 5' additions to the guide sequence also referred to herein as a protected guide molecule.

[0156] In one aspect, the invention provides for hybridizing a "protector RNA" to a sequence of the nucleic acid component wherein the "protector RNA" is an RNA strand complementary to the 3' end of the guide molecule to thereby generate a partially double-stranded guide molecule sequence. In an embodiment of the invention, protecting mismatched bases (i.e., the bases of the guide molecule which do not form part of the guide sequence) with a perfectly complementary protector sequence decreases the likelihood of target DNA binding to the mismatched basepairs at the 3' end. In one embodiment of the invention, additional sequences comprising an extended length may also be present within the sequence such that the guide or scaffold sequence comprises a protector sequence within the guide or scaffold sequence. This "protector sequence" ensures that the guide molecule comprises a "protected sequence" in addition to an "exposed sequence" (comprising the part of the guide sequence hybridizing to the target sequence). In one embodiment, the guide molecule is modified by the presence of the protector guide to comprise a secondary structure such as a hairpin. Advantageously there are three or four to thirty or more, e.g., about 10 or more, contiguous base pairs having complementarity to the protected sequence, the guide sequence or both. It is advantageous that the protected portion does not impede thermodynamics of the nucleic acid-guided nuclease and related system interacting with its target. By providing such an extension including a partially double stranded guide sequence, the guide sequence is considered protected and results in improved specific binding of the nucleic acid-guided nuclease/ guide sequence complex, while maintaining specific activity.

[0157] In one embodiment, use is made of a truncated guide sequence (tru- guide sequence), i.e. a guide sequence which comprises a guide sequence which is truncated in length with respect to the canonical guide sequence length. As described by Nowak et al. (Nucleic Acids Res (2016) 44 (20): 9555-9564), such guides may allow catalytically active nucleic acid-guided nuclease to bind its target without cleaving the target DNA. In one embodiment, a truncated guide sequence is used which allows the binding of the target but retains only nickase activity of the nucleic acid-guided nuclease .

[0158] In one embodiment, conjugation of triantennary N-acetyl galactosamine (GalNAc) to oligonucleotide components may be used to improve delivery, for example delivery to select cell types, for example hepatocytes (see International Patent Publication No. WO 2014/118272 incorporated herein by reference; Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961). This is considered to be a sugar-based particle and further details on other particle delivery systems and/or formulations are provided herein. GalNAc can therefore be considered to be a particle in the sense of the other particles described herein, such that general uses and other considerations, for instance delivery of said particles, apply to GalNAc particles as well. A solution-phase conjugation strategy may for example be used to attach triantennary GalNAc clusters (mol. wt. ~2000) activated as PFP (pentafluorophenyl) esters onto 5'-hexylamino modified oligonucleotides (5'-HA ASOs, mol. wt. ~8000 Da; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455). Similarly, poly(acrylate) polymers have been described for *in vivo* nucleic acid delivery (see WO2013158141 incorporated herein by reference). In further alternative embodiments, pre-mixing nucleic acid-guided nuclease nanoparticles (or protein complexes) with naturally occurring serum proteins may be used in order to improve delivery (Akinc A et al, 2010, Molecular Therapy vol. 18 no. 7, 1357–1364).

[0159] Screening techniques are available to identify delivery enhancers, for example by screening chemical libraries (Gilleron J. et al., 2015, Nucl. Acids Res. 43 (16): 7984-8001). Approaches have also been described for assessing the efficiency of delivery vehicles, such as lipid nanoparticles, which may be employed to identify effective delivery vehicles for components (see Sahay G. et al., 2013, Nature Biotechnology 31, 653–658).

PAM specificity

[0160] In one example embodiment, the Cas12b lack or substantially lack a PAM interacting (PI) domain. In an embodiment, the Cas12b may have a PI domain or a functional

fragment of a PI domain. In an embodiment, the Cas12b may achieve a target specificity by a non-protein domain. In an embodiment, the nucleic acid-guided nucleases may have helicase activity. In an embodiment, the nucleic acid-guided nucleases may have reduced helicase activity compared to Cas proteins known in the art. In an embodiment, the nucleic acid-guided nucleases may comprise additional components that contribute in mediating target recognition. In an embodiment, targeting specificity is obtained by a central hairpin structure in a guide molecule.

[0161] Examples of PAM sequences for the Cas12b herein include the 5'-YANTTN-3' where Y is T or C and N is any nucleotide. For example, the nucleic acid-guided nucleases may recognize PAM sequence TAATTA or CAATTA, etc.

[0162] The PAM interaction domain or PI domain as referred to herein is reported to be responsible for determining PAM specificity of Cas12b. By means of example, the PI domain is contained in the NUC lobe and forms an elongated structure comprising seven α -helices, a three-stranded antiparallel β -sheet, a five-stranded antiparallel β -sheet, and a two-stranded antiparallel β -sheet.

[0163] In some cases, where the nucleic acid-guided nucleases do have a PAM requirement, the precise sequence and length requirements for the PAM will differ depending on the nucleic acid-guided nucleases used. In some examples, PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different nucleic acid-guided nucleases orthologs have been identified and the skilled person will be able to identify further PAM sequences for use with a given nucleic acid-guided nucleases.

[0164] Further, associating a PAM Interacting (PI) domain (e.g., attaching or fusing) to a nucleic acid-guided nuclease may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas12b, genome engineering platform. nucleic acid-guided nucleases may be engineered to alter their PAM specificity, for example as described in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. The skilled person will understand that other Cas12b proteins may be modified analogously.

[0165] The crystal structure information (described in U.S. Provisional Patent Application Nos. 61/915,251 filed December 12, 2013, 61/930,214 filed on January 22, 2014, 61/980,012 filed April 15, 2014; and Nishimasu et al, "Crystal Structure of Cas9 in Complex with Guide

RNA and Target DNA,” Cell 156(5):935–949, DOI: [dx.doi.org/10.1016/j.cell.2014.02.001](https://doi.org/10.1016/j.cell.2014.02.001) (2014), each and all of which are incorporated herein by reference) provides structural information to truncate and create modular or multi-part CRISPR enzymes which may be incorporated into inducible composition. In particular, structural information is provided for *S. pyogenes* Cas9 (SpCas9), and this may be extrapolated to other Cas9 orthologs or Cas12b proteins (as well as homologs and orthologs thereof) or other nucleic acid-guided nucleases. In one embodiment, the conformational variations in the crystal structures of the CRISPR-Cas9 system or of components of the CRISPR-Cas9 provide important and critical information about the flexibility or movement of protein structure regions relative to nucleotide (RNA or DNA) structure regions that may be important for the function of other nucleic acid-guided nucleases and related systems. The structural information provided for Cas9 (e.g. *S. pyogenes* Cas9) as the nucleic acid-guided nuclease in the present application may be used to further engineer and optimize the other nucleic acid-guided nucleases and related system and this may be extrapolated to interrogate structure-function relationships in other nucleic acid-guided nucleases and related systems.

HDR Donor Templates

[0166] In one embodiment, the compositions and systems herein may further comprise one or more nucleic acid templates. In some cases, the nucleic acid template may comprise one or more polynucleotides. In certain cases, the nucleic acid template may comprise coding sequences for one or more polynucleotides. The nucleic acid template may be a DNA template.

[0167] The donor polynucleotide may be used for editing the target polynucleotide. In some cases, the donor polynucleotide comprises one or more mutations to be introduced into the target polynucleotide. Examples of such mutations include substitutions, deletions, insertions, or a combination thereof. The mutations may cause a shift in an open reading frame on the target polynucleotide. In some cases, the donor polynucleotide alters a stop codon in the target polynucleotide. For example, the donor polynucleotide may correct a premature stop codon. The correction may be achieved by deleting the stop codon or introduces one or more mutations to the stop codon. In other example embodiments, the donor polynucleotide addresses loss of function mutations, deletions, or translocations that may occur, for example, in certain disease contexts by inserting or restoring a functional copy of a gene, or functional fragment thereof, or a functional regulatory sequence or functional fragment of a regulatory sequence. A functional fragment refers to less than the entire copy of a gene by providing

sufficient nucleotide sequence to restore the functionality of a wild type gene or non-coding regulatory sequence (e.g. sequences encoding long non-coding RNA). In certain example embodiments, the systems disclosed herein may be used to replace a single allele of a defective gene or defective fragment thereof. In another example embodiment, the systems disclosed herein may be used to replace both alleles of a defective gene or defective gene fragment. A “defective gene” or “defective gene fragment” is a gene or portion of a gene that when expressed fails to generate a functioning protein or non-coding RNA with functionality of the corresponding wild-type gene. In certain example embodiments, these defective genes may be associated with one or more disease phenotypes. In certain example embodiments, the defective gene or gene fragment is not replaced but the systems described herein are used to insert donor polynucleotides that encode gene or gene fragments that compensate for or override defective gene expression such that cell phenotypes associated with defective gene expression are eliminated or changed to a different or desired cellular phenotype.

[0168] In an embodiment of the invention, the donor polynucleotide may include, but not be limited to, genes or gene fragments, encoding proteins or RNA transcripts to be expressed, regulatory elements, repair templates, and the like. According to the invention, the donor polynucleotides may comprise left end and right end sequence elements that function with transposition components that mediate insertion.

[0169] In certain cases, the donor polynucleotide manipulates a splicing site on the target polynucleotide. In some examples, the donor polynucleotide disrupts a splicing site. The disruption may be achieved by inserting the polynucleotide to a splicing site and/or introducing one or more mutations to the splicing site. In certain examples, the donor polynucleotide may restore a splicing site. For example, the polynucleotide may comprise a splicing site sequence.

[0170] The donor polynucleotide to be inserted may have a size from 10 base pair or nucleotides to 50 kb in length, e.g., from 50 to 40k, from 100 and 30 k, from 100 to 10000, from 100 to 300, from 200 to 400, from 300 to 500, from 400 to 600, from 500 to 700, from 600 to 800, from 700 to 900, from 800 to 1000, from 900 to 1100, from 1000 to 1200, from 1100 to 1300, from 1200 to 1400, from 1300 to 1500, from 1400 to 1600, from 1500 to 1700, from 600 to 1800, from 1700 to 1900, from 1800 to 2000 base pairs (bp) or nucleotides in length.

APPLICATIONS AND USES IN GENERAL

[0171] The 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.

[0172] Aspects of the invention thus also encompass methods and uses of the compositions and systems described herein in genome engineering, e.g. for altering or manipulating the 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*. In some examples, the target polynucleotides are target sequences within genomic DNA, including nuclear genomic DNA, mitochondrial DNA, or chloroplast DNA.

[0173] Typically, in the context of a nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide sequence 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, 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, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest).

[0174] In one embodiment, the present disclosure provides a method of targeting a polynucleotide, comprising contacting a sample (such as cell, population of cells, tissue, organ, or an organism) that comprises a target polynucleotide with the composition, systems, polynucleotide(s), or vector(s). The contacting may result in modification of a gene product or modification of the amount or expression of a gene product. In some examples, the target sequence of the polynucleotide is a disease-associated target sequence.

[0175] In one embodiment, the present disclosure provides a method of modifying target polynucleotides comprising delivering the composition, the one or more polynucleotides of 2, or one or more vectors to a cell or population of cells comprising the target polynucleotides, wherein the complex directs the reverse transcriptase to the target sequence and the reverse transcriptase facilitates insertion of the donor sequence from the guide RNA into the target polynucleotide.

[0176] 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.

[0177] The target polynucleotide of a 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 (proto-adjacent motif); that is, a short sequence recognized by the complex. The precise sequence and length requirements for the PAM differ depending on the Cas12b polypeptide used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). A skilled person will be able to identify further PAM sequences for use with a given Cas12b polypeptide. Further, engineering of the PAM Interacting domain may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas12b polypeptide nuclease, genome engineering platform. Cas12b polypeptide may be engineered to alter their PAM specificity.

[0178] 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.

[0179] Aspects of the invention relate to a method of targeting a polynucleotide, comprising contacting a sample that comprises the polynucleotide with a composition, system or Cas12b polypeptide as described in any embodiment herein, a delivery system comprising a composition, system or Cas12b polypeptide as described in any embodiment herein, a polynucleotide comprising a composition, system or Cas12b as described in any embodiment herein, a vector comprising a composition, system or Cas12b polypeptide as described in any embodiment herein, or a vector system comprising a composition, system or Cas12b polypeptide as described in any embodiment herein. In an embodiment, a target polynucleotide is contacted with at least two different composition, system or Cas12b polypeptide. In further embodiments, the two different Cas12b polypeptide nuclease have different target polynucleotide specificities, or degrees of specificity. In an embodiment, the two different Cas12b polypeptide have a different PAM specificity.

[0180] Also envisaged are methods of targeting a polynucleotide, comprising contacting a sample that comprises the polynucleotide with the composition and systems, vectors, polynucleotides, herein wherein contacting results in modification of a gene product or modification of the amount or expression of a gene product. In an embodiment, the expression of the targeted gene product is increased by the method. In an embodiment, the expression of the targeted gene product is increased by at least 10%, at least 15%, at least 20%, at least 25%, at least 30%, at least 35%, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, 100%. In an embodiment, the expression of the targeted gene product is increased at least 1.5-fold, at least 2-fold, at least 2.5-fold, at least 3-fold, at least 3.5-fold, at least 3.5-fold, at least 4-fold, at least 4.5-fold, at least 5-fold, at least 10-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 50-fold, at least 100-fold. In an embodiment, the expression

of the targeted gene product is reduced by at least 10%, by at least 15%, by at least 20%, by at least 25%, by at least 30%, by at least 35%, by at least 40%, by at least 45%, by at least 50%, by at least 55%, by at least 60%, by at least 65%, by at least 70%, by at least 75%, by at least 80%, by at least 85%, by at least 90%, by at least 95%, by at least 100%. In an embodiment, the expression of the targeted gene product is reduced at least 1.5-fold, at least 2-fold, at least 2.5-fold, at least 3-fold, at least 3.5-fold, at least 3.5-fold, at least 4-fold, at least 4.5-fold, at least 5-fold, at least 10-fold, at least 10-fold, at least 15-fold, at least 20-fold, at least 25-fold, at least 50-fold, at least 100-fold. In alternative embodiments, the expression of the targeted gene product is reduced by the method. In further embodiments, expression of the targeted gene may be completely eliminated, or may be considered eliminated as remnant expression levels of the targeted gene fall below the detection limit of methods known in the art that are used to quantify, detect, or monitor expression levels of genes.

[0181] In one embodiment, one or more polynucleotide molecules, vectors, or vector systems driving expression of one or more elements of a nucleic acid-targeting system or delivery systems comprising one or more elements of the 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. In an embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0182] In one embodiment, the host cell is a cell of a cell line. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In one embodiment, 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 one embodiment, a cell transiently transfected with the components of a composition or 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 complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In one embodiment, 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.

[0183] Further intended are isolated human cells or tissues, plants or non-human animals comprising one or more of the polynucleotide molecules, vectors, vector systems, or cells

described in any of the embodiments herein. In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including (isolated) stem cells, and progeny thereof.

[0184] In an embodiment, the plants or non-human animals comprise at least one of the compositions, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein at least one tissue type of the plant or non-human animal. In certain embodiment, non-human animals comprise at least one of the compositions, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein in at least one tissue type. In an embodiment, the presence of the compositions is transient, in that they are degraded over time. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is limited to certain tissue types or regions in the plant or non-human animal. In an embodiment, the expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent of a physiological cue. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells may be triggered by an exogenous molecule. In an embodiment, expression of the compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent on the expression of a non-Cas molecule in the plant or non-human animal.

[0185] 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 DNA or RNA (single or double stranded, linear or supercoiled). 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 DNA or 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 DNA or RNA-targeting effector protein complexed with a guide RNA hybridized to a target sequence within the target locus of interest.

[0186] In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method may comprise modifying a target polynucleotide using a nucleic

acid-targeting complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. 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 polynucleotide sequence. For example, the method can be used to cleave a disease polynucleotide in a cell. For example, an exogenous 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 polynucleotide. The exogenous 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 polynucleotide 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 recombination template are selected to promote recombination between the RNA sequence of interest and the recombination. The upstream sequence is a polynucleotide sequence that shares sequence similarity with the sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a polynucleotide sequence that shares sequence similarity with the polynucleotide sequence downstream of the targeted site of integration. The upstream and downstream sequences in the recombination template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted sequence. Preferably, the upstream and downstream sequences in the recombination template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted sequence. In some methods, the upstream and downstream sequences in the recombination template have about 99% or 100% sequence identity with the targeted 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 recombination 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 recombination 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 sequence by integrating a recombination 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 recombination template such that the template is integrated into the 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 DNA or RNA (e.g., mRNA or pre-mRNA). In some methods, a target can be inactivated to affect the modification of the expression in a cell. For example, upon the binding of a nucleic acid-targeting complex to a target sequence in a cell, the target 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 of a nucleic acid-targeting 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., 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 polynucleotide. Examples of target polynucleotide include a disease associated polynucleotide. A “disease-associated” polynucleotide refers to any polynucleotide 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 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 polynucleotide 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 translated products may be known or unknown, and may be at a normal or abnormal level. The target RNA of a nucleic acid-targeting complex can be any polynucleotide 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 polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA).

[0187] In one embodiment, the method may comprise allowing a composition to bind to the target DNA or RNA to effect cleavage of said target DNA or RNA thereby modifying the target DNA or 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 DNA or RNA. In one aspect, the invention provides a method of modifying expression of DNA or RNA in a eukaryotic cell. In one embodiment, the method comprises allowing a nucleic acid-targeting complex to bind to the DNA or RNA such that said binding results in increased or decreased expression of said DNA or RNA; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a gRNA or guide RNA. Similar considerations and conditions apply as above for methods of modifying a target DNA or 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 DNA or RNA in a eukaryotic cell, which may be *in vivo*, *ex vivo* or *in vitro*. In one embodiment, 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. The compositions as described in any embodiment herein may be used to detect nucleic acid identifiers. Nucleic acid identifiers are non-coding nucleic acids that may be used to identify a particular article. Example nucleic acid identifiers, such as DNA watermarks, are described in Heider and Barnekow. "DNA watermarks: A proof of concept" BMC Molecular Biology 9:40 (2008). The nucleic acid identifiers may also be a nucleic acid barcode. A nucleic-acid based barcode is a short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier for an associated molecule, such as a target molecule and/or target nucleic acid. A nucleic acid barcode can have a length of at least, for example, 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, 60, 70, 80, 90, or 100 nucleotides, and can be in single- or double-stranded form. One or more nucleic acid barcodes can be attached, or "tagged," to a target molecule and/or target nucleic acid. This attachment can be direct (for example, covalent or non-covalent binding of the barcode to the

target molecule) or indirect (for example, via an additional molecule, for example, a specific binding agent, such as an antibody (or other protein) or a barcode receiving adaptor (or other nucleic acid molecule). Target molecule and/or target nucleic acids can be labeled with multiple nucleic acid barcodes in combinatorial fashion, such as a nucleic acid barcode concatemer. Typically, a nucleic acid barcode is used to identify target molecules and/or target nucleic acids as being from a particular compartment (for example a discrete volume), having a particular physical property (for example, affinity, length, sequence, etc.), or having been subject to certain treatment conditions. Target molecule and/or target nucleic acid can be associated with multiple nucleic acid barcodes to provide information about all of these features (and more). Methods of generating nucleic acid- barcodes are disclosed, for example, in International Patent Application Publication No. WO/2014/047561.

[0188] In an embodiment, compositions induce a double strand break for the purpose of inducing HDR-mediated correction. In a further embodiment, two or more sequence RNAs complexing with Cas12b polypeptide nuclease or an ortholog or homolog thereof, may be used to induce multiplexed breaks for purpose of inducing HDR-mediated correction.

[0189] A recombination template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with compositions disclosed herein to alter the structure of a target position. In an embodiment, the target nucleic acid is modified to have some or all of the sequence of the recombination template nucleic acid, typically at or near cleavage site(s). In an embodiment, the recombination template nucleic acid is single stranded. In an alternate embodiment, the recombination template nucleic acid is double stranded. In an embodiment, the recombination template nucleic acid is DNA, e.g., double stranded DNA. In an alternate embodiment, the recombination template nucleic acid is single stranded DNA.

[0190] In one embodiment, a recombination template is provided to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a nucleic acid-targeting effector protein as a part of a nucleic acid-targeting complex.

[0191] A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. A recombination template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In one embodiment, the recombination template polynucleotide is complementary to a portion of

a polynucleotide comprising the target sequence. When optimally aligned, a recombination template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100 or more nucleotides). In one embodiment, when a recombination template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the recombination template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0192] In an embodiment, the recombination template nucleic acid alters the structure of the target position by participating in homologous recombination. In an embodiment, the recombination template nucleic acid alters the sequence of the target position. In an embodiment, the recombination template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

[0193] The recombination template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the recombination template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by an Cas12b polypeptide nuclease mediated cleavage event. In an embodiment, the recombination template nucleic acid may include sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas12b polypeptide nuclease mediated event and a second site on the target sequence that is cleaved in a second Cas12b polypeptide nuclease mediated event.

[0194] In an embodiment, the recombination template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation. In an embodiment, the recombination template nucleic acid can include sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

[0195] A recombination template nucleic acid having homology with a target position in a target gene may be used to alter the structure of a target sequence. The recombination template

sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The recombination template nucleic acid may include sequence which, when integrated, results in: decreasing the activity of a positive control element; increasing the activity of a positive control element; decreasing the activity of a negative control element; increasing the activity of a negative control element; decreasing the expression of a gene; increasing the expression of a gene; increasing resistance to a disorder or disease; increasing resistance to viral entry; correcting a mutation or altering an unwanted amino acid residue conferring, increasing, abolishing or decreasing a biological property of a gene product, e.g., increasing the enzymatic activity of an enzyme, or increasing the ability of a gene product to interact with another molecule.

[0196] The recombination template nucleic acid may include sequence which results in: a change in sequence of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or more nucleotides of the target sequence. In an embodiment, the recombination template nucleic acid may be 20 $\pm$ 10, 30 $\pm$ 10, 40 $\pm$ 10, 50 $\pm$ 10, 60 $\pm$ 10, 70 $\pm$ 10, 80 $\pm$ 10, 90 $\pm$ 10, 100 $\pm$ 10, 110 $\pm$ 10, 120 $\pm$ 10, 130 $\pm$ 10, 140 $\pm$ 10, 150 $\pm$ 10, 160 $\pm$ 10, 170 $\pm$ 10, 180 $\pm$ 10, 190 $\pm$ 10, 200 $\pm$ 10, 210 $\pm$ 10, of 220 $\pm$ 10 nucleotides in length. In an embodiment, the recombination template nucleic acid may be 30 $\pm$ 20, 40 $\pm$ 20, 50 $\pm$ 20, 60 $\pm$ 20, 70 $\pm$ 20, 80 $\pm$ 20, 90 $\pm$ 20, 100 $\pm$ 20, 110 $\pm$ 20, 120 $\pm$ 20, 130 $\pm$ 20, 140 $\pm$ 20, 150 $\pm$ 20, 160 $\pm$ 20, 170 $\pm$ 20, 180 $\pm$ 20, 190 $\pm$ 20, 200 $\pm$ 20, 210 $\pm$ 20, of 220 $\pm$ 20 nucleotides in length. In an embodiment, the recombination template nucleic acid is 10 to 1,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to 300, 50 to 200, or 50 to 100 nucleotides in length.

[0197] A recombination template nucleic acid comprises the following components: [5' homology arm]-[replacement sequence]-[3' homology arm]. The homology arms provide for recombination into the chromosome, thus replacing the undesired element, e.g., a mutation or signature, with the replacement sequence. In an embodiment, the homology arms flank the most distal cleavage sites. In an embodiment, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In an embodiment, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' from the 5' end of the replacement sequence. In an embodiment, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In an embodiment, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400,

500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 3' from the 3' end of the replacement sequence.

[0198] In an embodiment, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In one embodiment, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[0199] In an embodiment, a recombination template nucleic acids for correcting a mutation may be designed for use as a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

[0200] Mutating key residues in the DNA cleavage domains of the Cas12b polypeptide nuclease, results in the generation of a catalytically inactive Cas12b polypeptide nuclease. A catalytically inactive Cas12b polypeptide complexes with a guide RNA and localizes to the DNA sequence specified by that guide RNA's targeting domain, however, it does not cleave the target DNA. Fusion of the inactive Cas12b polypeptide to an effector domain, e.g., a transcription repression domain, enables recruitment of the effector to any DNA site specified by the guide RNA. In an embodiment, Cas12b polypeptide 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 Cas12b polypeptide can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene.

[0201] In an embodiment, a guide RNA can be targeted to a known transcription response elements (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 expression of the target DNA.

[0202] In some methods, a target polynucleotide can be inactivated to affect the modification of the expression in a cell. For example, upon the binding of a composition 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.

METHODS FOR MODIFYING TARGET SEQUENCES

[0203] 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 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 Cas12b polypeptide 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 composition to desired cell types. The methods herein include a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

[0204] In example embodiments, the present disclosure provides a method of modifying target polynucleotides comprising, delivering the composition, the one or more polynucleotides, or the one or more vectors described above to a cell, or population of cells, comprising the target polynucleotides, wherein the complex directs the reverse transcriptase to the target sequence and the reverse transcriptase facilitates insertion of a donor sequence encoded by the donor template from the scaffold sequence into the target polynucleotide.

[0205] In an embodiment, the present disclosure provides a method where insertion of the donor sequence (a) introduces one or more base edits, (b) corrects or introduces a premature stop codon, (c) disrupts a splice site, (d) inserts or restores a splice site, (e) inserts a gene or gene fragment at one or both alleles of the target polynucleotide, or (f) a combination thereof.

[0206] In an example embodiment, the present disclosure provides an isolated cell or progeny thereof comprising the modifications made using the methods provided herein.

METHODS FOR GENE EDITING

[0207] In another aspect, the Cas polypeptides disclosed herein may be engineered to have either a nickase, or to be catalytically inactive (“dead Cas, dCas”) and further engineered to associate with a heterologous functional domains.

[0208] The Cas12b polypeptide may be in a dead form, e.g. does not have nuclease or nickase activity. In one embodiment, the systems further comprising one or more functional domains, e.g., nucleotide deaminase, reverse transcriptase, non-LTR retrotransposon (and

protein encoded), polymerase, diversity generating element (and protein encoded) and integrases. In some examples, the systems further comprise one or more donor polynucleotides. The donor polynucleotides may be inserted to a target polynucleotide by the systems. The donor polynucleotide may be comprised in or coded by a nucleic acid template.

Cas12b Base Editing Systems

[0209] The present disclosure also provides for systems. In general, such a system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) associated (e.g., fused) with a Cas12b polypeptide nuclease, e.g., Cas12b protein. The Cas12b polypeptide nuclease may be a dead Cas12b polypeptide nuclease (such as a Cas12b polypeptide nickase, e.g., engineered from a Cas12b polypeptide nuclease). In certain examples, the nucleotide deaminase is a mutated form of an adenosine deaminase. The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities.

[0210] In some examples, the present disclosure provides an engineered, non-naturally occurring composition comprising: the nucleic acid-guided nuclease that is catalytically inactive, a nucleotide deaminase associated with or otherwise capable of forming a complex with the Cas12b protein, and a single guide RNA capable of forming a complex with the Cas12b protein and directing site-specific binding at a target sequence.

[0211] In one aspect, the present disclosure provides an engineered adenosine deaminase. The engineered adenosine deaminase may comprise one or more mutations herein. In one embodiment, the engineered adenosine deaminase has cytidine deaminase activity. In certain examples, the engineered adenosine deaminase has both cytidine deaminase activity and adenosine deaminase. In some cases, the modifications by base editors herein may be used for targeting post-translational signaling or catalysis. In one embodiment, compositions herein comprise nucleotide sequence comprising encoding sequences for one or more components of a base editing system. A base-editing system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) fused with a Cas12b polypeptide nuclease or a variant thereof. In some cases, the target polynucleotide is edited at one or more bases to introduce a G→A or C→T mutation.

[0212] In some cases, the adenosine deaminase is double-stranded RNA-specific adenosine deaminase (ADAR). Examples of ADARs include those described Yiannis A Savva et al., The ADAR protein family, *Genome Biol.* 2012; 13(12): 252, which is incorporated by reference in its entirety. In some examples, the ADAR may be hADAR1. In certain examples, the ADAR

may be hADAR2. The sequence of hADAR2 may be that described under Accession No. AF525422.1.

[0213] In some cases, the deaminase may be a deaminase domain, e.g., a deaminase domain of ADAR (“ADAR-D”). In one example, the deaminase may be the deaminase domain of hADAR2 (“hADAR2-D”), e.g., as described in Phelps KJ et al., Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res.* 2015 Jan;43(2):1123-32, which is incorporated by reference herein in its entirety. In a particular example, the hADAR2-D has a sequence comprising amino acid 299-701 of hADAR2-D, e.g., amino acid 299-701 of the sequence under Accession No. AF525422.1.

[0214] In certain examples, the system comprises a mutated form of an adenosine deaminase fused with a dead Cas12b polypeptide nuclease (e.g., a Cas12b polypeptide nickase). The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In

one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In one embodiment, the adenosine deaminase may

comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising one or more mutations of E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, fused with a dead Cas12b polypeptide nuclease or Cas12b polypeptide nickase. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, and S661T, fused with a dead Cas12b polypeptide nuclease or Cas12b polypeptide nickase. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, and S375N fused with a dead Cas12b polypeptide nuclease or Cas12b polypeptide nickase.

[0215] In one embodiment, the adenosine deaminase may be a tRNA-specific adenosine deaminase or a variant thereof. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: W23L, W23R, R26G, H36L, N37S, P48S, P48T, P48A, I49V, R51L, N72D, L84F, S97C, A106V, D108N, H123Y, G125A, A142N, S146C, D147Y, R152H, R152P, E155V, I156F, K157N, K161T, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: D108N based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine

deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, A142N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, A142N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, A142N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above. In one embodiment, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, P48S, W23R, P48A, R152P, A142N, based on amino acid sequence positions of *E. coli* TadA, and mutations in a homologous deaminase protein corresponding to the above.

[0216] In some examples, the base editing systems may comprise an intein-mediated trans-splicing system that enables *in vivo* delivery of a base editor, e.g., a split-intein cytidine base editors (CBE) or adenine base editor (ABE) engineered to trans-splice. Examples of such base editing systems include those described in Colin K.W. Lim et al., Treatment of a Mouse Model of ALS by *In Vivo* Base Editing, *Mol Ther.* 2020 Jan 14. pii: S1525-0016(20)30011-3. doi: 10.1016/j.ymthe.2020.01.005; and Jonathan M. Levy et al., Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses, *Nature Biomedical Engineering* volume 4, pages 97–110(2020), which are incorporated by reference herein in their entireties.

[0217] Examples of base editing systems include those described in International Patent Publication Nos. WO 2019/071048 (e.g. paragraphs [0933]-[0938]), WO 2019/084063 (e.g., paragraphs [0173]-[0186], [0323]-[0475], [0893]-[1094]), WO 2019/126716 (e.g., paragraphs [0290]-[0425], [1077]-[1084]), WO 2019/126709 (e.g., paragraphs [0294]-[0453]), WO 2019/126762 (e.g., paragraphs [0309]-[0438]), WO 2019/126774 (e.g., paragraphs [0511]-[0670]), Cox DBT, et al., RNA editing with CRISPR-Cas13, *Science.* 2017 Nov 24;358(6366):1019-1027; Abudayyeh OO, et al., A cytosine deaminase for programmable single-base RNA editing, *Science* 26 Jul 2019: Vol. 365, Issue 6451, pp. 382-386; Gaudelli NM et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, *Nature* volume 551, pages 464–471 (23 November 2017); Komor AC, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature.* 2016 May 19;533(7603):420-4; Jordan L. Doman et al., Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors, *Nat Biotechnol* (2020). doi.org/10.1038/s41587-020-0414-6; and Richter MF et al., Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity, *Nat Biotechnol* (2020). doi.org/10.1038/s41587-020-0453-z, which are incorporated by reference herein in their entireties and can be used to adapt to the Cas12b polypeptides.

Cas12b Prime Editing Systems

[0218] In one embodiment, the present disclosure provides compositions and systems may comprise a Cas12b polypeptide nuclease or a catalytically inactive form, one or more guide RNAs, and a reverse transcriptase. The systems may be used to insert a donor polynucleotide to a target polynucleotide. In some examples, the composition or system comprises a catalytically inactive Cas12b polypeptide nuclease, a reverse transcriptase associated with or

otherwise capable of forming a complex with the Cas12b polypeptide nuclease, and a guide RNA capable of forming a complex with the Cas12b polypeptide nuclease and directing site-specific binding of the complex to a target sequence of a target polynucleotide, the guide RNA further comprising a donor sequence for insertion into the target polynucleotide.

[0219] In some cases, the catalytically inactive Cas12b polypeptide is a nickase, e.g., a DNA nickase. In some cases, the Cas12b polypeptide has one or more mutations. In some examples, the Cas12b polypeptide comprises mutations corresponding to the mutations in the RuvC nuclease.

[0220] The Cas12b polypeptide may be associated with a reverse transcriptase. A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof. In certain aspects, the reverse transcriptase is Human immunodeficiency virus (HIV) RT, Avian myoblastosis virus (AMV) RT, Moloney murine leukemia virus (M-MLV) RT a group II intron RT, a group II intron-like RT, or a chimeric RT. In an embodiment, the RT comprises modified forms of these RTs, such as, engineered variants of Avian myoblastosis virus (AMV) RT, Moloney murine leukemia virus (M-MLV) RT, or Human immunodeficiency virus (HIV) RT (see, e.g., Anzalone, et al., Search-and-replace genome editing without double-strand breaks or donor DNA, *Nature*. 2019 Dec;576(7785):149-157).

[0221] In some examples, the compositions and systems may comprise the Cas12b protein disclosed herein; a reverse transcriptase (RT) polypeptide connected to or otherwise capable of forming a complex with the Cas12b polypeptide; and a guide RNA capable of forming a complex with the Cas12b polypeptide and comprising: a guide RNA capable of directing site-specific binding of the Cas12b polypeptide/RNP complex to a target sequence of a target polynucleotide; a 3' binding site region capable of binding to a cleaved upstream strand of the target polynucleotide; and a RT template sequence encoding an extended sequence, wherein the extended sequence comprises a variant region and a 3' homologous sequence capable of hybridization to the downstream cleaved strand of the target polynucleotide.

[0222] A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof. A wide variety of reverse transcriptases (RT) may be used in alternative embodiments of the present invention, including prokaryotic and eukaryotic RT, provided that the RT functions within the host to generate a donor polynucleotide sequence from the RNA template. If desired, the nucleotide sequence of a native RT may be modified, for example using known codon optimization techniques, so that expression within the desired host is optimized. A reverse

transcriptase (RT) is an enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed reverse transcription. Reverse transcriptases are used by retroviruses to replicate their genomes, by retrotransposon mobile genetic elements to proliferate within the host genome, by eukaryotic cells to extend the telomeres at the ends of their linear chromosomes, and by some non-retroviruses such as the hepatitis B virus, a member of the Hepadnaviridae, which are dsDNA-RT viruses. Retroviral RT has three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H, and DNA-dependent DNA polymerase activity. Collectively, these activities enable the enzyme to convert single-stranded RNA into double-stranded cDNA. In an embodiment, the RT domain of a reverse transcriptase is used in the present invention. The domain may include only the RNA-dependent DNA polymerase activity. In some examples, the RT domain is non-mutagenic, i.e., does not cause mutation in the donor polynucleotide (e.g., during the reverse transcriptase process). In some examples, the RT domain may be non-retron RT, e.g., a viral RT or a human endogenous RT. In some examples, the RT domain may be retron RT or DGRs RT. In some examples, the RT may be less mutagenic than a counterpart wildtype RT. In one embodiment, the RT herein is not mutagenic.

[0223] The reverse transcriptase may be fused to the C-terminus of a Cas12b polypeptide. Alternatively or additionally, the reverse transcriptase may be fused to the N-terminus of a Cas12b polypeptide. The fusion may be via a linker and/or an adaptor protein. In some examples, the reverse transcriptase may be an M-MLV reverse transcriptase or variant thereof. The M-MLV reverse transcriptase variant may comprise one or more mutations. For the examples, the M-MLV reverse transcriptase may comprise D200N, L603W, and T330P. In another example, the M-MLV reverse transcriptase may comprise D200N, L603W, T330P, T306K, and W313F. In a particular example, the fusion of Cas12b polypeptide and reverse transcriptase is Cas12b polypeptide (with a mutation corresponding to H840A of SpCas9) fused with M-MLV reverse transcriptase (D200N+L603W+T330P+T306K+W313F).

[0224] In one embodiment, the Cas12b polypeptide herein may target DNA using a guide sequence RNA containing a binding sequence that hybridizes to the target sequence on the DNA. The guide RNA may further comprise an editing sequence that contains new genetic information that replaces target DNA nucleotides. The small sizes of the Cas12b polypeptide herein may allow easier packaging and delivery of the prime editing system, e.g., with a viral vector, e.g., AAV or lentiviral vector.

[0225] A single-strand break (a nick) may be generated on the target DNA by the Cas12b polypeptide at the target site to expose a 3'-hydroxyl group, thus priming the reverse transcription of an edit-encoding extension on the ω RNA or guide directly into the target site. These steps may result in a branched intermediate with two redundant single-stranded DNA flaps: a 5' flap that contains the unedited DNA sequence, and a 3' flap that contains the edited sequence copied from the guide RNA. The 5' flaps may be removed by a structure-specific endonuclease, e.g., FEN122, which excises 5' flaps generated during lagging-strand DNA synthesis and long-patch base excision repair. The non-edited DNA strand may be nicked to induce bias DNA repair to preferentially replace the non-edited strand. Examples of prime editing systems and methods include those described in Anzalone AV *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA, *Nature*. 2019 Oct 21. doi: 10.1038/s41586-019-1711-4, which is incorporated by reference herein in its entirety.

[0226] The Cas12b polypeptide (e.g., the nickase form) may be used to prime-edit a single nucleotide on a target DNA. Alternatively or additionally, the Cas12b polypeptide may be used to prime-edit at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 1000 nucleotides on a target DNA.

[0227] In yet another embodiment, PRIME editing is used first to create a longer 3' region (e.g. 20 nucleotides). Examples of prime editing systems and methods include those described in Anzalone AV *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA, *Nature*. 2019 Oct 21. doi: 10.1038/s41586-019-1711-4, which is incorporated by reference herein in its entirety. In such cases, the system comprises a Cas12b polypeptide with nickase activity, a reverse transcriptase domain, and a DNA polymerase, and a guide molecule comprising a binding sequence capable of hybridizing to the target polynucleotide and a editing sequence. The generated region may be further extended on a DNA template as described herein. The latter may allow generation of a target-independent sequence, compatible with a generic donor sequence.

[0228] The Cas12b polypeptide is capable of generating a first cleavage in the target sequence and a second cleavage outside the target sequence on the target polynucleotide. In

some variations, a second Cas12b polypeptide-mediated cleavage in vicinity to the target site may be made, which may enable more efficient invasion of the extended DNA.

[0229] In some examples, the compositions and systems of the Cas12b polypeptide herein comprise: a reverse transcriptase (RT) polypeptide connected to or otherwise capable of forming a complex with the Cas12b polypeptide; a first guide sequence capable of forming a first Cas12b polypeptide-Reverse transcriptase complex with the Cas12b polypeptide and comprising: a guide RNA capable of directing site-specific binding of the first Cas12b polypeptide-Reverse transcriptase complex to a first target sequence of a target polynucleotide; a first binding site region capable of binding to a cleaved or nicked strand of the target polynucleotide; and a RT template sequence encoding a first extended sequence; a second guide RNA capable of forming a second Cas12b polypeptide-Reverse transcriptase complex with the Cas12b polypeptide and comprising: a guide RNA capable of directing site specific binding of the second Cas12b polypeptide-Reverse transcriptase complex to a second target sequence of the target polynucleotide; a second binding site region capable of binding to a cleaved or nicked strand of the target polynucleotide; and a RT template sequence encoding a second extended sequence.

[0230] In some cases, the compositions and systems may further comprise: a donor template; a third guide RNA capable of forming a Cas12b polypeptide-Reverse transcriptase complex- guide RNA with the Cas12b polypeptide and comprising: a guide RNA sequence capable of directing site-specific binding to a target sequence on the donor template; a third binding region capable of binding to a cleaved or nicked strand of the donor template; and a RT template encoding a third extended region complementary to the first extended region generated on the target polynucleotide; and a fourth gRNA or guide sequence capable of forming a Cas12b polypeptide-Reverse transcriptase complex with the Cas12b polypeptide and comprising: a guide sequence capable of directing site-specific binding to a second target sequence on the donor template; a fourth binding region capable of binding to a cleaved or nicked strand of the donor template; and a RT template encoding a fourth extended region complementary to the second extended region generated on the target polynucleotide.

[0231] In some cases, the compositions and systems may further comprise a site-specific recombinase, and wherein the first and second extended regions are complementary to each other and introduce a serine integrase recombination site; and a donor molecule comprising a

donor sequence for insertion into the target polypeptide and the complementary recombination site to the serine integrase recombination site.

[0232] In some examples, the compositions and systems may further comprise a recombinase. The recombinase is connected to or otherwise capable of forming a complex with the Cas12b polypeptide. In an embodiment, the complex is capable of inserting a recombination site in the DNA loci of interest by extension of RT templates that encode for the recombination site on the 3' extension of the guide sequences by the reverse transcriptase. In an embodiment, a donor template comprising a compatible recombination site is provided that can recombine unidirectionally with the inserted recombination site when a recombinase specific for the recombination site is also provided. In an embodiment, the donor template is a plasmid comprising the complementary recombination site and any sequence for insertion at the DNA loci of interest. In an embodiment, the recombinase is connected to or capable of forming a complex with the Cas12b polypeptide, such that all of the enzymatic proteins are brought into contact at the loci of interest. In an embodiment, the recombinase is codon optimized for eukaryotic cells (described further herein). In an embodiment, the recombinase includes a NLS (described further herein). In an embodiment, the recombinase is provided as a separate protein. The separate recombinase may form a dimer and bind to the donor template recombination site. The recombinase may be targeted to the loci of interest as a result of the insertion of the compatible recombination site that is also recognized by the recombinase. Thus, the recombinase may recognize the recombination site inserted at the DNA loci of interest and the recombination site on the donor and be targeted to the DNA loci of interest without any additional modifications to the recombinase.

[0233] In an embodiment, a second Cas12b complex connected to a recombinase is targeted to the DNA loci of interest. In an embodiment, the second Cas12b complex comprises a dead Cas12b protein (dCas12b, described further herein), such that the recombinase is targeted to the DNA loci of interest, but the target sequence is not further cleaved. In an embodiment, the dCas12b targets a sequence generated only after the insertion of the recombination site. In an embodiment, the recombinase recognizes and binds to the donor template recombination site and the inserted recombination site. In an embodiment, the recombinase forms a dimer with a recombinase provided as a separate protein.

[0234] As used herein, the term "Recombinase" refers to an enzyme that catalyzes recombination between two or more recombination sites (e.g., an acceptor and donor site).

Recombinases useful in the present invention catalyze recombination at specific recombination sites which are specific polynucleotide sequences that are recognized by a particular recombinase. “Uni-directional recombinases” or “integrases” refer to recombinase enzymes whose recognition sites are destroyed after the recombination has taken place. The term “integrase” refers to a type of recombinase. In other words, the sequence recognized by the recombinase is changed into one that is not recognized by the recombinase upon recombination. As a result, once a sequence is subjected to recombination by the uni-directional recombinase, the continued presence of the recombinase cannot reverse the previous recombination event.

[0235] “Recombination sites” are specific polynucleotide sequences that are recognized by the recombinase enzymes described herein. Typically, two different sites are involved (in regards to recombination termed “complementary sites”), one present in the target nucleic acid (e.g., a chromosome or episome of a eukaryote) and another on the nucleic acid that is to be integrated at the target recombination site. The terms “attB” and “attP,” which refer to attachment (or recombination) sites originally from a bacterial target (attachment site of bacteria) and a phage donor (attachment site of phage), respectively, are used herein although recombination sites for particular enzymes may have different names. The two attachment sites can share as little sequence identity as a few base pairs. The recombination sites typically include left and right arms separated by a core or spacer region. Thus, an attB recombination site consists of BOB', where B and B' are the left and right arms, respectively, and O is the core region. Similarly, attP is POP', where P and P' are the arms and O is again the core region. Upon recombination between the attB and attP sites, and concomitant integration of a nucleic acid at the target, the recombination sites that flank the integrated DNA are referred to as “attL” and “attR.” The attL and attR sites, using the terminology above, thus consist of BOP' and POB', respectively. In some representations herein, the “O” is omitted and attB and attP, for example, are designated as BB' and PP', respectively.

Cas12b Recombinase/Integrase Systems

[0236] The systems and compositions herein may comprise an CRISPR-Cas12b system, and one or more components of a recombinase or integrase. In an aspect, the Cas12b polypeptide is naturally catalytically inactive and utilized with one or more nucleic acid components to provide site-specific targeting, and the one or more components of the recombinase to introduce a modification. In an aspect, the Cas12b polypeptide may be

catalytically inactivated via mutation of one or more residues of a catalytic domain or via truncation, and utilized with one or more RNA components to provide site-specific targeting, and the one or more components of the recombinase introduce a modification. In an embodiment, the Cas12b polypeptide is naturally catalytically inactive. In one embodiment, a naturally inactive Cas12b is provided with a recombinase, e.g. an integrase, and optionally a reverse transcriptase.

[0237] A recombinase generally is an enzyme that mediates recombination, e.g. breaking and rejoining, of nucleic acids at specific points. DNA site-specific recombinases include serine integrases, which are phage-encoded site-specific recombinases that promote conservative recombination reactions between DNA substrates located on the phage (phage attachment site, *attP*) and bacterial attachment site, *attB*. In one embodiment, the recombinase is a serine integrase that drives a highly directions site-specific recombination.

[0238] In preferred embodiments, the recombinase mediates unidirectional site-specific recombination. In one embodiment, the recombinase is a serine recombinase (SR) also referred to as a serine integrase, encoded, for example, by IS607 family, Tn4451, and bacteriophage phiC31. See, generally, Smith MC, Thorpe HM: Diversity in the serine recombinases. *Mol Microbiol.* 2002, 44: 299-307. 10.1046/j.1365-2958.2002.02891.x; Li et al.,(2018) *J. Mol. Biol.* 430:21, 4401-4418.

[0239] In an embodiment, the recombinase is a tyrosine recombinase (YR) encoded by IS91, Helitron, IS200/IS605, Crypton or DIRS-retrotransposon families. See, generally, Goodwin TJ, Butler MI, Poulter T: Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology.* 2003, 149: 3099-3109. Doi:10.1099/mic.0.26529-0; Cappello J, Handelsman K, Lodish HF: Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell.* 1985, 43: 105-115. 10.1016/0092-8674(85)90016-9.

[0240] In an aspect, the recombinase provides site-specific integration of a template that can be provided with the composition, e.g. a donor oligonucleotide. Without being bound by theory, the recombinase allows for integration independent of payload size and can coordinate strand exchange and re-ligation across multiple cell types, allowing integration of long stretches of polynucleotides. In an exemplary embodiment, the serine recombinase is PhiC31 and the target is DNA. In an aspect, the phiC31 allows for integration of a target site comprising an *attP* or pseudoattP recognition site. See, e.g. systembio.com/wp-

content/uploads/phiC31_productsheet-1.pdf. In an embodiment utilizing phiC231, a donor oligonucleotide would be provided with an attB at sequence that facilitates attachment at the attP site of the target genome. Similar approaches of designing donor oligonucleotides with sequences complementary to attachment sites for a recombinase can be designed for use with the present invention. See, e.g. Li et al., (2018) J. Mol. Biol. 430:21, 4401-4418.

[0241] In preferred embodiments, the integrase mediates gene integration at diverse loci by directing insertion with an Cas12b nickase fused to both a reverse transcriptase and an integrase. In one embodiment, the integrase is a serine integrase, for example, BxbINT. See, generally, Yarnall *et al.*, Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases, Nature Biotechnology. 2022 Nov 24. doi: 10.1038/s41587-022-01527-4, which is incorporated by reference herein in its entirety. In Yarnall, Gootenberg, Abudayyeh, and colleagues show integration using a CRISPR-Cas9 nickase fused to a reverse transcriptase and serine integrase termed Programmable Addition via Site-specific Targeting Elements (PASTE) with delivery via a single dose of plasmids with functionality in non-dividing and primary cells, utilizing a guide RNA comprising an attB landing site, termed attachment site-containing guide RNA were used to insert sequences, including diverse cargo sequences that can be inserted across different loci, varying in size up to about 36 kb. Additional uses of the PASTE system included gene tagging, gene replacement, gene delivery, and protein production and secretion, approaches that are contemplated for use with the Cas12b nickase and integrase approach. In an aspect, the guide sequence RNA may comprise an attB landing site. In an aspect, the recombinase provides site-specific integration of a template that can be provided with the composition, e.g. a donor oligonucleotide.

[0242] Additional large serine integrases can be used with the Cas12b nickase, for example as identified and described in Durrant et al., Large-scale discovery of recombinases for integrating DNA into the human genome, doi:10.1101/2021.11.05.467528, incorporated herein by reference. Other integrases include BceINT, SscINT, SacINT. See Yarnall et al., 2022 at Fig. 5.

[0243] Without being bound by theory, the recombinase allows for integration independent of payload size and can coordinate strand exchange and re-ligation across multiple cell types, allowing integration of long stretches of polynucleotides. In an exemplary embodiment, the integrase is BxbINT and the target is DNA. In an aspect, the BxbINT allows for integration of a target site comprising an attP or pseudoattP recognition site. In an embodiment utilizing

BxbINT, a donor oligonucleotide would be provided with an attB at sequence that facilitates attachment at the attP site of the target genome. Similar approaches of designing donor oligonucleotides with sequences complementary to attachment sites for an integrase can be designed for use with the present invention, for example a circular double-strand DNA template containing the attP attachment site, or delivery of large cargo via an adenovirus or other viral vector, as described elsewhere herein. See, e.g. Yarnall et al., 2022 at Figs. 1a, 1b, and 6.

Cas12b Guided Excision-Transposition Systems

[0244] Embodiments disclosed herein provide an engineered or non-natural guided excision-transposition system. The engineered or non-natural guided excision-transposition system may comprise one or more components of a CRISPR-Cas12b system and one or more components of a Class II transposon. The components of the Cas12b guide RNA can direct the Class II transposon component(s) to retrotransposon to a target nucleic acid sequence and direct its transposition into a recipient polynucleotide.

[0245] For example, the engineered or non-natural guided excision-transposition systems that can include (a) a first Cas12b polypeptide; (b) a first Class II transposon polypeptide coupled to or otherwise capable of complexing with the first Cas12b polypeptide; (c) a first guide molecule capable of forming a first Cas12b-guide RNA complex with the first Cas12b protein and directing site-specific binding to a first target sequence of a first target polynucleotide; (d) a second Cas12b polypeptide; (e) a second Class II transposon polypeptide coupled to or otherwise capable of complexing with the second Cas12b polypeptide; (f) a second guide molecule capable of forming a second Cas12b-guide RNA complex with the first Cas12b protein and directing site-specific binding to a second target sequence of the first target polynucleotide; and (g) a Class II transposon polynucleotide comprising the first target polynucleotide and is capable of forming a complex with the first and second Cas12b polypeptide, the first and second guide molecules, and the first and second Class II transposon polypeptides.

[0246] In one embodiment, the engineered or non-natural guided excision-transposition system can include (h) a third guide molecule capable of complexing with the first Cas12b polypeptide and directing site-specific binding to a first target sequence of a second target polynucleotide, wherein the third guide molecule is optionally coupled to the first Cas12b polypeptide; (i) optionally, a first guide polynucleotide that encodes the third guide sequence;

(j) a fourth guide sequence capable of complexing with the second Cas12b polypeptide and directing site-specific binding to a second target sequence of the second target polynucleotide, wherein the fourth guide molecule is optionally coupled to the second Cas12b polypeptide; and
(k) optionally, a second ω RNA or guide molecule polynucleotide that encodes the fourth guide sequence.

[0247] In one embodiment, the first and the second Class II transposon polypeptides are capable of excising the first target polynucleotide from the Class II transposon polynucleotide. In one embodiment, the first and the second Class II transposon polypeptides are capable of transposing the first target polynucleotide in the second target polynucleotide. In one embodiment, the first target polynucleotide does not include one or more Class II transposon long terminal repeats.

[0248] The engineered or non-natural guided excision-transposition systems described herein can be based on a Class II transposon or Class II transposon system. The engineered or non-natural guided excision-transposition system may include a first target polynucleotide, also referred to as a donor polynucleotide or transposon and a second target polynucleotide, which is also referred to herein as a recipient polynucleotide. As used herein, “transposon” (also referred to as transposable element) refers to a polynucleotide sequence that is capable of moving from location in a genome to another. There are several classes of transposons. Transposons include retrotransposons (Class I transposons) and DNA transposons (Class II transposons). In some cases, retrotransposons require the transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. DNA transposons are those that do not require reverse transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide.

[0249] Any suitable transposon system can be used. Suitable transposon and systems thereof can include, but are not limited, to Sleeping Beauty transposon system (Tc1/mariner superfamily) (see e.g. Ivics et al. 1997. Cell. 91(4): 501-510), piggyBac (piggyBac superfamily) (see e.g. Li et al. 2013 110(25): E2279-E2287 and Yusa et al. 2011. PNAS. 108(4): 1531-1536), Tol2 (superfamily hAT), Frog Prince (Tc1/mariner superfamily) (see e.g. Miskey et al. 2003 Nucleic Acid Res. 31(23):6873-6881) and variants thereof.

[0250] In one embodiment, the first and/or second Class II transposon polypeptide is a DD[E/D] transposon or transposon polypeptide. In one embodiment, the first and/or the second

Class II transposon polynucleotide is a Tc1/mariner, PiggyBac, Frog Prince, Tn3, Tn5, hAT, CACTA, P, Mutator, PIF/Harbinger, Transib, or a Merlin/IS1016 transposon polynucleotide. In one embodiment, the first and/or second Class II transposon polypeptide is a Tc1/mariner, PiggyBac, Frog Prince, Tn3, Tn5, hAT, CACTA, P, Mutator, PIF/Harbinger, Transib, or a Merlin/IS1016 transposon polypeptide.

[0251] Suitable Class II transposon systems and components that can be utilized can also be and are not limited to those described in e.g. and without limitation, Han et al., 2013. BMC Genomics. 14:71, doi: 10.1186/1471-2164-14-71, Lopez and Garcia-Perez. 2010. Curr. Genomics. 11(2):115-128; Wessler. 2006. PNAS. 103(47): 176000-17601; Gao et al., 2017. Marine Genomics. 34:67-77; Bradic et al. 2014. Mobile DNA. 5(12) doi:10.1186/1759-8753-5-12; Li et al., 2013. PNAS. 110(25)E2279-E2287; Kebriaei et al. 2017. Trends in Genetics. 33(11): 852-870; Miskey et al. 2003. Nucleic Acid res. 31(23):6873-6881; Nicolas et al. 2015. Microbiol Spectr. 3(4) doi: 10.1128/microbiolspec.MDNA3-0060-2014); W.S. Reznikoff. 1993. Annu Rev. Microbiol. 47:945-963; Rubin et al. 2001. Genetics. 158(3): 949-957; Wicker et al. 2003. Plant Physiol. 132(1): 52-63; Majumdar and Rio. 2015. Microbiol. Spectr. 3(2) doi: 10.1128/microbiolspec.MDNA3-0004-2014; D. Lisch. 2002. Trends in Plant Sci. 7(11): 498-504; Sinzelle et al. 2007. PNAS. 105(12): 4715-4720; Han et al. 2014; Genome Biol. Evol. 6(7):1748-1757; Grzebelus et al. 2006; Mol. Genet. Genomics. 275(5):450-459; Zhang et al. 2004. Genetics. 166(2):971-986; Chen and Li. 2008. Gene. 408(1-2):51-63; and C. Feschotte. 2004. Mol. Biol. Evol. 21(9):1769-1780.

CRISPR Associated Transposases (CAST)

[0252] The Cas polypeptides disclosed herein may be used in CAST systems which comprise a Cas polypeptide, a guide molecule, a transposase, and a donor construct. The transposase is linked to or otherwise capable of forming a complex with the Cas polypeptide. The donor construct comprises a donor sequence to be inserted into a target polynucleotide and one or more transposase recognition elements. The transposase is capable of binding the donor construct and excising the donor template and directing insertion of the donor template into a target site on a target polynucleotide (e.g. genomic DNA). The guide molecule is capable of forming a CRISPR-Cas complex with the Cas polypeptide, and can be programmed to direct the entire CAST complex such that the transposase is positioned to insert the donor sequence at the target site on the target polynucleotide. For multimeric transposase, only those transposases needed for recognition of the donor construct and transposition of the donor

sequence into the target polypeptide may be required. The Cas may be naturally catalytically inactive or engineered to be catalytically inactive.

[0253] In one example embodiment, the CAST system is a Tn7-like CAST system, wherein the transposase comprises one or more polypeptides from a Tn7 or Tn7-like transposase. The Cas polypeptide

[0254] In one example embodiments, the Tn7 transposase may comprise TnsB, TnsC, and TniQ. In another example embodiment, the Tn7 transposase may comprise TnsB, TnsC, and TnsD. In certain example embodiments, the Tn7 transposase may comprise TnsD, TnsE, or both. As used herein, the terms “TnsAB”, “TnsAC”, “TnsBC”, or “TnsABC” refer to a transposon complex comprising TnsA and TnsB, TnsA and TnsC, TnsB and TnsC, TnsA and TnsB and TnsC, respectively. In these combinations, the transposases (TnsA, TnsB, TnsC) may form complexes or fusion proteins with each other. Similarly, the term TnsABC-TniQ refer to a transposon comprising TnsA, TnsB, TnsC, and TniQ, in a form of complex or fusion protein. An example Type 1f-Tn7 CAST system is described in Klompe et al. Nature, 2019, 571:219-224 and Vo et al. bioRxiv, 2021, doi.org/10.1101/2021.02.11.430876, which are incorporated herein by reference.

[0255] In one example embodiment, the CAST system is a Mu CAST system, wherein the transposase comprises one or more polypeptides of a Mu transposase. An example Mu CAST system is disclosed in WO/2021/041922 which is incorporated herein by reference.

Donor Polynucleotides

[0256] The system may further comprise one or more donor polynucleotides (e.g., for insertion into the target polynucleotide). A donor polynucleotide may be an equivalent of a transposable element that can be inserted or integrated to a target site. The donor polynucleotide may be or comprise one or more components of a transposon. A donor polynucleotide may be any type of polynucleotides, including, but not limited to, a gene, a gene fragment, a non-coding polynucleotide, a regulatory polynucleotide, a synthetic polynucleotide, etc. The donor polynucleotide may include a transposon left end (LE) and transposon right end (RE). The LE and RE sequences may be endogenous sequences for the CAST used or may be heterologous sequences recognizable by the CAST used, or the LE or RE may be synthetic sequences that comprise a sequence or structure feature recognized by the CAST and sufficient to allow insertion of the donor polynucleotide into the target polynucleotides. In certain example embodiments, the LE and RE sequences are truncated. In certain example embodiments may

be between 100-200 bps, between 100-190 base pairs, 100-180 base pairs, 100-170 base pairs, 100-160 base pairs, 100-150 base pairs, 100-140 base pairs, 100-130 base pairs, 100-120 base pairs, 100-110 base pairs, 20-100 base pairs, 20-90 base pairs, 20-80 base pairs, 20-70 base pairs, 20-60 base pairs, 20-50 base pairs, 20-40 base pairs, 20-30 base pairs, 50 to 100 base pairs, 60-100 base pairs, 70-100 base pairs, 80-100 base pairs, or 90-100 base pairs in length

[0257] The donor polynucleotide may be inserted at a position upstream or downstream of a PAM on a target polynucleotide. In some embodiments, a donor polynucleotide comprises a PAM sequence. Examples of PAM sequences include TTTN, ATTN, NGTN, RGTR, VGTD, or VGTR.

[0258] The donor polynucleotide may be inserted at a position between 10 bases and 200 bases, e.g., between 20 bases and 150 bases, between 30 bases and 100 bases, between 45 bases and 70 bases, between 45 bases and 60 bases, between 55 bases and 70 bases, between 49 bases and 56 bases or between 60 bases and 66 bases, from a PAM sequence on the target polynucleotide. In some cases, the insertion is at a position upstream of the PAM sequence. In some cases, the insertion is at a position downstream of the PAM sequence. In some cases, the insertion is at a position from 49 to 56 bases or base pairs downstream from a PAM sequence. In some cases, the insertion is at a position from 60 to 66 bases or base pairs downstream from a PAM sequence.

[0259] The donor polynucleotide may be used for editing the target polynucleotide. In some cases, the donor polynucleotide comprises one or more mutations to be introduced into the target polynucleotide. Examples of such mutations include substitutions, deletions, insertions, or a combination thereof. The mutations may cause a shift in an open reading frame on the target polynucleotide. In some cases, the donor polynucleotide alters a stop codon in the target polynucleotide. For example, the donor polynucleotide may correct a premature stop codon. The correction may be achieved by deleting the stop codon or introduces one or more mutations to the stop codon. In other example embodiments, the donor polynucleotide addresses loss of function mutations, deletions, or translocations that may occur, for example, in certain disease contexts by inserting or restoring a functional copy of a gene, or functional fragment thereof, or a functional regulatory sequence or functional fragment of a regulatory sequence. A functional fragment refers to less than the entire copy of a gene by providing sufficient nucleotide sequence to restore the functionality of a wild type gene or non-coding regulatory sequence (e.g. sequences encoding long non-coding RNA). In certain example

embodiments, the systems disclosed herein may be used to replace a single allele of a defective gene or defective fragment thereof. In another example embodiment, the systems disclosed herein may be used to replace both alleles of a defective gene or defective gene fragment. A “defective gene” or “defective gene fragment” is a gene or portion of a gene that when expressed fails to generate a functioning protein or non-coding RNA with functionality of a corresponding wild-type gene. In certain example embodiments, these defective genes may be associated with one or more disease phenotypes. In certain example embodiments, the defective gene or gene fragment is not replaced but the systems described herein are used to insert donor polynucleotides that encode gene or gene fragments that compensate for or override defective gene expression such that cell phenotypes associated with defective gene expression are eliminated or changed to a different or desired cellular phenotype.

[0260] In certain embodiments of the invention, the donor may include, but not be limited to, genes or gene fragments, encoding proteins or RNA transcripts to be expressed, regulatory elements, repair templates, and the like. According to the invention, the donor polynucleotides may comprise left end and right end sequence elements that function with transposition components that mediate insertion.

[0261] In certain cases, the donor polynucleotide manipulates a splicing site on the target polynucleotide. In some examples, the donor polynucleotide disrupts a splicing site. The disruption may be achieved by inserting the polynucleotide to a splicing site and/or introducing one or more mutations to the splicing site. In certain examples, the donor polynucleotide may restore a splicing site. For example, the polynucleotide may comprise a splicing site sequence.

[0262] The donor polynucleotide to be inserted may have a size from 10 bases to 50 kb in length, e.g., from 50 to 40 kb, from 100 to 30 kb, from 100 bases to 300 bases, from 200 bases to 400 bases, from 300 bases to 500 bases, from 400 bases to 600 bases, from 500 bases to 700 bases, from 600 bases to 800 bases, from 700 bases to 900 bases, from 800 bases to 1000 bases, from 900 bases to 1100 bases, from 1000 bases to 1200 bases, from 1100 bases to 1300 bases, from 1200 bases to 1400 bases, from 1300 bases to 1500 bases, from 1400 bases to 1600 bases, from 1500 bases to 1700 bases, from 600 bases to 1800 bases, from 1700 bases to 1900 bases, from 1800 bases to 2000 bases, from 1900 bases to 2100 bases, from 2000 bases to 2200 bases, from 2100 bases to 2300 bases, from 2200 bases to 2400 bases, from 2300 bases to 2500 bases, from 2400 bases to 2600 bases, from 2500 bases to 2700 bases, from 2600 bases to 2800 bases, from 2700 bases to 2900 bases, or from 2800 bases to 3000 bases in length.

[0263] The components in the systems herein may comprise one or more mutations that alter their (e.g., the transposase(s)) binding affinity to the donor polynucleotide. In some examples, the mutations increase the binding affinity between the transposase(s) and the donor polynucleotide. In certain examples, the mutations decrease the binding affinity between the transposase(s) and the donor polynucleotide. The mutations may alter the activity of the Cas and/or transposase(s).

[0264] In certain embodiments, the systems disclosed herein are capable of unidirectional insertion, that is the system inserts the donor polynucleotide in only one orientation.

[0265] Delivery mechanisms for CAST systems includes those discussed above for CRISPR-Cas systems.

CRISPR-Cas Topoisomerase Systems

[0266] The Cas polypeptides disclosed herein may be connected to one or more topoisomerase domains. In one embodiment, an engineered system for modifying a target polynucleotide comprising: a Cas polypeptide, a topoisomerase domain linked to or otherwise capable of associating with the Cas polypeptide, and a guide RNA capable of forming a complex with the Cas polypeptide and directing sequence-specific binding of the complex to a target sequence on a target polynucleotide; and a nucleic acid template comprising or encoding a donor polynucleotide to be inserted to a target sequence of the target polynucleotide.

[0267] Topoisomerases are a class of enzymes that modify the topological state of DNA via the breakage and rejoining of nucleic acid strands. In some cases, a topoisomerase may be a DNA topoisomerase, which is an enzyme that controls and alters the topologic states of DNA during transcription, and catalyzes the transient breaking and rejoining of a single strand of DNA which allows the strands to pass through one another, thus altering the topology of DNA.

[0268] In one embodiment, the topoisomerase domain is capable of ligating the donor polynucleotide with the target polynucleotide. The ligation may be achieved by sticky end or blunt end ligation. In an example, the donor polynucleotide may comprise an overhang comprising a sequence complementary to a region of the target polynucleotide. Examples of ligating the donor polynucleotide with the target polynucleotide include those of TOPO cloning, e.g., those described in “The Technology Behind TOPO Cloning,” at www.thermofisher.com/us/en/home/life-science/cloning/topo/topo-resources/the-technology-behind-topo-cloning.html.

[0269] In one embodiment, the topoisomerase domain may be associated with the donor polynucleotide. For example, the topoisomerase domain is covalently linked to the donor polynucleotide.

[0270] In one embodiment, a topoisomerase domain may be provided together with, e.g., associated (e.g., fused) with a Cas polypeptide). Alternatively or additionally, the topoisomerase domain may be on a molecule different from the Cas polypeptide. In some cases, the topoisomerase domain may be associated with a donor polynucleotide. For example, the topoisomerase domain may be pre-loaded covalently with a donor DNA molecule. Such design may allow for efficient ligation of only a specific cargo. The topoisomerase domain may ligate the donor polynucleotide (e.g., a DNA molecule) to a target site on a target polynucleotide (e.g., a free double-stranded DNA end). In one embodiment, the donor polynucleotide may have an overhang that comprises a sequence complementary to a region of the target polynucleotide. For example, the overhang may invade into the target polynucleotide at a cut site generated by the Cas12b polypeptide.

[0271] Examples of topoisomerases include type I, including type IA and type IB topoisomerases, which cleave a single strand of a double-stranded nucleic acid molecule, and type II topoisomerases (e.g., gyrases), which cleave both strands of a double-stranded nucleic acid molecule.

[0272] Type IA and IB topoisomerases cleave one strand of a double-stranded nucleic acid molecule. In some examples, the cleavage of a double-stranded nucleic acid molecule by type IA topoisomerases generates a 5' phosphate and a 3' hydroxyl at the cleavage site, with the type IA topoisomerase covalently binding to the 5' terminus of a cleaved strand. Cleavage of a double-stranded nucleic acid molecule by type IB topoisomerases may generate a 3' phosphate and a 5' hydroxyl at the cleavage site, with the type IB topoisomerase covalently binding to the 3' terminus of a cleaved strand.

[0273] Examples of Type IA topoisomerases include *E. coli* topoisomerase I, *E. coli* topoisomerase III, eukaryotic topoisomerase II, archeal reverse gyrase, yeast topoisomerase III, *Drosophila* topoisomerase III, human topoisomerase III, *Streptococcus pneumoniae* topoisomerase III, and the like, including other type IA topoisomerases. A DNA-protein adduct is formed with the enzyme covalently binding to the 5'-thymidine residue, with cleavage occurring between the two thymidine residues.

[0274] Examples of Type IB topoisomerases include the nuclear type I topoisomerases present in all eukaryotic cells and those encoded by *Vaccinia* and other cellular poxviruses. The eukaryotic type IB topoisomerases are exemplified by those expressed in yeast, *Drosophila* and mammalian cells, including human cells. Viral type IB topoisomerases are exemplified by those produced by the vertebrate poxviruses (*Vaccinia*, Shope fibroma virus, ORF virus, fowlpox virus, and molluscum contagiosum virus), and the insect poxvirus (*Amsacta moorei entomopoxvirus*).

[0275] Examples of Type II topoisomerases include, bacterial gyrase, bacterial DNA topoisomerase IV, eukaryotic DNA topoisomerase II, and T-even phage encoded DNA topoisomerases. Type II topoisomerases may have both cleaving and ligating activities. Substrate double-stranded nucleic acid molecules of type II topoisomerase can be prepared such that the type II topoisomerase can form a covalent linkage to one strand at a cleavage site. For example, calf thymus type II topoisomerase can cleave a substrate ds nucleic acid molecule containing a 5' recessed topoisomerase recognition site positioned three nucleotides from the 5' end, resulting in dissociation of the three nucleic acid molecule 5' to the cleavage site and covalent binding of the topoisomerase to the 5' terminus of the ds nucleic acid molecule. Furthermore, upon contacting such a type II topoisomerase-charged ds nucleic acid molecule with a second nucleic acid molecule containing a 3' hydroxyl group, the type II topoisomerase can ligate the sequences together, and then is released from the recombinant nucleic acid molecule.

[0276] In some examples, the topoisomerase is DNA topoisomerase I, e.g., a *Vaccinia* virus topoisomerase I. The topoisomerase may be pre-loaded with a donor polynucleotide. The *Vaccinia* virus topoisomerase may need a target comprising a 5' -OH group.

CRISPR-Cas Retrotransposon Systems

[0277] The systems and compositions herein may comprise a Cas polypeptide or, one or more e RNAs, and one or more components of a retrotransposon, e.g., a non-LTR retrotransposon. The one or more components of a retrotransposon include a retrotransposon protein and retrotransposon RNA. The systems and compositions may be used to insert a donor polynucleotide to a target polynucleotide. The systems and compositions may further comprise a donor polynucleotide.

[0278] In some examples, the present disclosure provides an engineered, non-naturally occurring composition comprising: a Cas polypeptide, a non-LTR retrotransposon protein

associated with or otherwise capable of forming a complex with the Cas polypeptide; a single guide RNA capable of forming a complex with the Cas polypeptide and directing site-specific binding to a target sequence of a target polynucleotide. The composition may further comprise a donor construct comprising a donor polynucleotide for insertion to the target polynucleotide and located between two binding elements capable of forming a complex with the non-LTR retrotransposon protein. In some cases, the Cas polypeptide is engineered to have nickase activity.

[0279] In some examples, the Cas polypeptide is fused to the N-terminus of the non-LTR retrotransposon protein. In some examples, the Cas polypeptide is fused to the C-terminus of the non-LTR retrotransposon protein.

[0280] The guides may direct the fusion protein to a target sequence 5' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site. The guides may direct the fusion protein to a target sequence 3' of the targeted insertion site, and wherein the Cas polypeptide generates a double-strand break at the targeted insertion site.

[0281] The donor polynucleotide may further comprise a polymerase processing element to facilitate 3' end processing of the donor polynucleotide sequence. The polymerase may be a DNA polymerase, e.g., DNA polymerase I. In some examples, the polymerase may be an RNA polymerase.

[0282] In some examples, the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both. In some examples, the homology region is from 1 to 50, from 5 to 30, from 8 to 25, e.g., 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, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 base pairs in length.

[0283] Native or wild-type non-LTR retrotransposons encode the protein machinery necessary for their self-mobilization. The non-LTR retrotransposon element comprises a DNA element integrated into a host genome. This DNA element may encode one or two open reading frames (ORFs). For example, the R2 element of *Bombyx mori* encodes a single ORF containing reverse transcriptase (RT) activity and a restriction enzyme-like (REL) domain. L1 elements encode two ORFs, ORF1 and ORF2. ORF1 contains a leucine zipper domain involved in protein-protein interactions and a C-terminal nucleic acid binding domain. ORF2 has a N-

terminal apurinic/apyrimidinic endonuclease (APE), a central RT domain, and a C-terminal cysteine histidine rich domain. An example replicative cycle of a non-LTR retrotransposon may comprise transcription of the full-length retrotransposon element to generate an mRNA active element (retrotransposon RNA). The active element mRNA is translated to generate the encoded retrotransposon proteins or polypeptides. A ribonucleoprotein complex comprising the active element and retrotransposon protein or polypeptide is formed and this RNP facilitates integration of the active element into the genome. The RNA-transposase complex nicks the genome. The 3' end of the nicked DNA serves as a primer to allow the reverse transcription of the transposon RNA into cDNA. Fourth, the transposase proteins integrate the cDNA into the genome.

[0284] Elements of these systems may be engineered to work within the context of the invention. For example, a non-LTR retrotransposon polypeptide may be fused to a site-specific nuclease. The binding elements that allow a non-LTR retrotransposon polypeptide to bind to the native retrotransposon DNA element, may be engineered into a donor construct to facilitate entry of a donor polynucleotide sequence into a target polypeptide.

[0285] In the present invention the protein component of the non-LTR retrotransposon may be connected to or otherwise engineered to form a complex with a site-specific nuclease. The retrotransposon RNA may be engineered to encode a donor polynucleotide sequence. Thus, in certain example embodiments, the Cas polypeptide nuclease, via formation of a Cas polypeptide nuclease complex with a guide RNA, directs the retrotransposon complex (e.g. the retrotransposon polypeptide(s) and retrotransposon RNA to a target sequence in a target polynucleotide, where the retrotransposon RNP complex facilitates integration of the donor polynucleotide sequence into the target polynucleotide. Accordingly, the one or more non-LTR retrotransposon components may comprise retrotransposon polypeptides, or function domains thereof, that facilitate binding of the retrotransposon RNA, reverse transcription of the retrotransposon RNA into cDNA, and/or integration of the donor polynucleotide into the target polynucleotide, as well as retrotransposon RNA elements modified to encode the donor polynucleotide sequence.

[0286] Examples of non-LTR retrotransposons include CRE, R2, R4, L1, RTE, Tad, R1, LOA, I, Jockey, CR1. In one example, the non-LTR retrotransposon is R2. In another example, the non-LTR retrotransposon is L1. Examples of non-LTR retrotransposons may include those described in Christensen SM et al., RNA from the 5' end of the R2 retrotransposon controls R2

protein binding to and cleavage of its DNA target site, *Proc Natl Acad Sci U S A.* 2006 Nov 21;103(47):17602-7; Eickbush TH et al, Integration, Regulation, and Long-Term Stability of R2 Retrotransposons, *Microbiol Spectr.* 2015 Apr;3(2):MDNA3-0011-2014. doi: 10.1128/microbiolspec.MDNA3-0011-2014; Han JS, Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions, *Mob DNA.* 2010 May 12;1(1):15. doi: 10.1186/1759-8753-1-15; Malik HS et al., The age and evolution of non-LTR retrotransposable elements, *Mol Biol Evol.* 1999 Jun;16(6):793-805, which are incorporated by reference herein in their entireties.

[0287] Examples of the non-LTR retrotransposon polypeptides also include R2 from *Clonorchis sinensis*, or *Zonotrichia albicollis*.

[0288] A non-LTR retrotransposon may comprise multiple retrotransposon polypeptides or polynucleotides encoding same. In one embodiment, the retrotransposon polypeptides may form a complex. For example, a non-LTR retrotransposon is a dimer, e.g., comprising two retrotransposon polypeptides forming a dimer. The dimer subunits may be connected or form a tandem fusion. A Cas12b polypeptide nuclease may be associated with (e.g., connected to) one or more subunits of such complex. In some examples, the non-LTR retrotransposon is a dimer of two retrotransposon polypeptides; one of the retrotransposon polypeptides comprises nuclease or nickase activity and is connected with a Cas12b polypeptide nuclease.

[0289] The retrotransposon polypeptides may comprise one or more modifications to, for example, enhance specificity or efficiency of donor polynucleotide recognition, target-primed template recognition (TPTR). The retrotransposon polypeptides may also comprise one or more truncations or excisions to remove domains or regions of wild-type protein to arrive at a minimal polypeptide that retain donor polynucleotide recognition and TPTR. In some example embodiments, the native endonuclease activity may be mutated to eliminate endonuclease activity.

[0290] In certain example embodiments, the modifications or truncations of the non-LTR retrotransposon peptide may be in a zinc finger region, a Myb region, a basic region, a reverse transcriptase domain, a cysteine-histidine rich motif, or an endonuclease domain.

[0291] A non-LTR retrotransposon may comprise polynucleotide encoding one or more retrotransposon RNA molecules. The polynucleotide may comprise one or more regulatory elements. The regulatory elements may be promoters. The regulatory elements and promoters

on the polynucleotides include those described throughout this application. For example, the polynucleotide may comprise a pol2 promoter, a pol3 promoter, or a T7 promoter.

[0292] In some cases, the polynucleotide encodes a retrotransposon RNA with at least a portion of its sequence complementary to a target sequence. For example, the 3' end of the retrotransposon RNA may be complementary to a target sequence. The RNA may be complementary to a portion of a nicked target sequence. In one embodiment, a retrotransposon RNA may comprise one or more donor polynucleotides. In certain cases, a retrotransposon RNA may encode one or more donor polynucleotides.

[0293] A retrotransposon RNA may be capable of binding to a retrotransposon polypeptide. Such retrotransposon RNA may comprise one or more elements for binding to the retrotransposon polypeptide. Examples of binding elements include hairpin structures, pseudoknots (e.g., a nucleic acid secondary structure containing at least two stem-loop structures in which half of one stem is intercalated between the two halves of another stem), stem loops, and bulges (e.g., unpaired stretches of nucleotides located within one strand of a nucleic acid duplex). In certain examples, the retrotransposon RNA comprises one or more hairpin structures. In some examples, the retrotransposon RNA comprises one or more pseudoknots. In certain examples, a retrotransposon RNA comprises a sequence encoding a donor polynucleotide and one or more binding elements for forming a complex with the retrotransposon polypeptide. The binding elements may be located on the 5' end or the 3' end.

[0294] In one embodiment, a retrotransposon RNA comprises a region capable of hybridizing with an overhang of a target polynucleotide at the target site. The overhang may be a stretch of single-stranded DNA. The overhang may function as a primer for reverse transcription of at least a portion of the retrotransposon RNA to a cDNA. In some cases, a region of the cDNA may be capable of hybridizing a second overhang of the target polynucleotide. The second overhang may function as a primer for the synthesis of a second strand to generate a double-stranded cDNA. The cDNA may comprise a donor polynucleotide sequence. The two overhangs may be from different strands of the target polynucleotide.

CRISPR-Cas Reverse Transcriptase Domain

[0295] The Cas polypeptides disclosed herein may be linked to one or more reverse transcriptase domains. In one embodiment, the systems comprise an engineered system for modifying a target polynucleotide comprising: an Cas polypeptide (or a nickase or dead variant thereof) a reverse transcriptase (RT) domain; a RNA template comprising or encoding a donor

polynucleotide to be inserted to a target sequence of the target polynucleotide; and a guide RNA molecule (i.e., a naturally single guide RNA molecule comprising a scaffold for reprogramming).

[0296] The reverse transcriptase may generate single-strand DNA based on the RNA template. The single-strand DNA may be generated by a non-retron, retron, or diversity generating retroelement (DGR). In some examples, the single-strand DNA may be generated from a self-priming RNA template. A self-priming RNA template may be used to generate a DNA without the need of a separate primer.

[0297] A reverse transcriptase domain may be a reverse transcriptase or a fragment thereof. A wide variety of reverse transcriptases (RT) may be used in alternative embodiments of the present invention, including prokaryotic and eukaryotic RT, provided that the RT functions within the host to generate a donor polynucleotide sequence from the RNA template. If desired, the nucleotide sequence of a native RT may be modified, for example using known codon optimization techniques, so that expression within the desired host is optimized. A reverse transcriptase (RT) is an enzyme used to generate complementary DNA (cDNA) from an RNA template, a process termed reverse transcription. Reverse transcriptases are used by retroviruses to replicate their genomes, by retrotransposon mobile genetic elements to proliferate within the host genome, by eukaryotic cells to extend the telomeres at the ends of their linear chromosomes, and by some non-retroviruses such as the hepatitis B virus, a member of the Hepadnaviridae, which are dsDNA-RT viruses. Retroviral RT has three sequential biochemical activities: RNA-dependent DNA polymerase activity, ribonuclease H, and DNA-dependent DNA polymerase activity. Collectively, these activities enable the enzyme to convert single-stranded RNA into double-stranded cDNA. In an embodiment, the RT domain of a reverse transcriptase is used in the present invention. The domain may include only the RNA-dependent DNA polymerase activity. In some examples, the RT domain is non-mutagenic, i.e., does not cause mutation in the donor polynucleotide (e.g., during the reverse transcriptase process). In some examples, the RT domain may be non-retron RT, e.g., a viral RT or a human endogenous RT. In some examples, the RT domain may be retron RT or DGRs RT. In some examples, the RT may be less mutagenic than a counterpart wildtype RT. In one embodiment, the RT herein is not mutagenic.

Retrons

[0298] In an embodiment, a donor template for homologous recombination is generated by use of a self-priming RNA template for reverse transcription. A non-limiting example of a self-priming reverse transcription system is the retron system. By the term “retron” it is meant a genetic element which encodes components enabling the synthesis of branched RNA-linked single stranded DNA (msDNA) and a reverse transcriptase. Retrons which encode msDNA are known in the art, for example, but not limited to U.S. Pat. No. 6,017,737; U.S. Pat. No. 5,849,563; U.S. Pat. No. 5,780,269; U.S. Pat. No. 5,436,141; U.S. Pat. No. 5,405,775; U.S. Pat. No. 5,320,958; CA 2,075,515; all of which are herein incorporated by reference.

[0299] In an embodiment, the reverse transcriptase domain is a retron RT domain. In an embodiment, the RNA template encodes a retron RNA template that is recognized and reverse transcribed by the retron reverse transcriptase domain. Conserved across many bacterial species, retrons are highly efficient reverse transcription systems of relatively unknown function. The retron system consists of the retron RT protein, as well as the msr and msd transcripts, which function as the primer and template sequences, respectively. All components of the retron system are expressed from a single open reading frame as a single transcript including the msr-msd and encoding the retron RT protein (Lampson, et al., 2005, Retrons, msDNA, and the bacterial genome. Cytogenet Genome Res 110:491–499). The msr element ORF of a retron provides for the RNA portion of the msDNA molecule, while the msd element ORF provides for the DNA portion of the msDNA molecule. The primary transcript from the msr-msd region is thought to serve as both a template and a primer to produce the msDNA. Synthesis of msDNA is primed from an internal rG residue of the RNA transcript using its 2'-OH group. Modification of msd, or msr may also be made to permit insertion of a RNA template encoding a donor polynucleotide within the msd without altering the functioning of or the production of msDNA. The RNA template encoding a donor polynucleotide sequence may be any length but is preferably less than about 5 kb nucleotides, or also less than about 2 kb, or also less than 500 bases, provided that an msDNA product is produced.

Diversity Generating Retroelement Systems

[0300] In an embodiment, the one or more functional domains may be a diversity generating retroelement(s) (e.g., DGR described in US20100041033A1). In one embodiment, the DGR may insert a donor polynucleotide with its homing mechanism. For example, the DGR may be associated with a catalytically inactive Cas12b protein (e.g., a dead Cas12b), and

integrate the single-strand DNA using a homing mechanism. In some examples, the DGR may be less mutagenic than a counterpart wild type DGR. In some examples, the DGR is not error-prone. In one embodiment, the DGR herein is not mutagenic. The non-mutagenic DGR may be a mutant of a wild type DGR. As used herein, the term “DGR” encompasses both diversity generating retroelement polynucleotides and proteins encoded by diversity generating retroelement polynucleotides. In some examples, DGR may be proteins encoded by diversity generating retroelement polynucleotides having reverse transcriptase activity. In some examples, DGR may be proteins encoded by diversity generating retroelement polynucleotides having reverse transcriptase activity and integrase activity. In some cases, the template or donor polynucleotide may be encoded by a diversity generating retroelement polynucleotide. In certain cases, the template may be a polynucleotide different from the diversity generating retroelement polynucleotide, e.g., provided as a separate construct or molecule.

[0301] In one embodiment, the DGR herein may also include a Group II intron (and any proteins and polynucleotides encoded), which are mobile ribozymes that self-splice from precursor RNAs to yield excised intron lariat RNAs, which then invade new genomic DNA sites by reverse splicing. Examples of Group II intron include those described in Lambowitz AM et al., Group II Introns: Mobile Ribozymes that Invade DNA, Cold Spring Harb Perspect Biol. 2011 Aug; 3(8): a003616.

[0302] In one embodiment, the diversity-generating retroelements (DGRs) are genetic elements that can produce targeted, massive variations in the genomes that carry these elements. In one embodiment, the DGR systems rely on error-prone reverse transcriptases to produce mutagenized cDNA (containing A-to-N mutations) from a template region (TR), to replace a segment called a variable region (VR) that is similar to the TR region—this process is called mutagenic retrohoming (see, e.g., Sharifi and Ye, MyDGR: a server for identification and characterization of diversity-generating retroelements. Nucleic Acids Res. 2019 Jul 2; 47(W1): W289–W294). DGRs may include a unique family of retroelements that generate sequence diversity of DNA. They exist widely in bacteria, archaea, phage and plasmid, and benefit their hosts by introducing variations and accelerating the evolution of target proteins (see, e.g., Yan et al., Discovery and characterization of the evolution, variation and functions of diversity-generating retroelements using thousands of genomes and metagenomes. BMC Genomics. 2019; 20: 595). The first DGR was discovered in a *Bordetella* phage, BPP-1. *Bordetella* causes the respiratory infection in humans and many other mammals, controlled by

the BvgAS signal transduction system. The surface of *Bordetella* is highly variable owing to the dynamic gene expression in the infectious cycle. The invasion of BPP-1 to *Bordetella* relies on the phage tail fiber protein Mtd. With the process of mutagenic reverse transcription and cDNA integration, DGR may introduce multiple nucleotide substitutions to Mtd gene and generates different receptor-binding molecules, thus making BPP-1 the ability to invade *Bordetellae* with diverse cell surfaces.

[0303] The systems may be used to generate an ssDNA donor using a retron- or DGR RT, which is then integrated by homologous recombination upon target cleavage or nicking using a Cas12b polypeptide. In one embodiment, the systems may comprise DGRs and/or Group-II intron reverse transcriptases. The homing mechanism of DGRs or Group-II introns may be used in modifying a target polynucleotide. The DGRs or Group-II introns reverse transcriptase may be guided to a target polynucleotide by tethering to a dead Cas12b nuclease, TALE, or ZF protein. In another embodiment, a non-retron/DGR reverse transcriptase (e.g. a viral RT) may be used for generating cDNA off of a self-priming RNA. In one embodiment, a ssDNA may be generated by an RT, but integrate it using a dead Cas12b polypeptide, creating an accessible R-loop instead of nicking/cleaving.

CRISPR-Cas Phosphatase Systems

[0304] In one example embodiment, a CRISPR-Cas phosphatase systems comprises a Cas polypeptide as disclosed herein, one or more guide RNAs, and phosphatase linked to or otherwise capable of associating with the Cas polypeptide. The systems herein may further comprise a phosphatase domain. A phosphatase is an enzyme capable of removing a phosphate group from a molecule e.g., a nucleic acid such as DNA. Examples of phosphatases include calf intestinal phosphatase, shrimp alkaline phosphatase, Antarctic phosphatase, and APEX alkaline phosphatase.

[0305] In some examples, the 5' -OH group of in the target polynucleotide may be generated by a phosphatase. A topoisomerase compatible with a 5' phosphate target may be used to generate stable loaded intermediates. In some cases, a Cas polypeptide that leaves a 5' OH after cleaving the target polynucleotide may be used. In some cases, the phosphatase domain may be associated with (e.g., fused to) the Cas protein. The phosphatase domain may be capable of generating a -OH group at a 5' end of the target polynucleotide. The phosphatase may be delivered separated from other components in the system, e.g., as a separate protein, on a separate vector from other components.

CRISPR-Cas Polymerase Systems

[0306] In one example embodiment, a CRISPR-Cas phosphatase systems comprises a Cas polypeptide as disclosed herein, one or more guide RNAs, and polymerases linked to or otherwise capable of associating with the Cas polypeptide. A polymerase refers to an enzyme that synthesizes chains of nucleic acids. The polymerase may be a DNA polymerase or an RNA polymerase.

[0307] In one embodiment, the systems comprise an engineered system for modifying a target polynucleotide comprising: an Cas polypeptide; a DNA polymerase domain; and a DNA template comprising a donor polynucleotide to be inserted to a target sequence of the target polynucleotide. In some examples, two or more of: the Cas protein; DNA polymerase domain; and DNA template may form a complex. In some examples, two or more of: the Cas protein; DNA polymerase domain; are comprised in a fusion protein. For example, the Cas polypeptide and DNA polymerase domain may be comprised in a fusion protein.

[0308] In one embodiment, the systems may comprise a Cas12b polypeptide (or variant thereof such as a dCas12b polypeptide or Cas12b polypeptide nickase) and a DNA polymerase (e.g. phi29, T4, T7 DNA polymerase). The systems may further comprise a single-stranded DNA or double-stranded DNA template. The DNA template may comprise i) a first sequence homologous to a target site of the Cas12b protein on the target polynucleotide, and/or ii) a second sequence homologous to another region of the target polynucleotide. In one embodiment, the template may be a synthetic single-stranded or PCR-generated DNA molecule, (optionally end-protected by modified nucleotides), or a viral genome (e.g. AAV). In another embodiment, the template is generated using a reverse transcriptase. When the system is delivered into a cell, an endogenous DNA polymerase in the cell may be used. Alternatively or additionally, an exogenous DNA polymerase may be expressed in the cell.

[0309] The DNA template may be end-protected by one or more modified nucleotides, or comprises a portion of a viral genome. In some embodiment, the DNA template comprises LNA or other modifications (e.g., at the 3' end). The presence of LNA and/or the modifications may lead to more efficient annealing with the 3' flap generated by Cas12b polypeptide cleavage.

[0310] Examples of DNA polymerase include Taq, Tne (exo ⁻), Tma (exo ⁻), Pfu (exo ⁻), Pwo (exo ⁻), Thermoanaerobacter thermohydrosulfuricus DNA polymerase, Thermococcus litoralis DNA polymerase I, E. coli DNA polymerase I, Taq DNA polymerase I, Tth DNA

polymerase I, *Bacillus stearothermophilus* (Bst) DNA polymerase I, *E. coli* DNA polymerase III, bacteriophage T5 DNA polymerase, bacteriophage M2 DNA polymerase, bacteriophage T4 DNA polymerase, bacteriophage T7 DNA polymerase, bacteriophage phi29 DNA polymerase, bacteriophage PRD1 DNA polymerase, bacteriophage phi15 DNA polymerase, bacteriophage phi21 DNA polymerase, bacteriophage PZE DNA polymerase, bacteriophage PZA DNA polymerase, bacteriophage Nf DNA polymerase, bacteriophage M2Y DNA polymerase, bacteriophage B103 DNA polymerase, bacteriophage SF5 DNA polymerase, bacteriophage GA-1 DNA polymerase, bacteriophage Cp-5 DNA polymerase, bacteriophage Cp-7 DNA polymerase, bacteriophage PR4 DNA polymerase, bacteriophage PR5 DNA polymerase, bacteriophage PR722 DNA polymerase and bacteriophage L17 DNA polymerase.

CRISPR-Cas Ligase Systems

[0311] In general, the systems comprise a Cas polypeptide and a ligase associated with the Cas polypeptide. The Cas12b polypeptide may be recruited to the target sequence by a guide or scaffold sequence RNA, and generate a break on the target sequence. The guide sequence RNA may further comprise a template sequence with desired mutations or other sequence elements. The template sequence may be ligated to the target sequence to introduce the mutations or other sequence elements to the nucleic acid molecule. The Cas12b polypeptide may be a nickase that generates a single-strand break on nucleic acid molecule, and the ligase may be a single-strand DNA ligase. In one embodiment, the systems comprise a pair of Cas12b polypeptide-ligases complexes, with two distinct guide sequences. Each Cas12b polypeptide-ligase complex, can target one strand of a double-stranded polynucleotide, and work together to effectively modify the sequence of the double-stranded polynucleotides.

[0312] In some examples, the Cas12b polypeptide is associated with a ligase or functional fragment thereof. The ligase may ligate a single-strand break (a nick) generated by the Cas12b polypeptide. In certain cases, the ligase may ligate a double-strand break generated by the Cas12b polypeptide. In certain examples, the Cas12b polypeptide is associated with a reverse transcriptase or functional fragment thereof.

[0313] The present invention further provides systems and methods of modifying a nucleic acid sequence using a pair of distinct Cas polypeptide-ligase guide RNA complexes, said systems and methods comprising: (a) an engineered Cas polypeptide connected to or complexed with a ligase; (b) two distinct guide RNA complexed with such Cas polypeptide-ligase protein complex to form a first and a second distinct Cas-ligase guide RNA complexes;

(c) the first Cas-ligase–guide or scaffold sequence RNA complex binding to one strand of a target double-stranded polynucleotide sequence, and the second Cas polypeptide-ligase- guide RNA complex binding to another strand of the target double-stranded polynucleotide sequence; (d) upon binding of the said complexes to the locus of interest the effector protein induces the modification of the sequences associated with or at the target locus of interest, whereby the two Cas polypeptide -ligase-guide RNA complexes work together on different strands of the double-stranded target sequence and modify the sequence.

[0314] One of the advantages of using such a “pair” of Cas polypeptide-ligase- guide RNA complexes includes high efficiency in modifying the sequence associated with or at the locus of interest of target double-stranded polynucleotides.

[0315] In one embodiment, the Cas polypeptide can be a nickase. In a preferred embodiment, a ligase is linked to the Cas polypeptide. The ligase can ligate the donor sequence to the target sequence. The ligase can be a single-strand DNA ligase or a double-strand DNA ligase. The ligase can be fused to the carboxyl-terminus of a Cas polypeptide, or to the amino-terminus of a Cas polypeptide.

[0316] As used herein the term “ligase” refers to an enzyme, which catalyzes the joining of breaks (e.g., double-stranded breaks or single-stranded breaks (“nicks”) between adjacent bases of nucleic acids. For example, a ligase may be an enzyme capable of forming intra- or inter-molecular covalent bonds between a 5' phosphate group and a 3' hydroxyl group. The term “ligate” refers to the reaction of covalently joining adjacent oligonucleotides through formation of an internucleotide linkage.

[0317] DNA ligases fall into two general categories: ATP-dependent DNA ligases (EC 6.5.1.1), and NAD (+) dependent DNA ligases (EC 6.5.1.2). NAD (+) dependent DNA ligases are found only in bacteria (and some viruses) while ATP-dependent DNA ligases are ubiquitous. The ATP-dependent DNA ligases can be divided into four classes: DNA ligase I, II, III, and IV. DNA ligase I links Okazaki fragments to form a continuous strand of DNA; DNA ligase II is an alternatively spliced form of DNA ligase III, found only in non-dividing cells; DNA ligase III is involved in base excision repair; and DNA ligase IV is involved in the repair of DNA double-strand breaks by non-homologous end joining (NHEJ). Amongst all ligases, there are two types of prokaryotic and one type of eukaryotic ligases that are particularly well suited for facilitating the blunt-ended, double-stranded DNA ligation: Prokaryotic DNA ligases (T3 and T4) and Eukaryotic DNA ligase (Ligase 1).

[0318] In some cases, the ligase is specific for double-stranded nucleic acids (e.g., dsDNA, dsRNA, RNA/DNA duplex). An example of a ligase specific for double-stranded DNA and DNA/RNA hybrids is T4 DNA ligase. In some cases, the ligase is specific for single-stranded nucleic acids (e.g., ssDNA, ssRNA). An example of such ligase is CircLigase II. In some cases, the ligase is specific for RNA/DNA duplexes. In some cases, the ligase is able to work on single-stranded, double-stranded, and/or RNA/DNA nucleic acids in any combination.

[0319] In some cases, the ligase may be a pan-ligase, which is a single ligase with the ability to ligate both DNA and RNA targets. The ligase may be specific for a target (e.g., DNA-specific or RNA-specific). In some cases, the ligase may be a dual ligase system that include DNA-specific, RNA-specific, and/or pan-ligases, in any combination.

[0320] Examples of ligases that can be used with the disclosure include T4 DNA Ligase, T3 DNA Ligase, T7 DNA Ligase, E. coli DNA Ligase, HiFi Taq DNA Ligase, 9° N™ DNA Ligase, Taq DNA Ligase, SplintR® Ligase (also known as PBCV-1 DNA Ligase or Chlorella virus DNA Ligase), Thermostable 5' AppDNA/RNA Ligase, T4 RNA Ligase, T4 RNA Ligase 2, T4 RNA Ligase 2 Truncated, T4 RNA Ligase 2 Truncated K227Q, T4 RNA Ligase 2, Truncated KQ, RtcB Ligase (joins single stranded RNA with a 3"-phosphate or 2',3'-cyclic phosphate to another RNA), CircLigase II, CircLigase ssDNA Ligase, CircLigase RNA Ligase, or Ampligase® Thermostable DNA Ligase, NAD-dependent ligases including Taq DNA ligase, Thermus filiformis DNA ligase, Escherichia coli DNA ligase, Tth DNA ligase, Thermus scotoductus DNA ligase (I and II), thermostable ligase, Ampligase thermostable DNA ligase, VanC-type ligase, 9° N DNA Ligase, Tsp DNA ligase, and novel ligases discovered by bioprospecting; ATP-dependent ligases including T4 RNA ligase, T4 DNA ligase, T3 DNA ligase, T7 DNA ligase, Pfu DNA ligase, DNA ligase I, DNA ligase III, DNA ligase IV, and novel ligases discovered by bioprospecting, and wild-type, mutant isoforms, and genetically engineered variants thereof. In a particular example, the ligase is a

[0321] In one embodiment, the examples of the ligases include those used in sequencing by synthesis or sequencing by ligation reactions.

[0322]

CRISPR-Cas Helitron Systems

[0323] The systems and compositions herein may comprise an Cas polypeptide as disclosed herein, one or more guide RNAs, and one or more components of a helitron. The

systems and compositions may be used to insert a donor polynucleotide to a target polynucleotide. The systems and compositions may further comprise a donor polynucleotide.

[0324] The term “helitron”, as used herein, refers to a polynucleotide (or nucleic acid segment), recognized as a transposon that captures and mobilizes gene fragments in eukaryotes. The term “helitron” as used herein refers to transposase that comprises an endonuclease domain and a C-terminal helicase domain. Helitrons are rolling-circle RNA transposons. In one embodiment, the helitron encodes a 1400 to about 2000 amino acid, or about 1800 amino acid multidomain transposase. In embodiments, the helitron comprises a hairpin near the 3′ end to function as a transposition terminator. In embodiments, the transposon comprises a RepHel motif comprising a replication initiator (Rep) and a DNA helicase (hel) domain. See, Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* 3, 893–926 (2015). In embodiments, the helitron comprises a Rep nuclease domain and C-terminal helicase domain and inserts between an AT dinucleotide in single strand DNA. In an aspect, the C-terminal helicase unwinds the DNA in a 5′ to 3′ direction. The HUH nuclease domain may comprise one or two active site tyrosine residues, in embodiments, is a 2 Tyrosine (Y2) HUH endonuclease domain. Helitrons can encompass helitron, proto-helitron and helitron2 type proteins, structures of which can be as described in Thomas et al., 2015 at Figures 1 and 3, incorporated specifically by reference. Particular organisms in which the helitron or helitrons have been found can include those in Table 1 of Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* 3, 893–926 (2015), incorporated herein by reference. Similarly, helitrons can be identified based at least in part on the Rep motif, and conserved residues in the helitrons, and according to the alignment sequence of Figure 2 of Thomas J. & Pritham E. J. Helitrons, the eukaryotic rolling-circle transposable elements. *Microbiol. Spectr.* 3, 893–926 (2015), specifically incorporated herein by reference.

[0325] The expression “helitron reaction” used herein refers to a reaction wherein a transposase inserts a donor polynucleotide sequence in or adjacent to an insertion site on a target polynucleotide. The insertion site may contain a sequence or secondary structure recognized by the helitron and/or an insertion motif sequence in the target polynucleotide into which the donor polynucleotide sequence may be inserted.

[0326] As described in Grabundzija 2018, the helitron terminal sequences contains a distinct ~150 base pairs (bp) long sequence with an absolutely conserved dinucleotide at the

end of left terminal sequence (LTS), and a tetranucleotide at the end of right terminal sequence (RTS) which is preceded by a palindromic sequence that can form a hairpin structure. Grabundzija et al., Nat. Commun. 2018; 9: 1278; doi:10.1035/s41467-018-03688-w.

[0327] The helitron end sequences may be responsible for identifying the donor polynucleotide for transposition. The helitron end sequences may be the DNA sequences used to perform a transposition reaction, the end sequences may be referred to herein as right terminal sequences and left terminal sequence. The donor polynucleotide can be configured to comprise a first and second helitron recognition sequence that are at least 80%, 85%, 90%, 95% 96%, 97%, 98%, 99% or 100% complementary to a left terminal sequence and/or a right terminal sequence of a polynucleotide encoding the helitron polypeptide.

[0328] In an aspect, the palindromic sequence may be located upstream of the right terminal sequence, for example, about 5, 10, 15, 20, 25, 30, 35 nucleotides upstream of the right terminal sequence end, or about 10 to 15 nucleotides upstream of the right terminal sequence end, about 10 to 12 nucleotides or about 11 nucleotides upstream of the right terminal sequence end. Ivana Grabundzija, Nat Commun. 2016; 7:10716, doi:10.1038/ncomms10716, incorporated herein by reference.

[0329] Exemplary helitrons can be identified using software, for example (EAHelitron) that has been used to identify helitrons in a wide range of plant genomes. See, Hu, K., Xu, K., Wen, J. *et al. Helitron* distribution in Brassicaceae and whole Genome Helitron density as a character for distinguishing plant species. *BMC Bioinformatics* 20, 354 (2019). doi: 10.1186/s12859-019-2945-8, incorporated herein by reference.

[0330] The helitron may be derived from a eukaryote. In an aspect, the helitron is derived from a mammalian genome, in an aspect, vespertilionid bats, e.g. Helibat. In embodiments, the helitron is derived from derived from a Helibat1 transposon. In embodiments, the helitron is Helraiser, the full DNA sequence of the consensus transposon, including left terminal and right terminal sequences as well as hairpin identified is provided in Grabundzija, 2016 at Supplementary Figure 1, specifically incorporated herein by reference. In an aspect, the helitron is flanked by left and right terminal sequences of the transposon. In an aspect, the left terminal sequence and right terminal sequence terminates with the conserved 5'-TC/CTAG-3' motif. In an embodiment, the helitron may comprise a palindromic sequence that is about 10 to about 35, or about 5-25 bp or about 19-bp-long palindromic sequence with the potential to form a hairpin structure.

[0331] Elements of these systems may be engineered to work within the context of the invention. For example, a helitron polypeptide may be fused to a polypeptide capable of generating an R-loop. Fusion may be by any appropriate linker, in an exemplary embodiment, XTEN16. The binding elements that allow a helitron polypeptide to bind, for example, the use of sequences complementary to the right terminal sequence and the left terminal sequence of the helitron may be engineered into a donor construct to facilitate entry of a donor polynucleotide sequence into a target polynucleotide.

[0332] In certain example embodiments, the Cas12b polypeptide, via formation of complex with a guide RNA, directs the helitron polypeptide to a target sequence in a target polynucleotide, where the helitron facilitates integration of a donor polynucleotide sequence into the target polynucleotide.

[0333] The helitron polypeptides may also comprise one or more truncations or excisions to remove domains or regions of wild-type protein to arrive at a minimal polypeptide, alter functionality according to the system in which the helitron is used, or mutated to enhance or diminish particular activities associated with the helitron, i.e., nuclease activity or helicase activity.

Multiplexing

[0334] In one embodiment, Cas polypeptide nucleases may be used in a multiplex (tandem) targeting approach. For example, Cas polypeptide nuclease herein can employ more than one RNA guide without losing activity. This may enable the use of the Cas polypeptide, systems or complexes as defined herein for targeting multiple DNA targets, genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a conserved nucleotide sequence as defined herein. The position of the different guide RNAs in the tandem does not influence the activity.

[0335] In one aspect, the Cas polypeptide nucleases may be used for tandem or multiplex targeting. It is to be understood that any of the Cas polypeptide, complexes, or compositions herein elsewhere may be used in such an approach. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the multiplex or tandem targeting approach further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0336] In one aspect, the invention provides for the use of a Cas polypeptide, complex or system as defined herein for targeting multiple gene loci. In one embodiment, this can be established by using multiple (tandem or multiplex) guide RNAs. In an embodiment, a double nickase system is provided, wherein two or more Cas nickases are provided for modifying multiple target polynucleotides. In an aspect, the guide RNA specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. In one embodiment, the guide RNAs target locations on opposite strands of the same double stranded DNA molecule. In an embodiment, the guide RNAs target locations on the same strand DNA molecule. In an embodiment, the two or more guide RNAs directs sequence-specific binding of the Cas system to sense and antisense strands of the target sequence and introduce one or more double strand break(s) to the target sequence.

[0337] In one aspect, the invention provides methods for using one or more elements of a Cas12b polypeptide, complex or system as defined herein for tandem or multiplex targeting, wherein said system herein comprises multiple guide RNA. Said guide RNA are separated by a nucleotide sequence, such as a conserved nucleotide sequence as defined herein elsewhere.

[0338] The Cas polypeptide, compositions, systems, or complexes as defined herein provide an effective means for modifying multiple target polynucleotides. The Cas polypeptide, system or complex as defined herein has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) one or more target polynucleotides in a multiplicity of cell types. As such the Cas polypeptide, system or complex as defined herein of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis, including targeting multiple gene loci within a single system.

[0339] In one aspect, the present disclosure provides a Cas polypeptide, system or complex as defined herein, having a Cas polypeptide having at least one destabilization domain associated therewith, and multiple guide RNAs that target multiple nucleic acid molecules such as DNA molecules, whereby each of said multiple guide RNAs specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. In one embodiment, the Cas polypeptide used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the Cas12b polypeptide used for multiplex targeting is a dead Cas polypeptide nuclease. The inventors have found that the Cas polypeptide as described

herein may enable improved and/or direct access to one or more nucleotides involved in the DNA:RNA duplex.

[0340] Each nucleic acid molecule target, e.g., DNA molecule can encode a gene product or encompass a gene locus. Using multiple crRNA or guide RNAs hence enables the targeting of multiple gene loci or multiple genes. In one embodiment the Cas polypeptide may cleave the DNA molecule encoding the gene product. In one embodiment expression of the gene product is altered. The Cas polypeptide and the guide RNAs do not naturally occur together. The present disclosure comprehends the guide RNAs comprising tandemly arranged guide sequences. The present disclosure further comprehends coding sequences for the Cas polypeptide being codon optimized for expression in a eukaryotic cell. In an embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. Expression of the gene product may be decreased. The Cas12b polypeptide may form part of a system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing to a target sequence in a genomic locus of interest in a cell. In one embodiment, the functional system or complex binds to the multiple target sequences. In one embodiment, the functional system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and In one embodiment, there may be an alteration of gene expression. In one embodiment, the functional system or complex may comprise further functional domains. In one embodiment, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences).

Inducible Systems

[0341] In one embodiment, a Cas polypeptide nuclease 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 Cas12b polypeptide 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 Cas12b polypeptide, 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 Provisional Application Nos. 61/736,465 and US 61/721,283, and International Patent Publication No. WO 2014/018423 A2 which is hereby incorporated by reference in its entirety.

Self-Inactivating Systems

[0342] Once all copies of a gene in the genome of a cell have been edited, continued expression of the system in that cell is no longer necessary. Indeed, sustained expression would be undesirable in case of off-target effects at unintended genomic sites, etc. Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a self-Inactivating system that relies on the use of a non-coding guide target sequence within the vector itself. Thus, after expression begins, the system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self-inactivating system includes additional RNA (e.g., guide RNA) that targets the coding sequence for the Cas polypeptide itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following: (a) within the promoter driving expression of the non-coding RNA elements, (b) within the promoter driving expression of the Cas polypeptide gene, (c) within 100bp of the ATG translational start codon in the Cas polypeptide coding sequence, (d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in the AAV genome.

[0343] In some aspects, a single guide RNA is provided that is capable of hybridization to a sequence downstream of a Cas polypeptide start codon, whereby after a period of time there is a loss of the Cas polypeptide nuclease expression. In some aspects, one or more guide RNA(s) are provided that are capable of hybridization to one or more coding or non-coding regions of the polynucleotide encoding the system, whereby after a period of time there is an inactivation of one or more, or in some cases all, of the system. In some aspects of the system, and not to be limited by theory, the cell may comprise a plurality of complexes, wherein a first subset of complexes comprise a first guide RNA capable of targeting a genomic locus or loci

to be edited, and a second subset of complexes comprise at least one second guide RNA capable of targeting the polynucleotide encoding the system, wherein the first subset of complexes mediate editing of the targeted genomic locus or loci and the second subset of complexes eventually inactivate the system, thereby inactivating further expression in the cell.

[0344] The various coding sequences (Cas polypeptide and guide RNAs) can be included on a single vector or on multiple vectors. For instance, it is possible to encode the enzyme on one vector and the various RNA sequences on another vector, or to encode the enzyme and one ω RNA or guide RNA on one vector, and the remaining guide sequence RNA on another vector, or any other permutation. In general, a system using a total of one or two different vectors is preferred.

[0345] Where multiple vectors are used, it is possible to deliver them in unequal numbers, and ideally with an excess of a vector which encodes the first guide sequence RNA relative to the second guide RNA, thereby assisting in delaying final inactivation of the system until genome editing has had a chance to occur.

[0346] The first guide RNA can target any target sequence of interest within a genome, as described elsewhere herein. The second guide sequence RNA targets a sequence within the vector which encodes the Cas polypeptide, and thereby inactivates the enzyme's expression from that vector. Thus, the target sequence in the vector must be capable of inactivating expression. Suitable target sequences can be, for instance, near to or within the translational start codon for the Cas polypeptide coding sequence, in a non-coding sequence in the promoter driving expression of the non-coding RNA elements, within the promoter driving expression of the Cas polypeptide gene, within 100bp of the ATG translational start codon in the Cas polypeptide nuclease coding sequence, and/or within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in the AAV genome. A double stranded break near this region can induce a frame shift in the Cas polypeptide nuclease coding sequence, causing a loss of protein expression. An alternative target sequence for the "self-inactivating" ω RNA or guide RNA would aim to edit/inactivate regulatory regions/sequences needed for the expression of the system or for the stability of the vector. For instance, if the promoter for the Cas12b polypeptide coding sequence is disrupted then transcription can be inhibited or prevented. Similarly, if a vector includes sequences for replication, maintenance or stability then it is possible to target these. For instance, in a AAV vector a useful target sequence is within the iTR. Other useful sequences to target can be promoter sequences, polyadenylation sites, etc.

[0347] Furthermore, if the guide RNAs are expressed in array format, the “self-inactivating” guide RNAs that target both promoters simultaneously will result in the excision of the intervening nucleotides from within the Cas12b polypeptide expression construct, effectively leading to its complete inactivation. Similarly, excision of the intervening nucleotides will result where the guide RNAs target both ITRs, or targets two or more other components simultaneously. Self-inactivation as explained herein is applicable, in general, with systems in order to provide regulation of the systems. For example, self-inactivation as explained herein may be applied to the repair of mutations, for example expansion disorders, as explained herein. As a result of this self-inactivation, repair may be only transiently active.

[0348] Addition of non-targeting nucleotides to the 5' end (e.g. 1-10 nucleotides, preferably 1-5 nucleotides) of the “self-inactivating” guide RNA can be used to delay its processing and/or modify its efficiency as a means of ensuring editing at the targeted genomic locus prior to shut down.

[0349] In one aspect of the self-inactivating AAV system, plasmids that co-express one or more guide RNA targeting genomic sequences of interest (e.g. 1-2, 1-5, 1-10, 1-15, 1-20, 1-30) may be established with “self-inactivating” guide RNAs that target an Cas polypeptide sequence at or near the engineered ATG start site (e.g. within 5 nucleotides, within 15 nucleotides, within 30 nucleotides, within 50 nucleotides, within 100 nucleotides). A regulatory sequence in the U6 promoter region can also be targeted with an guide RNA. The U6-driven guide RNAs may be designed in an array format such that multiple guide RNA sequences can be simultaneously released. When first delivered into target tissue/cells (left cell) ω RNA or guide RNAs begin to accumulate while Cas12b polypeptide levels rise in the nucleus. Cas polypeptide nuclease complexes with all of the guide RNAs to mediate genome editing and self-inactivation of the Cas12b polypeptide plasmids.

[0350] One aspect of a self-inactivating system is expression of singly or in tandem array format from 1 up to 4 or more different guide sequences, e.g., up to about 20 or about 30 guide sequences. Each individual self-inactivating guide sequence may target a different target. Such may be processed from, e.g. one chimeric pol3 transcript. Pol3 promoters such as U6 or H1 promoters may be used. Pol2 promoters such as those mentioned throughout herein. Inverted terminal repeat (iTR) sequences may flank the Pol3 promoter - ω RNA or guide RNA(s)-Pol2 promoter- Cas polypeptide.

[0351] One aspect of a tandem array transcript is that one or more guide(s) or scaffold sequences edit the one or more target(s) while one or more self-inactivating guides or scaffold sequences inactivate the system. Thus, for example, the described system for repairing expansion disorders may be directly combined with the self-inactivating system described herein. Such a system may, for example, have two ω RNA or guides directed to the target region for repair as well as at least a third ω RNA or guide directed to self-inactivation of the Cas12b polypeptide or systems.

[0352] The guide RNA may be a control guide. For example, it may be engineered to target a nucleic acid sequence encoding the Cas12b polypeptide itself, as described in U.S. Patent Publication No. US2015232881A1, the disclosure of which is hereby incorporated by reference. In one embodiment, a system or composition may be provided with just the guide RNA engineered to target the nucleic acid sequence encoding the Cas polypeptide. In addition, the system or composition may be provided with the guide RNA engineered to target the nucleic acid sequence encoding the Cas12b polypeptide, as well as nucleic acid sequence encoding the Cas12b polypeptide and, optionally a second ω RNA or guide RNA and, further optionally, a repair template. The second guide RNA may be the primary target of the system or composition (such a therapeutic, diagnostic, knock out etc. as defined herein). In this way, the system or composition is self-inactivating. This is exemplified in relation to Cas in US2015232881A1 (also published as WO2015070083 (A1) referenced elsewhere herein, and may be extrapolated to other Cas12b polypeptides, e.g. orthologous Cas12b polypeptides.

DETECTION COMPOSITIONS AND METHODS OF DETECTION

[0353] In another aspect, embodiments disclosed herein are directed to polynucleotide detection compositions, systems and methods. The detection composition may comprise any of the Cas polypeptides and any one or more nucleic acid components discussed above. In the addition, the compositions and system may comprise a detection construct. In one example embodiment, the detection construct comprises at least a portion of single-stranded polynucleotide. The one or more nucleic acid components are configured to bind a target sequence on a target polypeptide. Binding of the CRISPR-Cas complex to the target sequence activates Cas cleavage activity and may further activate Cas collateral activity whereby Cas subsequently cleaves non-target single-stranded polynucleotides in a nucleic acid components-independent fashion. Accordingly, the detection construct can be configured so that a

detectable signal is generated upon cleavage of the single-stranded portion of the detection constructs thereby indicating the present of the target sequence in a sample. Example detection constructs are discussed in further detail below. In further example embodiments, the compositions may further comprise amplification reagents. Amplification reagents may comprise primers and polymerase and/or reverse transcriptases needed to amplify the target sequence. In one example embodiment, the amplification reagents are isothermal amplification reagents. In other example embodiments, the compositions and systems may further comprise quick extraction solutions that allow for detection of target sequences in crude samples or with minimal purification prior to amplification and/or detection.

Detection Construct

[0354] The systems and methods described herein comprise a detection construct. As used herein, a “detection construct” refers to a molecule that can be cleaved or otherwise deactivated by an activated CRISPR-Cas system protein described herein. The term “detection construct” may also be referred to in the alternative as a “masking construct.” Depending on the nuclease activity of the Cas polypeptide and the methods utilized, the masking construct may be an RNA-based masking construct or a DNA-based masking construct. The Nucleic Acid-based masking constructs comprises a nucleic acid element that is cleavable by a Cas polypeptide. Cleavage of the nucleic acid element releases agents or produces conformational changes that allow a detectable signal to be produced. Example constructs demonstrating how the nucleic acid element may be used to prevent or mask generation of detectable signal are described below and embodiments of the invention comprise variants of the same. Prior to cleavage, or when the masking construct is in an ‘active’ state, the masking construct blocks the generation or detection of a positive detectable signal. In one embodiment, detection constructs are designed for cutting motifs of particular Cas polypeptide.

[0355] It will be understood that in certain example embodiments a minimal background signal may be produced in the presence of an active masking construct. 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 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 one embodiment a first signal may be detected when the masking agent is present or when a CRISPR-Cas system has not been activated (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, or upon activation of the Cas polypeptide. The positive detectable signal, then, is a signal detected upon activation of the Cas polypeptide, and may be, in a colorimetric or fluorescent assay, a decrease in fluorescence or color relative to a control or an increase in fluorescence or color relative to a control, depending on the configuration of the lateral flow substrate, and as described further herein.

[0356] In certain example embodiments, the masking construct may comprise an HCR initiator sequence and a cutting motif, or a cleavable structural element, such as a loop or hairpin, that prevents the initiator from initiating the HCR reaction. The cutting motif may be preferentially cut by one of the activated Cas polypeptides. Upon cleavage of the cutting motif or structure element by an activated Cas polypeptide, the initiator is then released to trigger the HCR reaction, detection thereof indicating the presence of one or more targets in the sample. In certain example embodiments, the masking construct comprises a hairpin with an RNA loop. When an activated Cas polypeptide cuts the RNA loop, the initiator can be released to trigger the HCR reaction.

[0357] 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 an RNA interference pathway, such as a short hairpin RNA (shRNA) or small interfering RNA (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, aptamer, 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. In preferred embodiments, the masking constructs comprise two or more detectable signals, for example, fluorescent signals, that can be read on different channels of a fluorimeter.

[0358] In specific embodiments, the masking construct comprises a silencing RNA that suppresses generation of a gene product encoded by a reporting construct, wherein the gene product generates the detectable positive signal when expressed.

[0359] 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 purposes. 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 and the RNA or DNA aptamers are degraded.

[0360] 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 an RNA- or DNA-based aptamer that can be cleaved by the activated effector protein upon detection of a target molecule.

[0361] 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 DNA or 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 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 herein.

[0362] In certain example embodiments, the masking construct may comprise a ribozyme. Ribozymes are RNA molecules having catalytic properties. 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 the reaction generating a negative control signal, or preventing generation of a positive detectable signal, is removed thereby allowing a positive detectable signal to be generated. 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.

[0363] In one embodiment, the masking construct may be a ribozyme that generates a negative detectable signal, and wherein a positive detectable signal is generated when the ribozyme is deactivated.

[0364] 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 DNA or RNA aptamers to the protein. Upon activation of the effector proteins disclosed herein, the DNA or RNA aptamers are cleaved or degraded to an extent that they no longer inhibit the protein's ability to generate the detectable signal. In certain example embodiments, the aptamer is a thrombin inhibitor aptamer. In certain example embodiments the thrombin inhibitor aptamer has a sequence of GGGAACAAAGCUGAAGUACUUACCC (SEQ ID NO: 41). When this aptamer is cleaved, thrombin will become active and will cleave a peptide colorimetric or fluorescent substrate. In certain example embodiments, the colorimetric substrate is para-nitroanilide (pNA) covalently linked to the peptide substrate for thrombin. Upon cleavage by thrombin, pNA is released and becomes yellow in color and easily visible to the eye. In certain example

embodiments, the fluorescent substrate is 7-amino-4-methylcoumarin a blue fluorophore that can be detected using a fluorescence detector. Inhibitory aptamers may also be used for horseradish peroxidase (HRP), beta-galactosidase, or calf alkaline phosphatase (CAP) and within the general principals laid out above.

[0365] In one embodiment, RNase or DNase activity is detected colorimetrically via cleavage of enzyme-inhibiting aptamers. One potential mode of converting DNase or RNase activity into a colorimetric signal is to couple the cleavage of a DNA or RNA aptamer with the re-activation of an enzyme that is capable of producing a colorimetric output. In the absence of RNA or DNA cleavage, the intact aptamer will bind to the enzyme target and inhibit its activity. The advantage of this readout system is that the enzyme provides an additional amplification step: once liberated from an aptamer via collateral activity (e.g. Cas collateral activity), the colorimetric enzyme will continue to produce colorimetric product, leading to a multiplication of signal.

[0366] In one embodiment, an existing aptamer that inhibits an enzyme with a colorimetric readout is used. Several aptamer/enzyme pairs with colorimetric readouts exist, such as thrombin, protein C, neutrophil elastase, and subtilisin. These proteases have colorimetric substrates based upon pNA and are commercially available. In one embodiment, a novel aptamer targeting a common colorimetric enzyme is used. Common and robust enzymes, such as beta-galactosidase, horseradish peroxidase, or calf intestinal alkaline phosphatase, could be targeted by engineered aptamers designed by selection strategies such as SELEX. Such strategies allow for quick selection of aptamers with nanomolar binding efficiencies and could be used for the development of additional enzyme/aptamer pairs for colorimetric readout.

[0367] In one embodiment, the masking construct may be a DNA or RNA aptamer and/or may comprise a DNA or RNA-tethered inhibitor.

[0368] In one embodiment, the masking construct may comprise a DNA or RNA oligonucleotide to which a detectable ligand and a masking component are attached.

[0369] In one embodiment, RNase or DNase activity is detected colorimetrically via cleavage of RNA-tethered inhibitors. Many common colorimetric enzymes have competitive, reversible inhibitors: for example, beta-galactosidase can be inhibited by galactose. Many of these inhibitors are weak, but their effect can be increased by increases in local concentration. By linking local concentration of inhibitors to DNase RNase activity, colorimetric enzyme and inhibitor pairs can be engineered into DNase and RNase sensors. The colorimetric DNase or

RNAse sensor based upon small-molecule inhibitors involves three components: the colorimetric enzyme, the inhibitor, and a bridging RNA or DNA that is covalently linked to both the inhibitor and enzyme, tethering the inhibitor to the enzyme. In the uncleaved configuration, the enzyme is inhibited by the increased local concentration of the small molecule; when the DNA or RNA is cleaved (e.g. by Cas collateral cleavage), the inhibitor will be released, and the colorimetric enzyme will be activated.

[0370] In one embodiment, the aptamer or DNA- or RNA-tethered inhibitor may sequester an enzyme, wherein the enzyme generates a detectable signal upon release from the aptamer or DNA or RNA tethered inhibitor by acting upon a substrate. In one embodiment, the aptamer may be an inhibitor aptamer that inhibits an enzyme and prevents the enzyme from catalyzing generation of a detectable signal from a substance. In one embodiment, the DNA- or RNA-tethered inhibitor may inhibit an enzyme and may prevent the enzyme from catalyzing generation of a detectable signal from a substrate.

[0371] In one embodiment, RNAse activity is detected colorimetrically via formation and/or activation of G-quadruplexes. G quadruplexes in DNA can complex with heme (iron (III)-protoporphyrin IX) to form a DNzyme with peroxidase activity. When supplied with a peroxidase substrate (e.g. ABTS: (2,2'-Azinobis [3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt)), the G-quadruplex-heme complex in the presence of hydrogen peroxide causes oxidation of the substrate, which then forms a green color in solution. An example G-quadruplex forming DNA sequence is: GGGTAGGGCGGGTTGGGA (SEQ ID NO: 42). By hybridizing an additional DNA or RNA sequence, referred to herein as a “staple,” to this DNA aptamer, formation of the G-quadruplex structure will be limited. Upon collateral activation, the staple will be cleaved allowing the G quadruplex to form and heme to bind. This strategy is particularly appealing because color formation is enzymatic, meaning there is additional amplification beyond collateral activation.

[0372] In one embodiment, the masking construct may comprise an RNA oligonucleotide designed to bind a G-quadruplex forming sequence, wherein a G-quadruplex structure is formed by the G-quadruplex forming sequence upon cleavage of the masking construct, and wherein the G-quadruplex structure generates a detectable positive signal.

[0373] 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 DNA- or RNA-based aptamer that can be cleaved by the activated effector protein upon detection of a target molecule.

[0374] 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 or DNA. Upon activation of the effector proteins disclosed herein, the RNA or DNA 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 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.

[0375] When the RNA or DNA bridge is cut by the activated Cas polypeptide, the aforementioned color shift is observed. In certain example embodiments the particles are colloidal metals. In certain other example embodiments, the colloidal metal is a colloidal gold. In certain example embodiments, the colloidal nanoparticles are 15 nm gold nanoparticles (AuNPs). Due to the unique surface properties of colloidal gold nanoparticles, maximal absorbance is observed at 520 nm when fully dispersed in solution and appear red in color to the naked eye. Upon aggregation of AuNPs, they exhibit a red-shift in maximal absorbance and appear darker in color, eventually precipitating from solution as a dark purple aggregate.

In certain example embodiments the nanoparticles are modified to include DNA linkers extending from the surface of the nanoparticle. Individual particles are linked together by single-stranded RNA (ssRNA) or single-stranded DNA (ssDNA) bridges that hybridize on each end to at least a portion of the DNA linkers. Thus, the nanoparticles will form a web of linked particles and aggregate, appearing as a dark precipitate. Upon activation of the Cas polypeptides disclosed herein, the ssRNA or ssDNA bridge will be cleaved, releasing the Au NPS from the linked mesh and produce a visible red color. Example DNA linkers and bridge sequences are listed below. Thiol linkers on the end of the DNA linkers may be used for surface conjugation to the AuNPS. Other forms of conjugation may be used. In certain example embodiments, two populations of AuNPs may be generated, one for each DNA linker. This will help facilitate proper binding of the ssRNA bridge with proper orientation. In certain example embodiments, a first DNA linker is conjugated by the 3' end while a second DNA linker is conjugated by the 5' end.

[0376] In certain other example embodiments, the masking construct may comprise an RNA or DNA 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 or DNA 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 or DNA 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.

[0377] In certain other example embodiments, the masking construct may comprise one or more RNA oligonucleotides to which are attached one or more metal nanoparticles, such as gold nanoparticles. In one embodiment, the masking construct comprises a plurality of metal

nanoparticles crosslinked by a plurality of RNA or DNA oligonucleotides forming a closed loop. In one embodiment, the masking construct comprises three gold nanoparticles crosslinked by three RNA or DNA oligonucleotides forming a closed loop. In one embodiment, the cleavage of the RNA or DNA oligonucleotides by the Cas polypeptide leads to a detectable signal produced by the metal nanoparticles.

[0378] In certain other example embodiments, the masking construct may comprise one or more RNA or DNA oligonucleotides to which are attached one or more quantum dots. In one embodiment, the cleavage of the RNA or DNA oligonucleotides by the Cas polypeptide leads to a detectable signal produced by the quantum dots.

[0379] 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 or DNA. 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 or DNA 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 certain example embodiments the quantum dot is streptavidin conjugated. RNA or DNA are attached via biotin linkers and recruit quenching molecules with the sequences /5Biosg/UCUCGUACGUUC/3IAbRQSp/ (SEQ ID NO: 43) or /5Biosg/UCUCGUACGUUCUCUCGUACGUUC/3IAbRQSp/ (SEQ ID NO: 44) where /5Biosg/ is a biotin tag and /3IAbRQSp/ is an Iowa black quencher (Iowa Black FQ). Upon cleavage, by the activated effectors disclosed herein the quantum dot will fluoresce visibly.

[0380] In specific embodiments, the detectable ligand may be a fluorophore and the masking component may be a quencher molecule.

[0381] 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).

[0382] In certain example embodiments, the masking construct comprises the use of intercalating dyes which change their absorbance in response to cleavage of long RNAs or DNAs to short nucleotides. Several such dyes exist. For example, pyronine-Y will complex with RNA and form a complex that has an absorbance at 572 nm. Cleavage of the RNA results in loss of absorbance and a color change. Methylene blue may be used in a similar fashion, with changes in absorbance at 688 nm upon RNA cleavage. Accordingly, in certain example embodiments the masking construct comprises an RNA and intercalating dye complex that changes absorbance upon the cleavage of RNA by the effector proteins disclosed herein.

[0383] In certain example embodiments, the masking construct may comprise an initiator for an HCR reaction. See e.g. Dirks and Pierce. PNAS 101, 15275-15728 (2004). HCR reactions utilize the potential energy in two hairpin species. When a single-stranded initiator having a portion of complementary to a corresponding region on one of the hairpins is released into the previously stable mixture, it opens a hairpin of one species. This process, in turn, exposes a single-stranded region that opens a hairpin of the other species. This process, in turn, exposes a single stranded region identical to the original initiator. The resulting chain reaction may lead to the formation of a nicked double helix that grows until the hairpin supply is exhausted. Detection of the resulting products may be done on a gel or colorimetrically. Example colorimetric detection methods include, for example, those disclosed in Lu et al. "Ultra-sensitive colorimetric assay system based on the hybridization chain reaction-triggered enzyme cascade amplification ACS Appl Mater Interfaces, 2017, 9(1):167-175, Wang et al. "An enzyme-free colorimetric assay using hybridization chain reaction amplification and split

aptamers” Analyst 2015, 150, 7657-7662, and Song et al. “Non-covalent fluorescent labeling of hairpin DNA probe coupled with hybridization chain reaction for sensitive DNA detection.” Applied Spectroscopy, 70(4): 686-694 (2016).

[0384] In certain example embodiments, the masking construct suppresses generation of a detectable positive signal until cleaved, or modified by an activated Cas polypeptide. In one embodiment, the masking construct may suppress generation of a detectable positive signal by masking the detectable positive signal, or generating a detectable negative signal instead.

Amplification Reagents

[0385] In certain example embodiments, target RNAs and/or DNAs may be amplified prior to activating the CRISPR effector protein. Any suitable RNA or DNA amplification technique may be used. In certain example embodiments, the RNA or DNA amplification is an isothermal amplification. In certain example embodiments, the isothermal amplification may be nucleic-acid sequenced-based amplification (NASBA), recombinase polymerase amplification (RPA), loop-mediated isothermal amplification (LAMP), strand displacement amplification (SDA), helicase-dependent amplification (HDA), or nicking enzyme amplification reaction (NEAR). In certain example embodiments, non-isothermal amplification methods may be used which include, but are not limited to, PCR, multiple displacement amplification (MDA), rolling circle amplification (RCA), ligase chain reaction (LCR), or ramification amplification method (RAM).

[0386] In certain example embodiments, the RNA or DNA amplification is NASBA, which is initiated with reverse transcription of target RNA by a sequence-specific reverse primer to create an RNA/DNA duplex. RNase H is then used to degrade the RNA template, allowing a forward primer containing a promoter, such as the T7 promoter, to bind and initiate elongation of the complementary strand, generating a double-stranded DNA product. The RNA polymerase promoter-mediated transcription of the DNA template then creates copies of the target RNA sequence. Importantly, each of the new target RNAs can be detected by the guide RNAs thus further enhancing the sensitivity of the assay. Binding of the target RNAs by the guide RNAs then leads to activation of the CRISPR effector protein and the methods proceed as outlined above. The NASBA reaction has the additional advantage of being able to proceed under moderate isothermal conditions, for example at approximately 41°C, making it suitable for systems and devices deployed for early and direct detection in the field and far from clinical laboratories.

[0387] In certain other example embodiments, a recombinase polymerase amplification (RPA) reaction may be used to amplify the target nucleic acids. RPA reactions employ recombinases which are capable of pairing sequence-specific primers with homologous sequence in duplex DNA. If target DNA is present, DNA amplification is initiated and no other sample manipulation such as thermal cycling or chemical melting is required. The entire RPA amplification system is stable as a dried formulation and can be transported safely without refrigeration. RPA reactions may also be carried out at isothermal temperatures with an optimum reaction temperature of 37-42o C. The sequence specific primers are designed to amplify a sequence comprising the target nucleic acid sequence to be detected. In certain example embodiments, an RNA polymerase promoter, such as a T7 promoter, is added to one of the primers. This results in an amplified double-stranded DNA product comprising the target sequence and an RNA polymerase promoter. After, or during, the RPA reaction, an RNA polymerase is added that will produce RNA from the double-stranded DNA templates. The amplified target RNA can then in turn be detected by the CRISPR effector system. In this way target DNA can be detected using the embodiments disclosed herein. RPA reactions can also be used to amplify target RNA. The target RNA is first converted to cDNA using a reverse transcriptase, followed by second strand DNA synthesis, at which point the RPA reaction proceeds as outlined above.

[0388] An embodiment of the invention may comprise nickase-based amplification. The nicking enzyme may be a CRISPR protein. Accordingly, the introduction of nicks into dsDNA can be programmable and sequence-specific. An embodiment of the invention, may comprise two guides designed to target opposite strands of a dsDNA target. The nicked strands may then be extended by a polymerase. In an embodiment, the locations of the nicks are selected such that extension of the strands by a polymerase is towards the central portion of the target duplex DNA between the nick sites. In certain embodiments, primers are included in the reaction capable of hybridizing to the extended strands followed by further polymerase extension of the primers to regenerate two dsDNA pieces: a first dsDNA that includes the first strand guide site or both the first and second strand guide sites, and a second dsDNA that includes the second strand guide site or both the first and second strand guide sites. These pieces continue to be nicked and extended in a cyclic reaction that exponentially amplifies the region of the target between nicking sites.

[0389] The amplification can be isothermal and selected for temperature. In one embodiment, the amplification proceeds rapidly at 37 degrees. In other embodiments, the temperature of the isothermal amplification may be chosen by selecting a polymerase (e.g. Bsu, Bst, Phi29, klenow fragment etc.) operable at a different temperature.

[0390] Thus, whereas nicking isothermal amplification techniques use nicking enzymes with fixed sequence preference (e.g. in nicking enzyme amplification reaction or NEAR), which requires denaturing of the original dsDNA target to allow annealing and extension of primers that add the nicking substrate to the ends of the target, use of a CRISPR nickase wherein the nicking sites can be programmed via guide RNAs means that no denaturing step is necessary, enabling the entire reaction to be truly isothermal. This also simplifies the reaction because these primers that add the nicking substrate are different than the primers that are used later in the reaction, meaning that NEAR requires two primer sets (i.e. 4 primers) while Cpf1 nicking amplification only requires one primer set (i.e. two primers). This makes nicking Cpf1 amplification much simpler and easier to operate without complicated instrumentation to perform the denaturation and then cooling to the isothermal temperature.

[0391] Accordingly, in certain example embodiments the systems disclosed herein may include amplification reagents. Different components or reagents useful for amplification of nucleic acids are described herein. For example, an amplification reagent as described herein may include a buffer, such as a Tris buffer. A Tris buffer may be used at any concentration appropriate for the desired application or use, for example including, but not limited to, a concentration of 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM, 12 mM, 13 mM, 14 mM, 15 mM, 25 mM, 50 mM, 75 mM, 1 M, or the like. One of skill in the art will be able to determine an appropriate concentration of a buffer such as Tris for use with the present invention.

[0392] Other components of a biological or chemical reaction may include a cell lysis component in order to break open or lyse a cell for analysis of the materials therein. A cell lysis component may include, but is not limited to, a detergent, a salt as described above, such as NaCl, KCl, ammonium sulfate $[(\text{NH}_4)_2\text{SO}_4]$, or others. Detergents that may be appropriate for the invention may include Triton X-100, sodium dodecyl sulfate (SDS), CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), ethyl trimethyl ammonium bromide, nonyl phenoxypolyethoxyethanol (NP-40). Concentrations of detergents may depend on the particular application and may be specific to the reaction in some

cases. Amplification reactions may include dNTPs and nucleic acid primers used at any concentration appropriate for the invention, such as including, but not limited to, a concentration of 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM, 500 nM, 550 nM, 600 nM, 650 nM, 700 nM, 750 nM, 800 nM, 850 nM, 900 nM, 950 nM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 150 mM, 200 mM, 250 mM, 300 mM, 350 mM, 400 mM, 450 mM, 500 mM, or the like. Likewise, a polymerase useful in accordance with the invention may be any specific or general polymerase known in the art and useful or the invention, including Taq polymerase, Q5 polymerase, or the like.

[0393] In some embodiments, amplification reagents as described herein may be appropriate for use in hot-start amplification. Hot start amplification may be beneficial in some embodiments to reduce or eliminate dimerization of adaptor molecules or oligos, or to otherwise prevent unwanted amplification products or artifacts and obtain optimum amplification of the desired product. Many components described herein for use in amplification may also be used in hot-start amplification. In some embodiments, reagents or components appropriate for use with hot-start amplification may be used in place of one or more of the composition components as appropriate. For example, a polymerase or other reagent may be used that exhibits a desired activity at a particular temperature or other reaction condition. In some embodiments, reagents may be used that are designed or optimized for use in hot-start amplification, for example, a polymerase may be activated after transposition or after reaching a particular temperature. Such polymerases may be antibody-based or aptamer-based. Polymerases as described herein are known in the art. Examples of such reagents may include, but are not limited to, hot-start polymerases, hot-start dNTPs, and photo-caged dNTPs. Such reagents are known and available in the art. One of skill in the art will be able to determine the optimum temperatures as appropriate for individual reagents.

[0394] Amplification of nucleic acids may be performed using specific thermal cycle machinery or equipment and may be performed in single reactions or in bulk, such that any desired number of reactions may be performed simultaneously. In some embodiments, amplification may be performed using microfluidic or robotic devices, or may be performed using manual alteration in temperatures to achieve the desired amplification. In some embodiments, optimization may be performed to obtain the optimum reactions conditions for

the particular application or materials. One of skill in the art will understand and be able to optimize reaction conditions to obtain sufficient amplification.

[0395] In certain embodiments, detection of DNA with the methods or systems of the invention requires transcription of the (amplified) DNA into RNA prior to detection.

[0396] It will be evident that detection methods of the invention can involve nucleic acid amplification and detection procedures in various combinations. The nucleic acid to be detected can be any naturally occurring or synthetic nucleic acid, including but not limited to DNA and RNA, which may be amplified by any suitable method to provide an intermediate product that can be detected. Detection of the intermediate product can be by any suitable method including but not limited to binding and activation of a CRISPR protein which produces a detectable signal moiety by direct or collateral activity.

LAMP-based Isothermal Amplification

[0397] In certain example embodiments, the LAMP amplification reagents may include primers to SARS-COV2. LAMP reagents may further comprise colorimetric and/or fluorescent detection reagents, such as hydroxy naphthol blue (see, e.g. Goto, M., et al., Colorimetric detection of loop-mediated isothermal amplification reaction by using hydroxy naphthol blue. *Biotechniques*, 2009. 46(3): p. 167-72.), leuco triphenylmethane dyes (see, e.g. Miyamoto, S., et al., Method for colorimetric detection of double-stranded nucleic acid using leuco triphenylmethane dyes. *Anal Biochem*, 2015. 473: p. 28-33) and pH-sensitive dyes (see, e.g. Tanner, N.A., Y. Zhang, and T.C. Evans, Jr., Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. *Biotechniques*, 2015. 58(2): p. 59-68); as well as fluorescent detection (see, e.g. Yu et al., *Clinical Chemistry*, hvaa102, doi:10.1093/clinchem/hvaa102 12 May 2020), including use of quenching probes (see, e.g. Shirato *et al.*, *J Virol Methods*. 2018 Aug;258:41-48. doi: 10.1016/j.jviromet.2018.05.006). An overview of LAMP methods, including OSD-LAMP, for sequence-specific detection is described in Becherer et al., *Anal. Methods*, 2020,12, 717-746, doi: 10.1039/C9AY02246E, incorporated herein by reference.

[0398] In embodiments, the LAMP amplification reagents can comprise oligonucleotide strand displacement (OSD) probes. As used herein, oligonucleotide strand displacement probes are also referred to herein as oligonucleotide strand exchange probes or one-step strand displacement probes. The general concept of the use of OSD exchange is depicted in Figure 1 of Bhadra *et al.*, High-surety isothermal amplification and detection of SARS-CoV-2,

including with crude enzymes, doi:10.1101/2020.04.13.039941. OSD probes rely on the binding enthalpy between the target-binding probe and amplicon of the LAMP reaction yielding a strand exchange reaction, leading to an easily read change in fluorescent signal. As a result, the results of a LAMP reaction can be visually or optically read fluorogenic OSD probes.

[0399] In an aspect, the OSD probes comprise a sequence specific for a target molecule. The OSD probes may comprise a pre-hybridized nucleic acid sequence, strand wherein the target sequence is 5, 6, 7, 8, 9, 10, 11, 12 or 13 nucleotides longer than the strand to which it is hybridized, allowing for sequence-specific interaction with a complementary target, with the OSD undergoing strand exchange and yielding a change in fluorescent signal.

[0400] In an aspect, the OSD probes are provided at a concentration of about 50nM to 200nM, about 75nM to 150nM, less than or equal to 200nM, 190nM, 180nM, 170nM, 160nM, 150nM, 140nM, 130nM, 120nM, 110nM, 100nM, 90nM, 80nM, 75nM, 65nM, or 50nM. Probes can be designed to be complementary to the loop region between the F1c and F2 primer binding sites for the LAMP primers, this can be referred to as the long toehold region. The complementary portion can be between about 9 and 14 nucleotides long, more preferably 11-12 nucleotides long. In an aspect, the longer strand of the OSD is labeled with a fluorescent molecule at the 5' or 3' end of the strand. In an aspect, the label is provided on the end opposite the designed complementary target region (long toehold region). The short strand is prepared with a quencher on one end of the probe, and can be designed to comprise a region complementary to a portion of the long strand. The OSD probes can be provided as part of LAMP reagents as described herein, which may comprise their use on any of the devices, cartridges or in any of the compositions as provided herein, including being provided as a lyophilized reagent in some instances.

Extraction Solutions

[0401] In certain aspects, embodiments disclosed herein are directed to compositions and kits that consolidate extraction-free lysis and amplification of target nucleic acids into a single reaction volume. In certain example embodiments, the extraction-free lysis reagents can be used to extract nucleic acids from cells and/or viral particles. In contrast to existing protocols, the extraction-free lysis solution does not require isolation of the nucleic acid prior to further amplification. The extraction-free lysis reagents may be mixed with amplification reagents such as standard RT-PCR amplification reactions.

[0402] In one embodiment, extraction-free lysis solution and isothermal amplification reagents may be lyophilized in a single reaction volume, to be reconstituted by addition of a sample to be assayed. In certain other embodiments, the extraction-free lysis solution and isothermal amplification reagents may be lyophilized and stored on a cartridge or lateral flow strip, as discussed in further detail below.

[0403] In certain example embodiments, the single lysis reaction compositions and kits may further comprise one or more Cas polypeptides possessing collateral activity and a detection construct. Pairing with one or more Cas polypeptides may increase sensitivity or specificity of the assay. In certain example embodiments, the one or more Cas polypeptides may be thermostable Cas polypeptides. Example Cas polypeptides are disclosed in further detail below.

[0404] In certain example embodiments, the single lysis amplification reaction compositions and kits may comprise optimized primers and/or one or more additives. In an aspect, the design optimizes the primers used in the amplification. In particular aspects, the isothermal amplification is used alone. In another aspect, the isothermal amplification is used with CRISPR-Cas systems. In either approach, design considerations can follow a rational design for optimization of the reactions. In an example, varying additives with specific primers, target, Cas polypeptide, temperature, and other additive concentrations within the reaction can be identified. Optimization can be made with the goal of reducing the number of steps and buffer exchanges that have to occur in the reaction, simplifying the reaction and reducing the risks of contamination at transfer steps. In an aspect, addition of inhibitors, such as proteinase K can be considered so that buffer exchanges can be reduced. Similarly, optimizing the salt levels as well as the type of salt utilized can further facilitate and optimize the one-pot detections disclosed herein. In an aspect, potassium chloride can be utilized rather than sodium chloride when such amplification approaches are used with bead concentration in a lysis step.

[0405] In one embodiment, the compositions and kits may further comprise nucleic acid binding bead. The bead may be used to capture, concentrate or otherwise enrich for particular material. The bead may be magnetic and may be provided to capture nucleic acid material. In another aspect, the bead is a silica bead. Beads may be utilized in an extraction step of the methods disclosed herein. Beads can be optionally used with the methods described herein, including with the one-pot methods that allow for concentration of viral nucleic acids from large volume samples, such as saliva or swab samples to allow for a single one-pot reaction

method. Concentration of desired target molecules can be increased by about 10-fold, 50-fold, 100-fold, 200-fold, 500-fold, 800-fold, 1000-fold, 1500-fold, 2000-fold, 2500-fold, 3000-fold, or more.

[0406] Magnetic beads in a PEG and salt solution are preferred in an aspect, and in embodiments bind to viral RNA and/or DNA which allows for concentration and lysis concurrently. Silica beads can be used in another aspect. Capture moieties such as oligonucleotide functionalized beads are envisioned for use. The beads may be used with the extraction reagents, allowed to incubate with a sample and the lysis/extraction buffer, thereby concentrating target molecules on the beads. When used with a cartridge device detailed elsewhere herein, a magnet can be activated and the beads collected, with optional flushing of the extraction buffer and one or more washes performed. Advantageously, the beads can be used in the one-pot methods and systems without additional washings of the beads, allowing for a more efficient process without increased risks of contamination in multi-step processes. Beads can be utilized with the isothermal amplifications detailed herein, and the beads can flow into an amplification chamber of the cartridge or be maintained in the pot for the amplification step. Upon heating, nucleic acid can be released off the beads.

Diagnostic Devices

[0407] The systems described herein can be embodied on diagnostic devices. A number of substrates and configurations may be used. The devices may be capable of defining multiple individual discrete volumes within the device. As used herein an “individual discrete volume” refers to a discrete space, such as a container, receptacle, or other defined volume or space that can be defined by properties that prevent and/or inhibit migration of target molecules, for example a volume or space defined by physical properties such as walls, for example the walls of a well, tube, or a surface of a droplet, which may be impermeable or semipermeable, or as defined by other means such as chemical, diffusion rate limited, electro-magnetic, or light illumination, or any combination thereof that can contain a sample within a defined space. Individual discrete volumes may be identified by molecular tags, such as nucleic acid barcodes. By “diffusion rate limited” (for example diffusion defined volumes) is meant spaces that are only accessible to certain molecules or reactions because diffusion constraints effectively defining a space or volume as would be the case for two parallel laminar streams where diffusion will limit the migration of a target molecule from one stream to the other. By “chemical” defined volume or space is meant spaces where only certain target molecules can

exist because of their chemical or molecular properties, such as size, where for example gel beads may exclude certain species from entering the beads but not others, such as by surface charge, matrix size or other physical property of the bead that can allow selection of species that may enter the interior of the bead. By “electro-magnetically” defined volume or space is meant spaces where the electro-magnetic properties of the target molecules or their supports such as charge or magnetic properties can be used to define certain regions in a space such as capturing magnetic particles within a magnetic field or directly on magnets. By “optically” defined volume is meant any region of space that may be defined by illuminating it with visible, ultraviolet, infrared, or other wavelengths of light such that only target molecules within the defined space or volume may be labeled. One advantage to the use of non-walled, or semipermeable discrete volumes is that some reagents, such as buffers, chemical activators, or other agents may be passed through the discrete volume, while other materials, such as target molecules, may be maintained in the discrete volume or space. Typically, a discrete volume will include a fluid medium, (for example, an aqueous solution, an oil, a buffer, and/or a media capable of supporting cell growth) suitable for labeling of the target molecule with the indexable nucleic acid identifier under conditions that permit labeling. Exemplary discrete volumes or spaces useful in the disclosed methods include droplets (for example, microfluidic droplets and/or emulsion droplets), hydrogel beads or other polymer structures (for example poly-ethylene glycol di-acrylate beads or agarose beads), tissue slides (for example, fixed formalin paraffin embedded tissue slides with particular regions, volumes, or spaces defined by chemical, optical, or physical means), microscope slides with regions defined by depositing reagents in ordered arrays or random patterns, tubes (such as, centrifuge tubes, microcentrifuge tubes, test tubes, cuvettes, conical tubes, and the like), bottles (such as glass bottles, plastic bottles, ceramic bottles, Erlenmeyer flasks, scintillation vials and the like), wells (such as wells in a plate), plates, pipettes, or pipette tips among others. In certain embodiments, the compartment is an aqueous droplet in a water-in-oil emulsion. In specific embodiments, any of the applications, methods, or systems described herein requiring exact or uniform volumes may employ the use of an acoustic liquid dispenser.

[0408] In some embodiments, the individual discrete volumes may be droplets.

[0409] In certain example embodiments, the device comprises a flexible material substrate on which a number of spots may be defined. Flexible substrate materials suitable for use in diagnostics and biosensing are known within the art. The flexible substrate materials may be

made of plant derived fibers, such as cellulosic fibers, or may be made from flexible polymers such as flexible polyester films and other polymer types. Within each defined spot, reagents of the system described herein are applied to the individual spots. Each spot may contain the same reagents except for a different guide RNA or set of guide RNAs, or where applicable, a different detection aptamer to screen for multiple targets at once. Thus, the systems and devices herein may be able to screen samples from multiple sources (e.g. multiple clinical samples from different individuals) for the presence of the same target, or a limited number of targets, or aliquots of a single sample (or multiple samples from the same source) for the presence of multiple different targets in the sample. In certain example embodiments, the elements of the systems described herein are freeze dried onto the paper or cloth substrate. Example flexible material-based substrates that may be used in certain example devices are disclosed in Pardee et al. Cell. 2016, 165(5):1255-66 and Pardee et al. Cell. 2014, 159(4):950-54. Suitable flexible material-based substrates for use with biological fluids, including blood are disclosed in International Patent Application Publication No. WO/2013/071301 entitled "Paper based diagnostic test" to Shevkoplyas et al. U.S. Patent Application Publication No. 2011/0111517 entitled "Paper-based microfluidic systems" to Siegel et al. and Shafiee et al. "Paper and Flexible Substrates as Materials for Biosensing Platforms to Detect Multiple Biotargets" Scientific Reports 5:8719 (2015). Further flexible based materials, including those suitable for use in wearable diagnostic devices are disclosed in Wang et al. "Flexible Substrate-Based Devices for Point-of-Care Diagnostics" Cell 34(11):909-21 (2016). Further flexible based materials may include nitrocellulose, polycarbonate, methylethyl cellulose, polyvinylidene fluoride (PVDF), polystyrene, or glass (see e.g., US20120238008). In certain embodiments, discrete volumes are separated by a hydrophobic surface, such as but not limited to wax, photoresist, or solid ink.

[0410] In some embodiments, a dosimeter or badge may be provided that serves as a sensor or indicator such that the wearer is notified of exposure to certain microbes or other agents. For example, the systems described herein may be used to detect a particular pathogen. Likewise, aptamer-based embodiments disclosed above may be used to detect both polypeptide as well as other agents, such as chemical agents, to which a specific aptamer may bind. Such a device may be useful for surveillance of soldiers or other military personnel, as well as clinicians, researchers, hospital staff, and the like, in order to provide information relating to exposure to potentially dangerous agents as quickly as possible, for example for biological or chemical

warfare agent detection. In other embodiments, such a surveillance badge may be used for preventing exposure to dangerous microbes or pathogens in immunocompromised patients, burn patients, patients undergoing chemotherapy, children, or elderly individuals.

[0411] In specific embodiments, each individual discrete volume further comprises one or more detection aptamers comprising a masked RNA polymerase promoter binding site or a masked primer binding site. As such, each individual discrete volume may further comprise nucleic acid amplification reagents.

[0412] In specific embodiments, the target molecule may be a target DNA and the individual discrete volumes further comprise a primer that binds the target DNA and comprises an RNA polymerase promoter.

[0413] Samples sources that may be analyzed using the systems and devices described herein include biological samples of a subject or environmental samples. Environmental samples may include surfaces or fluids. The biological samples may include, but are not limited to, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, spinal fluid, cerebrospinal fluid, a swab from skin or a mucosal membrane, or combination thereof. In an example embodiment, the environmental sample is taken from a solid surface, such as a surface used in the preparation of food or other sensitive compositions and materials.

[0414] In other example embodiments, the elements of the systems described herein may be place on a single use substrate, such as swab or cloth that is used to swab a surface or sample fluid. For example, the system could be used to test for the presence of a pathogen on a food by swabbing the surface of a food product, such as a fruit or vegetable. Similarly, the single use substrate may be used to swab other surfaces for detection of certain microbes or agents, such as for use in security screening. Single use substrates may also have applications in forensics, where the CRISPR systems are designed to detect, for example identifying DNA SNPs that may be used to identify a suspect, or certain tissue or cell markers to determine the type of biological matter present in a sample. Likewise, the single use substrate could be used to collect a sample from a patient – such as a saliva sample from the mouth – or a swab of the skin. In other embodiments, a sample or swab may be taken of a meat product on order to detect the presence of absence of contaminants on or within the meat product.

[0415] Near-real-time microbial diagnostics are needed for food, clinical, industrial, and other environmental settings (see e.g., Lu TK, Bowers J, and Koeris MS., Trends Biotechnol. 2013 Jun;31(6):325-7). In certain embodiments, the present invention is used for rapid

detection of foodborne pathogens using guide RNAs specific to a pathogen (e.g., *Campylobacter jejuni*, *Clostridium perfringens*, *Salmonella* spp., *Escherichia coli*, *Bacillus cereus*, *Listeria monocytogenes*, *Shigella* spp., *Staphylococcus aureus*, *Staphylococcal enteritis*, *Streptococcus*, *Vibrio cholerae*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*, *Brucella* spp., *Corynebacterium ulcerans*, *Coxiella burnetii*, or *Plesiomonas shigelloides*).

[0416] In certain embodiments, the device is or comprises a flow strip. For instance, a lateral flow strip allows for detection by color. The reporter is modified to have a first molecule (such as for instance FITC) attached to the 5' end and a second molecule (such as for instance biotin) attached to the 3' end (or vice versa). The lateral flow strip is designed to have two capture lines with anti-first molecule (e.g. anti-FITC) antibodies hybridized at the first line and anti-second molecule (e.g. anti-biotin) antibodies at the second downstream line. As the reaction flows down the strip, uncleaved reporter will bind to anti-first molecule antibodies at the first capture line, while cleaved reporters will liberate the second molecule and allow second molecule binding at the second capture line. Second molecule sandwich antibodies, for instance conjugated to nanoparticles, such as gold nanoparticles, will bind any second molecule at the first or second line and result in a strong readout/signal (e.g. color). As more reporter is cleaved, more signal will accumulate at the second capture line and less signal will appear at the first line. In certain aspects, the invention relates to the use of a follow strip as described herein for detecting nucleic acids or polypeptides. In certain aspects, the invention relates to a method for detecting nucleic acids or polypeptides with a flow strip as defined herein, e.g. (lateral) flow tests or (lateral) flow immunochromatographic assays.

[0417] The embodiments disclosed herein are directed to lateral flow detection devices that comprise CRISPR-Cas systems. The device may comprise a lateral flow substrate for detecting a Cas collateral reaction. Substrates suitable for use in lateral flow assays are known in the art. These may include but are not necessarily limited to membranes or pads made of cellulose and/or glass fiber, polyesters, nitrocellulose, or absorbent pads (J Saudi Chem Soc 19(6):689-705; 2015). The CRISPR-Cas system, i.e. one or more CRISPR-Cas systems and corresponding reporter constructs are added to the lateral flow substrate at a defined reagent portion of the lateral flow substrate, typically on one end of the lateral flow substrate. Reporting constructs used within the context of the present invention comprise a first molecule and a second molecule linked by an DNA linker. The lateral flow substrate further comprises a

sample portion. The sample portion may be equivalent to, continuous with, or adjacent to the reagent portion. The lateral flow strip further comprises a first capture line, typically a horizontal line running across the device, but other configurations are possible. The first capture region is proximate to and on the same end of the lateral flow substrate as the sample loading portion. A first binding agent that specifically binds the first molecule of the reporter construct is fixed or otherwise immobilized to the first capture region. The second capture region is located towards the opposite end of the lateral flow substrate from the first binding region. A second binding agent is fixed or otherwise immobilized at the second capture region. The second binding agent specifically binds the second molecule of the reporter construct, or the second binding agent may bind a detectable ligand. For example, the detectable ligand may be a particle, such as a colloidal particle, that when it aggregates can be detected visually. The particle may be modified with an antibody that specifically binds the second molecule on the reporter construct. If the reporter construct is not cleaved, it will facilitate accumulation of the detectable ligand at the first binding region. If the reporter construct is cleaved the detectable ligand is released to flow to the second binding region. In such an embodiment, the second binding agent is an agent capable of specifically or non-specifically binding the detectable ligand on the antibody on the detectable ligand. Examples of suitable binding agents for such an embodiment include, but are not limited to, protein A and protein G.

[0418] Lateral support substrates may be located within a housing (see for example, “Rapid Lateral Flow Test Strips” Merck Millipore 2013). The housing may comprise at least one opening for loading samples and a second single opening or separate openings that allow for reading of detectable signal generated at the first and second capture regions.

[0419] The CRISPR-Cas system may be freeze-dried to the lateral flow substrate and packaged as a ready to use device, or the CRISPR-Cas system may be added to the reagent portion of the lateral flow substrate at the time of using the device. Samples to be screened are loaded at the sample loading portion of the lateral flow substrate. The samples must be liquid samples or samples dissolved in an appropriate solvent, usually aqueous. The liquid sample reconstitutes the CRISPR-Cas such that a collateral reaction can occur. The liquid sample begins to flow from the sample portion of the substrate towards the first and second capture regions. Intact reporter construct is bound at the first capture region by binding between the first binding agent and the first molecule. Likewise, the detection agent will begin to collect at the first binding region by binding to the second molecule on the intact reporter construct. If

target molecule(s) are present in the sample, the Cas collateral effect is activated. As activated Cas comes into contact with the bound reporter construct, the reporter constructs are cleaved, releasing the second molecule to flow further down the lateral flow substrate towards the second binding region. The released second molecule is then captured at the second capture region by binding to the second binding agent, where additional detection agent may also accumulate by binding to the second molecule. Accordingly, if the target molecule(s) is not present in the sample, a detectable signal will appear at the first capture region, and if the target molecule(s) is present in the sample, a detectable signal will appear at the location of the second capture region.

[0420] Specific binding-integrating molecules comprise any members of binding pairs that can be used in the present invention. Such binding pairs are known to those skilled in the art and include, but are not limited to, antibody-antigen pairs, enzyme-substrate pairs, receptor-ligand pairs, and streptavidin-biotin. In addition to such known binding pairs, novel binding pairs may be specifically designed. A characteristic of binding pairs is the binding between the two members of the binding pair.

[0421] Oligonucleotide Linkers having molecules on either end may comprise DNA if the Cas polypeptide has DNA collateral activity. Oligonucleotide linkers may be single stranded or double stranded, and in certain embodiments, they could contain both RNA and DNA regions. Oligonucleotide linkers may be of varying lengths, such as 5-10 nucleotides, 10-20 nucleotides, 20-50 nucleotides, or more.

[0422] In some embodiments, the polypeptide identifier elements include affinity tags, such as hemagglutinin (HA) tags, Myc tags, FLAG tags, V5 tags, chitin binding protein (CBP) tags, maltose-binding protein (MBP) tags, GST tags, poly-His tags, and fluorescent proteins (for example, green fluorescent protein (GFP), yellow fluorescent protein (YFP), cyan fluorescent protein (CFP), dsRed, mCherry, Kaede, Kindling, and derivatives thereof, FLAG tags, Myc tags, AU1 tags, T7 tags, OLLAS tags, Glu-Glu tags, VSV tags, or a combination thereof. Other Affinity tags are well known in the art. Such labels can be detected and/or isolated using methods known in the art (for example, by using specific binding agents, such as antibodies, that recognize a particular affinity tag). Such specific binding agents (for example, antibodies) can further contain, for example, detectable labels, such as isotope labels and/or nucleic acid barcodes such as those described herein.

[0423] In certain example embodiments, a lateral flow device comprises a lateral flow substrate comprising a first end for application of a sample. The first region is loaded with a detectable ligand, such as those disclosed herein, for example a gold nanoparticle. The gold nanoparticle may be modified with a first antibody, such as an anti-FITC antibody. The first region also comprises a detection construct. In one example embodiment, a DNA detection construct and a CRISPR-Cas system as disclosed herein. In one example embodiment, and for purposes of further illustration, the DNA construct may comprise a FAM molecule on a first end of the detection construction and a biotin on a second end of the detection construct. Upstream of the flow of solution from the first end of the lateral flow substrate is a first test band. The test band may comprise a biotin ligand. Accordingly, when the DNA detection construct is present in its initial state, i.e., in the absence of target, the FAM molecule on the first end will bind the anti-FITC antibody on the gold nanoparticle, and the biotin on the second end of the DNA construct will bind the biotin ligand allowing for the detectable ligand to accumulate at the first test, generating a detectable signal. Generation of a detectable signal at the first band indicates the absence of the target ligand. In the presence of target, the CRISPR-Cas forms and the Cas is activated resulting in cleavage of the DNA detection construct. In the absence of intact DNA detection construct the colloidal gold will flow past the second strip. The lateral flow device may comprise a second band, upstream of the first band. The second band may comprise a molecule capable of binding the antibody-labeled colloidal gold molecule, for example an anti-rabbit antibody capable of binding a rabbit anti-FITC antibody on the colloidal gold. Therefore, in the presence of one or more targets, the detectable ligand will accumulate at the second band, indicating the presence of the one or more targets in the sample.

[0424] In certain example embodiments, the device is a microfluidic device that generates and/or merges different droplets (i.e. individual discrete volumes). For example, a first set of droplets may be formed containing samples to be screened and a second set of droplets formed containing the elements of the systems described herein. The first and second set of droplets are then merged and then diagnostic methods as described herein are carried out on the merged droplet set. Microfluidic devices disclosed herein may be silicone-based chips and may be fabricated using a variety of techniques, including, but not limited to, hot embossing, molding of elastomers, injection molding, LIGA, soft lithography, silicon fabrication and related thin film processing techniques. Suitable materials for fabricating the microfluidic devices include,

but are not limited to, cyclic olefin copolymer (COC), polycarbonate, poly(dimethylsiloxane) (PDMS), and poly(methylacrylate) (PMMA). In one embodiment, soft lithography in PDMS may be used to prepare the microfluidic devices. For example, a mold may be made using photolithography which defines the location of flow channels, valves, and filters within a substrate. The substrate material is poured into a mold and allowed to set to create a stamp. The stamp is then sealed to a solid support, such as but not limited to, glass. Due to the hydrophobic nature of some polymers, such as PDMS, which absorbs some proteins and may inhibit certain biological processes, a passivating agent may be necessary (Schoffner et al. *Nucleic Acids Research*, 1996, 24:375-379). Suitable passivating agents are known in the art and include, but are not limited to, silanes, parylene, n-Dodecyl-b-D-matoside (DDM), pluronic, Tween-20, other similar surfactants, polyethylene glycol (PEG), albumin, collagen, and other similar proteins and peptides.

[0425] In certain example embodiments, the system and/or device may be adapted for conversion to a flow-cytometry readout in or allow to all of sensitive and quantitative measurements of millions of cells in a single experiment and improve upon existing flow-based methods, such as the PrimeFlow assay. In certain example embodiments, cells may be cast in droplets containing unpolymerized gel monomer, which can then be cast into single-cell droplets suitable for analysis by flow cytometry. A detection construct comprising a fluorescent detectable label may be cast into the droplet comprising unpolymerized gel monomer. Upon polymerization of the gel monomer to form a bead within a droplet. Because gel polymerization is through free-radical formation, the fluorescent reporter becomes covalently bound to the gel. The detection construct may be further modified to comprise a linker, such as an amine. A quencher may be added post-gel formation and will bind via the linker to the reporter construct. Thus, the quencher is not bound to the gel and is free to diffuse away when the reporter is cleaved by the Cas polypeptide. Amplification of signal in droplet may be achieved by coupling the detection construct to a hybridization chain reaction (HCR initiator) amplification. DNA/RNA hybrid hairpins may be incorporated into the gel which may comprise a hairpin loop that has a RNase sensitive domain. By protecting a strand displacement toehold within a hairpin loop that has a RNase sensitive domain, HCR initiators may be selectively deprotected following cleavage of the hairpin loop by the CRISPR-Cas system. Following deprotection of HCR initiators via toehold mediated strand displacement, fluorescent HCR monomers may be washed into the gel to enable signal amplification where the initiators are deprotected.

[0426] An example of microfluidic device that may be used in the context of the invention is described in Hour et al. “Direct Detection and drug-resistance profiling of bacteremias using inertial microfluidics” *Lap Chip*. 15(10):2297-2307 (2016).

[0427] In systems described herein, may further be incorporated into wearable medical devices that assess biological samples, such as biological fluids, of a subject outside the clinic setting and report the outcome of the assay remotely to a central server accessible by a medical care professional. The device may include the ability to self-sample blood, such as the devices disclosed in U.S. Patent Application Publication No. 2015/0342509 entitled “Needle-free Blood Draw to Peeters et al., U.S. Patent Application Publication No. 2015/0065821 entitled “Nanoparticle Phoresis” to Andrew Conrad.

[0428] In some embodiments, the individual discrete volumes are microwells.

[0429] In certain example embodiments, the device may comprise individual wells, such as microplate wells. The size of the microplate wells may be the size of standard 6, 24, 96, 384, 1536, 3456, or 9600 sized wells. In certain example embodiments, the elements of the systems described herein may be freeze dried and applied to the surface of the well prior to distribution and use.

[0430] The devices disclosed herein may further comprise inlet and outlet ports, or openings, which in turn may be connected to valves, tubes, channels, chambers, and syringes and/or pumps for the introduction and extraction of fluids into and from the device. The devices may be connected to fluid flow actuators that allow directional movement of fluids within the microfluidic device. Example actuators include, but are not limited to, syringe pumps, mechanically actuated recirculating pumps, electroosmotic pumps, bulbs, bellows, diaphragms, or bubbles intended to force movement of fluids. In certain example embodiments, the devices are connected to controllers with programmable valves that work together to move fluids through the device. In certain example embodiments, the devices are connected to the controllers discussed in further detail below. The devices may be connected to flow actuators, controllers, and sample loading devices by tubing that terminates in metal pins for insertion into inlet ports on the device.

[0431] As shown herein the elements of the system are stable when freeze dried, therefore embodiments that do not require a supporting device are also contemplated, i.e. the system may be applied to any surface or fluid that will support the reactions disclosed herein and allow for detection of a positive detectable signal from that surface or solution. In addition to freeze-

drying, the systems may also be stably stored and utilized in a pelletized form. Polymers useful in forming suitable pelletized forms are known in the art.

[0432] In some embodiments, the individual discrete volumes are defined on a solid substrate. In some embodiments, the individual discrete volumes are spots defined on a substrate. In some embodiments, the substrate may be a flexible materials substrate, for example, including, but not limited to, a paper substrate, a fabric substrate, or a flexible polymer-based substrate. In specific embodiments, the flexible materials substrate is a paper substrate or a flexible polymer-based substrate.

[0433] In certain embodiments, the Cas polypeptide is bound to each discrete volume in the device. Each discrete volume may comprise a different nucleic acid component specific for a different target molecule. In certain embodiments, a sample is exposed to a solid substrate comprising more than one discrete volume each comprising a nucleic acid component specific for a target molecule. Not being bound by a theory, each nucleic acid component will capture its target molecule from the sample and the sample does not need to be divided into separate assays. Thus, a valuable sample may be preserved. The effector protein may be a fusion protein comprising an affinity tag. Affinity tags are well known in the art (e.g., HA tag, Myc tag, Flag tag, His tag, biotin). The effector protein may be linked to a biotin molecule and the discrete volumes may comprise streptavidin. In other embodiments, the CRISPR effector protein is bound by an antibody specific for the effector protein. Methods of binding a CRISPR enzyme has been described previously (see, e.g., US20140356867A1).

[0434] The devices disclosed herein may also include elements of point of care (POC) devices known in the art for analyzing samples by other methods. See, for example St John and Price, “Existing and Emerging Technologies for Point-of-Care Testing” (Clin Biochem Rev. 2014 Aug; 35(3): 155–167).

[0435] The present invention may be used with a wireless lab-on-chip (LOC) diagnostic sensor system (see e.g., US patent number 9,470,699 “Diagnostic radio frequency identification sensors and applications thereof”). In certain embodiments, the present invention is performed in a LOC controlled by a wireless device (e.g., a cell phone, a personal digital assistant (PDA), a tablet) and results are reported to said device.

[0436] Radio frequency identification (RFID) tag systems include an RFID tag that transmits data for reception by an RFID reader (also referred to as an interrogator). In a typical RFID system, individual objects (e.g., store merchandise) are equipped with a relatively small

tag that contains a transponder. The transponder has a memory chip that is given a unique electronic product code. The RFID reader emits a signal activating the transponder within the tag through the use of a communication protocol. Accordingly, the RFID reader is capable of reading and writing data to the tag. Additionally, the RFID tag reader processes the data according to the RFID tag system application. Currently, there are passive and active type RFID tags. The passive type RFID tag does not contain an internal power source, but is powered by radio frequency signals received from the RFID reader. Alternatively, the active type RFID tag contains an internal power source that enables the active type RFID tag to possess greater transmission ranges and memory capacity. The use of a passive versus an active tag is dependent upon the particular application.

[0437] Lab-on-the chip technology is well described in the scientific literature and consists of multiple microfluidic channels, input or chemical wells. Reactions in wells can be measured using radio frequency identification (RFID) tag technology since conductive leads from RFID electronic chip can be linked directly to each of the test wells. An antenna can be printed or mounted in another layer of the electronic chip or directly on the back of the device. Furthermore, the leads, the antenna and the electronic chip can be embedded into the LOC chip, thereby preventing shorting of the electrodes or electronics. Since LOC allows complex sample separation and analyses, this technology allows LOC tests to be done independently of a complex or expensive reader. Rather a simple wireless device such as a cell phone or a PDA can be used. In one embodiment, the wireless device also controls the separation and control of the microfluidics channels for more complex LOC analyses. In one embodiment, a LED and other electronic measuring or sensing devices are included in the LOC-RFID chip. Not being bound by a theory, this technology is disposable and allows complex tests that require separation and mixing to be performed outside of a laboratory.

[0438] In preferred embodiments, the LOC may be a microfluidic device. The LOC may be a passive chip, wherein the chip is powered and controlled through a wireless device. In certain embodiments, the LOC includes a microfluidic channel for holding reagents and a channel for introducing a sample. In certain embodiments, a signal from the wireless device delivers power to the LOC and activates mixing of the sample and assay reagents. Specifically, in the case of the present invention, the system may include a masking agent, CRISPR effector protein, and guide RNAs specific for a target molecule. Upon activation of the LOC, the microfluidic device may mix the sample and assay reagents. Upon mixing, a sensor detects a

signal and transmits the results to the wireless device. In certain embodiments, the unmasking agent is a conductive RNA molecule. The conductive RNA molecule may be attached to the conductive material. Conductive molecules can be conductive nanoparticles, conductive proteins, metal particles that are attached to the protein or latex or other beads that are conductive. In certain embodiments, if DNA or RNA is used then the conductive molecules can be attached directly to the matching DNA or RNA strands. The release of the conductive molecules may be detected across a sensor. The assay may be a one step process.

[0439] Since the electrical conductivity of the surface area can be measured precisely quantitative results are possible on the disposable wireless RFID electro-assays. Furthermore, the test area can be very small allowing for more tests to be done in a given area and therefore resulting in cost savings. In certain embodiments, separate sensors each associated with a different CRISPR effector protein and guide RNA immobilized to a sensor are used to detect multiple target molecules. Not being bound by a theory, activation of different sensors may be distinguished by the wireless device.

[0440] In addition to the conductive methods described herein, other methods may be used that rely on RFID or Bluetooth as the basic low cost communication and power platform for a disposable RFID assay. For example, optical means may be used to assess the presence and level of a given target molecule. In certain embodiments, an optical sensor detects unmasking of a fluorescent masking agent.

[0441] In certain embodiments, the device of the present invention may include handheld portable devices for diagnostic reading of an assay (see e.g., Vashist et al., *Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management*, *Diagnostics* 2014, 4(3), 104-128; mReader from Mobile Assay; and Holomic Rapid Diagnostic Test Reader).

[0442] As noted herein, certain embodiments allow detection via colorimetric change which has certain attendant benefits when embodiments are utilized in POC situations and or in resource poor environments where access to more complex detection equipment to readout the signal may be limited. However, portable embodiments disclosed herein may also be coupled with hand-held spectrophotometers that enable detection of signals outside the visible range. An example of a hand-held spectrophotometer device that may be used in combination with the present invention is described in Das et al. "Ultra-portable, wireless smartphone spectrophotometer for rapid, non-destructive testing of fruit ripeness." *Nature Scientific*

Reports. 2016, 6:32504, DOI: 10.1038/srep32504. Finally, in certain embodiments utilizing quantum dot-based masking constructs, use of a hand held UV light, or other suitable device, may be successfully used to detect a signal owing to the near complete quantum yield provided by quantum dots.

Methods for Detecting Nucleic Acids

[0443] The low cost and adaptability of the assay platform lends itself to a number of applications including (i) general RNA/DNA quantitation, (ii) rapid, multiplexed RNA/DNA and protein expression detection, and (iii) sensitive detection of target nucleic acids, peptides, and proteins in both clinical and environmental samples. Additionally, the systems disclosed herein may be adapted for detection of transcripts within biological settings, such as cells. Given the highly specific nature of the CRISPR effectors described herein, it may be possible to track allelic specific expression of transcripts or disease-associated mutations in live cells.

[0444] In some embodiments, methods include detecting target nucleic acids in samples, comprising distributing a sample or set of samples into one or more individual discrete volumes comprising a CRISPR-Cas system as described herein. The sample or set of samples may then be incubated under conditions sufficient to allow binding of the one or more nucleic acid components to one or more target molecules, and the Cas polypeptide may be activated via binding of the one or more nucleic acid component to the one or more target molecules, wherein activating the CRISPR effector protein results in modification of the detection construct such that a detectable positive signal is generated. The one or more detectable positive signals may then be detected, with detection indicating the presence of one or more target molecules in the sample.

[0445] In some embodiments, methods of the invention include detecting polypeptides in samples, comprising distributing a sample or set of samples into a set of individual discrete volumes comprising peptide detection aptamers and a Cas polypeptide as described herein. The sample or set of samples may then be incubated under conditions sufficient to allow binding of the peptide detection aptamers to the one or more target molecules, wherein binding of the aptamer to a corresponding target molecule exposes the RNA polymerase binding site or primer binding site resulting in generation of a trigger RNA. The Cas may then be activated via binding of the one or more nucleic acid components to the trigger RNA, wherein activating the Cas polypeptide results in modification of the detection construct such that a detectable positive

signal is produced. The detectable positive signal may then be detected, with detection of the detectable positive signal indicating the presence of one or more target molecules in a sample.

[0446] In certain example embodiments, a single guide sequence specific to a single target is placed in separate volumes. Each volume may then receive a different sample or aliquot of the same sample. In certain example embodiments, multiple nucleic acid components each to separate target may be placed in a single well such that multiple targets may be screened in a different well. In order to detect multiple nucleic acid component in a single volume, in certain example embodiments, multiple Cas polypeptides with different specificities may be used.

[0447] In embodiments, different Cas orthologs with different sequence specificities may be used. Cutting motifs may be used to take advantage of the sequence specificities of different orthologs. The detection construct can comprise a cutting motif preferentially cut by a given Cas ortholog. A cutting motif sequence can be a particular nucleotide base, a repeat nucleotide base in a homopolymer, or a heteropolymer of bases. The cutting motif can be a dinucleotide sequence, a trinucleotide sequence or more complex motifs comprising 4, 5, 6, 7, 8, 9, or 10 nucleotide motifs. For example, one orthologue may preferentially cut A, while others preferentially cut C, G, U/ T. Accordingly, detection constructs completely comprising, or comprised of a substantial portion, of a single nucleotide may be generated, each with a different fluorophore that can be detected at differing wavelengths. In this way up to four different targets may be screened in a single individual discrete volume. In certain other example embodiments, different orthologues with different nucleotide editing preferences may be used in combination with a Cas13 or Cas12.

[0448] In addition to single base editing preferences, additional detection constructs can be designed based on other motif cutting preferences of Cas12, and Cas13 orthologs. For example, Cas13 or Cas12 orthologs may preferentially cut a dinucleotide sequence, a trinucleotide sequence or more complex motifs comprising 4, 5, 6, 7, 8, 9, or 10 nucleotide motifs. As an example, LwaCas13a showed strong preference for a hexanucleotide motif sequences, with CcaCas13b showing strong preference for other hexanucleotide motifs. Thus the upper bound for multiplex assays using the embodiments disclosed herein is primarily limited by the number of distinguishable detectable labels and the detection channels needed to detect them. In certain example embodiments, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 25, 27, 28, 29, or 30 different targets are detected.

[0449] In specific embodiments, the target molecule may be a target DNA and the method may further comprise binding the target DNA with a primer comprising an RNA polymerase site, as described herein.

[0450] In specific embodiments, the one or more nucleic acid component may be designed to detect a single nucleotide polymorphism in a target RNA or DNA, or a splice variant of an RNA transcript.

[0451] A sample for use with the invention may be a biological or environmental sample, such as a food sample (fresh fruits or vegetables, meats), a beverage sample, a paper surface, a fabric surface, a metal surface, a wood surface, a plastic surface, a soil sample, a freshwater sample, a wastewater sample, a saline water sample, exposure to atmospheric air or other gas sample, or a combination thereof. For example, household/commercial/industrial surfaces made of any materials including, but not limited to, metal, wood, plastic, rubber, or the like, may be swabbed and tested for contaminants. Soil samples may be tested for the presence of pathogenic bacteria or parasites, or other microbes, both for environmental purposes and/or for human, animal, or plant disease testing. Water samples such as freshwater samples, wastewater samples, or saline water samples can be evaluated for cleanliness and safety, and/or potability, to detect the presence of, for example, *Cryptosporidium parvum*, *Giardia lamblia*, or other microbial contamination. In further embodiments, a biological sample may be obtained from a source including, but not limited to, a tissue sample, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, cerebrospinal fluid, ascites, pleural effusion, seroma, pus, bile, aqueous or vitreous humor, transudate, exudate, or swab of skin or a mucosal membrane surface. In some particular embodiments, an environmental sample or biological samples may be crude samples and/or the one or more target molecules may not be purified or amplified from the sample prior to application of the method. Identification of microbes may be useful and/or needed for any number of applications, and thus any type of sample from any source deemed appropriate by one of skill in the art may be used in accordance with the invention.

[0452] In some embodiments, the one or more nucleic acid components may be designed to bind to cell free nucleic acids. In some embodiments, the one or more nucleic acid components may be designed to detect a single nucleotide polymorphism in a target RNA or DNA, or a splice variant of an RNA transcript. In some embodiments, the one or more guide

RNAs are designed to bind to one or more target molecules that are diagnostic for a disease state, as described herein.

[0453] In some embodiments, the disease state may be an infection, an organ disease, a blood disease, an immune system disease, a cancer, a brain and nervous system disease, an endocrine disease, a pregnancy or childbirth-related disease, an inherited disease, or an environmentally-acquired disease.

[0454] In certain example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting the presence of one or more microbial agents in a sample, such as a biological sample obtained from a subject. In certain example embodiments, the microbe may be a bacterium, a fungus, a yeast, a protozoa, a parasite, or a virus. Accordingly, the methods disclosed herein can be adapted for use in other methods (or in combination) with other methods that require quick identification of microbe species, monitoring the presence of microbial proteins (antigens), antibodies, antibody genes, detection of certain phenotypes (e.g. bacterial resistance), monitoring of disease progression and/or outbreak, and antibiotic screening. Because of the rapid and sensitive diagnostic capabilities of the embodiments disclosed here, detection of microbe species type, down to a single nucleotide difference, and the ability to be deployed as a POC device, the embodiments disclosed herein may be used guide therapeutic regimens, such as selection of the appropriate antibiotic or antiviral. The embodiments disclosed herein may also be used to screen environmental samples (air, water, surfaces, food etc.) for the presence of microbial contamination.

[0455] Disclosed is a method to identify microbial species, such as bacterial, viral, fungal, yeast, or parasitic species, or the like. Particular embodiments disclosed herein describe methods and systems that will identify and distinguish microbial species within a single sample, or across multiple samples, allowing for recognition of many different microbes. The present methods allow the detection of pathogens and distinguishing between two or more species of one or more organisms, e.g., bacteria, viruses, yeast, protozoa, and fungi or a combination thereof, in a biological or environmental sample, by detecting the presence of a target nucleic acid sequence in the sample. A positive signal obtained from the sample indicates the presence of the microbe. Multiple microbes can be identified simultaneously using the methods and systems of the invention, by employing the use of more than one effector protein, wherein each effector protein targets a specific microbial target sequence. In this way, a multi-level analysis can be performed for a particular subject in which any number of microbes can

be detected at once. In some embodiments, simultaneous detection of multiple microbes may be performed using a set of probes that can identify one or more microbial species.

[0456] Multiplex analysis of samples enables large-scale detection of samples, reducing the time and cost of analyses. However, multiplex analyses are often limited by the availability of a biological sample. In accordance with the invention, however, alternatives to multiplex analysis may be performed such that multiple effector proteins can be added to a single sample and each masking construct may be combined with a separate quencher dye. In this case, positive signals may be obtained from each quencher dye separately for multiple detection in a single sample.

[0457] Disclosed herein are methods for distinguishing between two or more species of one or more organisms in a sample. The methods are also amenable to detecting one or more species of one or more organisms in a sample.

[0458] In some embodiments, the methods provide for detection of disease states that are characterized by the presence or absence of an antibiotic or drug resistance or susceptibility gene or transcript or polypeptide, preferably in a pathogen or a cell.

Devices for Detection Assays

[0459] In one embodiment, the detection assay can be provided on a cartridge or chip. In an aspect, the cartridge can comprise one or more ampoules and one or more wells that are communicatively coupled, allowing for the transfer, exchange or movement of reagents and sample with or without the use of beads through the chambers of the cartridge and facilitating detection assays utilizing systems/devices for facilitating the detection assay on the cartridge.

Cartridge

[0460] The cartridge, also referred to herein as a chip, according to the present invention comprises a series of components of ampoules and chambers that are communicatively coupled with one or more other components on the cartridge. The coupling is typically a fluidic communication, for example, via channels. The cartridge may comprise a membrane that seals one or more of the chambers and/or ampoules. In an aspect, the membrane allows for storage of reagents, buffers and other solid or fluid components which cover and seal the cartridge. The membrane can be configured to be punctured, pierced, or otherwise released from sealing or covering one or more components of the cartridge by a means for releasing reagents.

[0461] As noted above, certain embodiments enable the use of nucleic acid binding beads to concentrate target nucleic acid but that do not require elution of the isolated nucleic acid.

Thus, in certain example embodiments, the cartridge may further comprise an activatable magnet, such as an electro-magnet. A means for activating the magnet may be located on the device, or the means for supplying the magnet or activating the magnet on the cartridge may be provided by a second device, such as those disclosed in further detail below.

Ampoules

[0462] The ampoules, also referred to as blisters, allow for storage and release of reagents throughout the cartridge. Ampoules can include liquid or solid reagents, for example, lysis reagents in one ampoule and reaction reagents in another ampoule. The reagents can be as described elsewhere herein and can be adapted for the use in the cartridge. The ampoule may be sealed by a film that allows for the bursting, puncture, or other release of the contents of the ampoules. See, e.g. Becker, H. & Gärtner, C. Microfluidics-enabled diagnostic systems: markets, challenges, and examples. In *Microchip Diagnostics: Methods and Protocols* (eds Taly, V. *et al.*) (Springer, New York, 2017); Czurratis *et al.*, doi: 10.1088/0960-1317/25/4/045002. Considerations for ampoules can include as discussed in, for example, Smith, S., et al., Blister pouches for effective reagent storage on microfluidic chips for blood cell counting. *Microfluid Nanofluid* 20, 163 (2016). DOI:10.1007/s10404-016-1830-2. In an aspect, the seal is a frangible seal formed of a composite-layer film that is assembled to the cartridge main body. While referred to herein as an ampoule, the ampoule may comprise a cavity on a chip which comprises a sealed film that is opened by the release means.

Chambers

[0463] The chambers on the chip may be located and sized for fluidic communication via channels or other communication means with ampoules and/or other chambers on the chip.

Means for Reading the Results of the Assay

[0464] A means for reading the results of the assay can be provided in the system. The means for reading the results of the assay will depend in part on the type of detectable signal generated by the assay. In particular embodiments, the assay generates a detectable fluorescent or color readout. In these instances, the means for reading the results of the assay will be an optic means, for example a single channel or multi-channel optical means such as a fluorimeter, colorimeter, or other spectroscopic sensor.

[0465] A combination of means for reading the results of the assay can be utilized, and may include readings such as turbidity, temperature, magnetic, radio, or electrical properties and or optical properties, including scattering, polarization effects, etc.

[0466] The system may further comprise a user interface for programming the device and/or readout of the results of the assay. The user interface may comprise an LED screen. The system can be further configured for a USB port that can allow for docking of four or more devices.

[0467] In an aspect, the system comprises a means for activating a magnet that is disposed within or on the cartridge.

Lateral Flow Devices

[0468] In one embodiment, the detection assay can be provided on a lateral flow device, see, e.g. International Publication WO 2019/071051, incorporated herein by reference for exemplary lateral flow devices. The lateral flow device can be adapted to detect one or more coronaviruses and/or other viruses in combination of the coronavirus. The lateral flow device may comprise a flexible substrate, such as a paper substrate or a flexible polymer-based substrate, which can include freeze-dried reagents for detection assays with a visual readout of the assay results. See, WO 2019/071051 at [0145]-[0151] and Example 2, specifically incorporated herein by reference. In an aspect, lyophilized reagents can include preferred excipients that aid in rate of reaction, specificity, or other variable, for example, trehalose, histidine, and/or glycine. In one embodiment, a coronavirus assay can be utilized with isothermal amplification reagents, allowing amplification without complex instrumentation that may be unavailable in the field. Accordingly, the assay can be adapted for field diagnostics, including use of visual readout on a lateral flow device, rapid, sensitive detection and can be deployed for early and direct detection. Colorimetric detection can be utilized and may be particularly suited for field deployable applications, as described in International Application PCT/US2019/015726, published as WO2019/148206. In particular, colorimetric detection can be as described in WO2019/148206 at Figures 102, 105, 107-111 and [00306]-[00324], incorporated herein by reference and may be utilized with the CRISPR-Cas systems.

[0469] In one embodiment, the invention provides a lateral flow device comprising a substrate comprising a first end and a second end. The first end may comprise a sample loading portion, a first region comprising a detectable ligand, two or more CRISPR-Cas systems, two or more detection constructs, and one or more first capture regions, each comprising a first binding agent. The substrate may also comprise two or more second capture regions between the first region of the first end and the second end, each second capture region comprising a different binding agent. Each of the two or more CRISPR-Cas systems may comprise a Cas

polypeptide and one or more nucleic acid component molecules, each nucleic acid component molecule sequence configured to bind one or more target molecules.

[0470] The device may comprise a lateral flow substrate for detecting a reaction between a Cas polypeptide and a target molecule triggering collateral, non-specific cleavage of detection construct. Substrates suitable for use in lateral flow assays are known in the art. These may include, but are not necessarily limited to, membranes or pads made of cellulose and/or glass fiber, polyesters, nitrocellulose, or absorbent pads (*J Saudi Chem Soc* 19(6):689-705; 2015), and other embodiments further described herein. The detection system, i.e. one or more CRISPR-Cas systems and corresponding reporter constructs are added to the lateral flow substrate at a defined reagent portion of the lateral flow substrate, typically on one end of the lateral flow substrate. Reporting constructs used within the context of the present invention can comprise a first molecule and a second molecule linked by an RNA or DNA linker. The lateral flow substrate further comprises a sample portion. The sample portion may be equivalent to, continuous with, or adjacent to the reagent portion. In an aspect, the lateral flow substrate can be contained within a further device. In an aspect, the lateral flow substrate can be utilized for visual readout of a detectable signal in one-pot reactions, e.g. wherein steps of extracting, amplifying and detecting are performed in an individual discrete volume.

Lateral Flow Substrate

[0471] In certain example embodiments, a lateral flow device comprises a lateral flow substrate on which detection can be performed. Substrates suitable for use in lateral flow assays are known in the art. These may include, but are not necessarily limited to, membranes or pads made of cellulose and/or glass fiber, polyesters, nitrocellulose, or absorbent pads (*J Saudi Chem Soc* 19(6):689-705; 2015).

[0472] Lateral support substrates comprise a first and second end, and one or more capture regions that each comprise binding agents. The first end may comprise a sample loading portion, a first region comprising a detectable ligand, two or more CRISPR-Cas systems, two or more detection constructs, and one or more first capture regions, each comprising a first binding agent. The substrate may also comprise two or more second capture regions between the first region of the first end and the second end, each second capture region comprising a different binding agent. Each of the two or more CRISPR-Cas systems may comprise a Cas polypeptide and one or more nucleic acid component molecules, each nucleic acid component configured to bind one or more target molecules. The lateral flow substrates may be configured

to detect a reaction wherein collateral, non-specific cleavage is triggered upon binding and cleavage of a target molecule in the reaction by the Cas polypeptide.

[0473] Lateral support substrates may be located within a housing (see for example, “Rapid Lateral Flow Test Strips” Merck Millipore 2013). The housing may comprise at least one opening for loading samples and a second single opening or separate openings that allow for reading of detectable signal generated at the first and second capture regions.

[0474] The embodiments disclosed herein can be prepared in freeze-dried format for convenient distribution and point-of-care (POC) applications. Such embodiments are useful in multiple scenarios in human health including, for example, viral detection, bacterial strain typing, sensitive genotyping, and detection of disease-associated cell free DNA. Accordingly, the lateral substrate comprising one or more of the elements of the system, including detectable ligands, CRISPR-Cas systems, detection constructs and binding agents may be freeze-dried to the lateral flow substrate and packaged as a ready to use device. Alternatively, all or a portion of the elements of the system may be added to the reagent portion of the lateral flow substrate at the time of using the device.

First End and Second End of the Substrate

[0475] The substrate of the lateral flow device comprises a first and second end. The CRISPR-Cas system, i.e. one or more CRISPR-Cas systems and corresponding reporter constructs are added to the lateral flow substrate at a defined reagent portion of the lateral flow substrate, typically on a first end of the lateral flow substrate. Reporting constructs used within the context of the present invention comprise a first molecule and a second molecule linked by an RNA or DNA linker. The lateral flow substrate further comprises a sample portion. The sample portion may be equivalent to, continuous with, or adjacent to the reagent portion.

[0476] In certain example embodiments, the first end comprises a first region. The first region comprises a detectable ligand, two or more CRISPR-Cas systems, two or more detection constructs, and one or more first capture regions, each comprising a first binding agent.

Capture Regions

[0477] The lateral flow substrate can comprise one or more capture regions. In embodiments the first end of the lateral flow substrate comprises one or more first capture regions, with two or more second capture regions between the first region of the first end of the substrate and the second end of the substrate. The capture regions may be provided as a capture line, typically a horizontal line running across the device, but other configurations are

possible. The first capture region is proximate to and on the same end of the lateral flow substrate as the sample loading portion.

Binding Agents

[0478] Specific binding-integrating molecules comprise any members of binding pairs that can be used in the present invention. Such binding pairs are known to those skilled in the art and include, but are not limited to, antibody-antigen pairs, enzyme-substrate pairs, receptor-ligand pairs, and streptavidin-biotin. In addition to such known binding pairs, novel binding pairs may be specifically designed. A characteristic of binding pairs is the binding between the two members of the binding pair.

[0479] A first binding agent that specifically binds the first molecule of the reporter construct is fixed or otherwise immobilized to the first capture region. The second capture region is located towards the opposite end of the lateral flow substrate from the first capture region. A second binding agent is fixed or otherwise immobilized at the second capture region. The second binding agent specifically binds the second molecule of the reporter construct, or the second binding agent may bind a detectable ligand. For example, the detectable ligand may be a particle, such as a colloidal particle, that when it aggregates can be detected visually, and generates a detectable positive signal. The particle may be modified with an antibody that specifically binds the second molecule on the reporter construct. If the reporter construct is not cleaved, it will facilitate accumulation of the detectable ligand at the first binding region. If the reporter construct is cleaved the detectable ligand is released to flow to the second binding region. In such an embodiment, the second binding region comprises a second binding agent capable of specifically or non-specifically binding the detectable ligand on the antibody of the detectable ligand. Binding agents can be, for example, antibodies, that recognize a particular affinity tag. Such binding agents can further contain, for example, detectable labels, such as isotope labels and/or nucleic acid barcodes. A barcode is a short sequence of nucleotides (for example, DNA, RNA, or combinations thereof) that is used as an identifier. A nucleic acid barcode may have a length of 4-100 nucleotides and be either single or double-stranded. Methods for identifying cells with barcodes are known in the art. Accordingly, nucleic acid component molecules of the CRISPR-Cas systems described herein may be used to detect the barcode.

Detectable Ligands

[0480] The first region is loaded with a detectable ligand, such as those disclosed herein, for example a gold nanoparticle. The detectable ligand may be a particle, such as a colloidal particle, that when it aggregates can be detected visually. The particle may be modified with an antibody that specifically binds the second molecule on the reporter construct. If the reporter construct is not cleaved, it will facilitate accumulation of the detectable ligand at the first binding region. If the reporter construct is cleaved the detectable ligand is released to flow to the second binding region. In such an embodiment, the second binding agent is an agent capable of specifically or non-specifically binding the detectable ligand on the antibody on the detectable ligand. Examples of suitable binding agents for such an embodiment include, but are not limited to, protein A and protein G. In some examples, the detectable ligand is a gold nanoparticle, which may be modified with a first antibody, such as an anti-FITC antibody.

Lateral Flow Detection Constructs

[0481] The first region also comprises a detection construct. In one example embodiment, a RNA detection construct and a CRISPR-Cas system (a Cas polypeptide and one or more nucleic acid component molecules configured to bind to one or more target sequences) as disclosed herein. In one example embodiment, and for purposes of further illustration, the RNA construct may comprise a FAM molecule on a first end of the detection construction and a biotin on a second end of the detection construct. Upstream of the flow of solution from the first end of the lateral flow substrate is a first test band. The test band may comprise a biotin ligand. Accordingly, when the RNA detection construct is present in its initial state, i.e. in the absence of target, the FAM molecule on the first end will bind the anti-FITC antibody on the gold nanoparticle, and the biotin on the second end of the RNA construct will bind the biotin ligand allowing for the detectable ligand to accumulate at the first test, generating a detectable signal. Generation of a detectable signal at the first band indicates the absence of the target ligand. In the presence of target, the CRISPR-Cas forms and the Cas polypeptide is activated resulting in cleavage of the detection construct. In the absence of intact RNA detection construct the colloidal gold will flow past the second strip. The lateral flow device may comprise a second band, upstream of the first band. The second band may comprise a molecule capable of binding the antibody-labeled colloidal gold molecule, for example an anti-rabbit antibody capable of binding a rabbit anti-FITC antibody on the colloidal gold. Therefore, in

the presence of one or more targets, the detectable ligand will accumulate at the second band, indicating the presence of the one or more targets in the sample.

[0482] In one embodiment, the first end of the lateral flow device comprises two detection constructs and each of the two detection constructs comprises an RNA or DNA oligonucleotide, comprising a first molecule on a first end and a second molecule on a second end. The first molecule and the second molecule may be linked by an RNA or DNA linker.

[0483] In one embodiment, the first molecule on the first end of the first detection construct may be FAM and the second molecule on the second end of the first detection construct may be biotin, or vice versa. In one embodiment, the first molecule on the first end of the second detection construct may be FAM and the second molecule on the second end of the second detection construct may be Digoxigenin (DIG), or vice versa.

[0484] In one embodiment, the first end may comprise three detection constructs, wherein each of the three detection constructs comprises an RNA or DNA oligonucleotide, comprising a first molecule on a first end and a second molecule on a second end. In specific embodiments, the first and second molecules on the detection constructs comprise Tye 665 and Alexa 488; Tye 665 and FAM, and Tye 665 and Digoxigenin (DIG), respectively.

[0485] In one embodiment, the first end of the lateral flow device comprises two or more CRISPR-Cas systems, also referred to as a CRISPR-Cas system. In one embodiment, such a CRISPR-Cas system may include a Cas polypeptide and one or more nucleic acid component molecules configured to bind to one or more target sequences.

Samples

[0486] When utilizing the detection systems with a lateral flow substrate, samples to be screened are loaded at the sample loading portion of the lateral flow substrate. The samples must be liquid samples or samples dissolved in an appropriate solvent, usually aqueous. The liquid sample reconstitutes the detection reagents such that a detection reaction can occur. The liquid sample begins to flow from the sample portion of the substrate towards the first and second capture regions.

[0487] A sample for use with the invention may be a biological or environmental sample, such as a surface sample, a fluid sample, or a food sample (fresh fruits or vegetables, meats). Food samples may include a beverage sample, a paper surface, a fabric surface, a metal surface, a wood surface, a plastic surface, a soil sample, a freshwater sample, a wastewater sample, a saline water sample, exposure to atmospheric air or other gas sample, or a combination thereof.

For example, household/commercial/industrial surfaces made of any materials including, but not limited to, metal, wood, plastic, rubber, or the like, may be swabbed and tested for contaminants. Soil samples may be tested for the presence of pathogenic bacteria or parasites, or other microbes, both for environmental purposes and/or for human, animal, or plant disease testing. Water samples such as freshwater samples, wastewater samples, or saline water samples can be evaluated for cleanliness and safety, and/or potability, to detect the presence of, for example, *Cryptosporidium parvum*, *Giardia lamblia*, or other microbial contamination. In further embodiments, a biological sample may be obtained from a source including, but not limited to, a tissue sample, saliva, blood, plasma, sera, stool, urine, sputum, mucous, lymph, synovial fluid, spinal fluid, cerebrospinal fluid, ascites, pleural effusion, seroma, pus, bile, aqueous or vitreous humor, transudate, exudate, or swab of skin or a mucosal membrane surface. In some particular embodiments, an environmental sample or biological samples may be crude samples and/or the one or more target molecules may not be purified or amplified from the sample prior to application of the method. Identification of microbes may be useful and/or needed for any number of applications, and thus any type of sample from any source deemed appropriate by one of skill in the art may be used in accordance with the invention.

[0488] In particular embodiments, the methods and systems can be utilized for direct detection from patient samples. In an aspect, the methods and systems can further allow for direct detection from patient samples with a visual readout to further facilitate field-deployability. In an aspect, a field deployable version can include, for example the lateral flow devices and systems as described herein, and/or colorimetric detection. The methods and systems can be utilized to distinguish multiple viral species and strains and identify clinically relevant mutations, important with viral outbreaks such as the coronavirus outbreak in Wuhan (2019-nCoV). In an aspect, the sample is from a nasopharyngeal swab or a saliva sample. See., e.g. Wyllie et al., “Saliva is more sensitive for SARS-CoV-2 detection in COVID-19 patients than nasopharyngeal swabs,” DOI: 10.1101/2020.04.16.20067835.

Methods for Detecting and/or Quantifying Target Nucleic Acids

[0489] In one embodiment, the invention provides methods for detecting target nucleic acids in a sample. Such methods may comprise contacting a sample with the first end of a lateral flow device as described herein. The first end of the lateral flow device may comprise a sample loading portion, wherein the sample flows from the sample loading portion of the substrate towards the first and second capture regions and generates a detectable signal.

[0490] 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, as described elsewhere herein.

[0491] In one embodiment, the lateral flow device may be capable of detecting two different target nucleic acid sequences. In one embodiment, this detection of two different target nucleic acid sequences may occur simultaneously.

[0492] In one embodiment, the absence of target nucleic acid sequences in a sample elicits a detectable fluorescent signal at each capture region. In such instances, the absence of any target nucleic acid sequences in a sample may cause a detectable signal to appear at the first and second capture regions.

[0493] In one embodiment, the lateral flow device as described herein is capable of detecting three different target nucleic acid sequences. In specific embodiments, when the target nucleic acid sequences are absent from the sample, a fluorescent signal may be generated at each of the three capture regions. In such exemplary embodiments, a fluorescent signal may be absent at the capture region for the corresponding target nucleic acid sequence when the sample contains one or more target nucleic acid sequences.

[0494] Samples to be screened are loaded at the sample loading portion of the lateral flow substrate. The samples must be liquid samples or samples dissolved in an appropriate solvent, usually aqueous. The liquid sample reconstitutes the system reagents such that a detection reaction can occur. Intact reporter construct is bound at the first capture region by binding between the first binding agent and the first molecule. Likewise, the detection agent will begin to collect at the first binding region by binding to the second molecule on the intact reporter construct. If target molecule(s) are present in the sample, the Cas polypeptide collateral effect is activated. As activated Cas polypeptide comes into contact with the bound reporter construct, the reporter constructs are cleaved, releasing the second molecule to flow further down the lateral flow substrate towards the second binding region. The released second molecule is then captured at the second capture region by binding to the second binding agent, where additional detection agent may also accumulate by binding to the second molecule. Accordingly, if the target molecule(s) is not present in the sample, a detectable signal will appear at the first capture region, and if the target molecule(s) is present in the sample, a detectable signal will appear at the location of the second capture region.

[0495] In one embodiment, the invention provides a method for quantifying target nucleic acids in samples comprising distributing a sample or set of samples into one or more individual discrete volumes comprising two or more CRISPR-Cas systems as described herein. The method may comprise using HDA to amplify one or more target molecules in the sample or set of samples, as described herein. The method may further comprise incubating the sample or set of samples under conditions sufficient to allow binding of the nucleic acid component molecules to one or more target molecules. The method may further comprise activating the Cas polypeptide via binding of the nucleic acid component molecules to the one or more target molecules. Activating the Cas polypeptide may result in modification of the detection construct such that a detectable positive signal is generated. The method may further comprise detecting the one or more detectable positive signals, wherein detection indicates the presence of one or more target molecules in the sample. The method may further comprise comparing the intensity of the one or more signals to a control to quantify the nucleic acid in the sample. The steps of amplifying, incubating, activating, and detecting may all be performed in the same individual discrete volume.

[0496] An “individual discrete volume” is a discrete volume or discrete space, such as a container, receptacle, or other defined volume or space that can be defined by properties that prevent and/or inhibit migration of nucleic acids and reagents necessary to carry out the methods disclosed herein, for example a volume or space defined by physical properties such as walls, for example the walls of a well, tube, or a surface of a droplet, which may be impermeable or semipermeable, or as defined by other means such as chemical, diffusion rate limited, electro-magnetic, or light illumination, or any combination thereof. By “diffusion rate limited” (for example diffusion defined volumes) is meant spaces that are only accessible to certain molecules or reactions because diffusion constraints effectively defining a space or volume as would be the case for two parallel laminar streams where diffusion will limit the migration of a target molecule from one stream to the other. By “chemical” defined volume or space is meant spaces where only certain target molecules can exist because of their chemical or molecular properties, such as size, where for example gel beads may exclude certain species from entering the beads but not others, such as by surface charge, matrix size or other physical property of the bead that can allow selection of species that may enter the interior of the bead. By “electro-magnetically” defined volume or space is meant spaces where the electro-magnetic properties of the target molecules or their supports, such as charge or magnetic properties, can

be used to define certain regions in a space such as capturing magnetic particles within a magnetic field or directly on magnets. By “optically” defined volume is meant any region of space that may be defined by illuminating it with visible, ultraviolet, infrared, or other wavelengths of light such that only target molecules within the defined space or volume may be labeled. One advantage to the use of non-walled, or semipermeable is that some reagents, such as buffers, chemical activators, or other agents may be passed in Applicants’ through the discrete volume, while other material, such as target molecules, may be maintained in the discrete volume or space. Typically, a discrete volume will include a fluid medium, (for example, an aqueous solution, an oil, a buffer, and/or a media capable of supporting cell growth) suitable for labeling of the target molecule with the indexable nucleic acid identifier under conditions that permit labeling. Exemplary discrete volumes or spaces useful in the disclosed methods include droplets (for example, microfluidic droplets and/or emulsion droplets), hydrogel beads or other polymer structures (for example poly-ethylene glycol diacrylate beads or agarose beads), tissue slides (for example, fixed formalin paraffin embedded tissue slides with particular regions, volumes, or spaces defined by chemical, optical, or physical means), microscope slides with regions defined by depositing reagents in ordered arrays or random patterns, tubes (such as, centrifuge tubes, microcentrifuge tubes, test tubes, cuvettes, conical tubes, and the like), bottles (such as glass bottles, plastic bottles, ceramic bottles, Erlenmeyer flasks, scintillation vials and the like), wells (such as wells in a plate), plates, pipettes, or pipette tips among others. In certain example embodiments, the individual discrete volumes are the wells of a microplate. In certain example embodiments, the microplate is a 96 well, a 384 well, or a 1536 well microplate.

[0497] Incubating the sample at either the amplification step or the extraction steps as described herein can be performed using heat sources known in the art. Advantageously, the heat source can be readily commercially available heating sources that do not require complicated instrumentation. Exemplary heating systems can include heating blocks, incubators, and/or water baths with temperatures maintained by commercially available sous-vide cookers. In this way, sample diagnostics can be performed without the requirement of expensive and proprietary equipment found primarily in diagnostic laboratory and hospital settings.

[0498] In certain example embodiments, paper-based microfluidics may be used for transfer of samples or reagents. For example, paper strips having wax barrier printed at a

defined distance from the end of a paper dipstick may be used to define a volume of reagent or sample to be transferred. For example, a wax barrier may be printed across a paper dipstick to define a microliter volume such that when the dipstick is transferred into a volume of a reagent or sample only a microliter of said reagent or sample is absorbed onto the dipstick. The dipstick may be placed in a second reagent mix, where the reagent or sample will diffuse into the reaction mixture. Such components allow for preparation and use of the assay without specialized equipment such as pipettors.

[0499] Optical means may be used to assess the presence and level of a given target molecule. In one embodiment, an optical sensor detects unmasking of a fluorescent masking agent. In one embodiment, the device of the present invention may include handheld portable devices for diagnostic reading of an assay (see e.g., Vashist et al., Commercial Smartphone-Based Devices and Smart Applications for Personalized Healthcare Monitoring and Management, *Diagnostics* 2014, 4(3), 104-128; mReader from Mobile Assay; and Holomic Rapid Diagnostic Test Reader).

[0500] As noted herein, certain embodiments allow detection via colorimetric change which has certain attendant benefits when embodiments are utilized in POC situations and or in resource poor environments where access to more complex detection equipment to readout the signal may be limited. However, portable embodiments disclosed herein may also be coupled with hand-held spectrophotometers that enable detection of signals outside the visible range. An example of a hand-held spectrophotometer device that may be used in combination with the present invention is described in Das *et al.* "Ultra-portable, wireless smartphone spectrophotometer for rapid, non-destructive testing of fruit ripeness." *Nature Scientific Reports*. 2016, 6:32504, DOI: 10.1038/srep32504. Finally, In one embodiment utilizing quantum dot-based masking constructs, use of a hand-held UV light, or other suitable device, may be successfully used to detect a signal owing to the near complete quantum yield provided by quantum dots.

Amplifying Target Molecules

[0501] The step of amplifying one or more target molecules can comprise amplification systems known in the art. In one embodiment, amplification is isothermal. In certain example embodiments, target RNAs and/or DNAs may be amplified prior to activating the Cas polypeptide. Any suitable RNA or DNA amplification technique may be used. In certain embodiments, the amplifying step may take less than about 1 hour, 50 minutes, 40 minutes, 30

minutes, 25 minutes, 20 minutes or 15 minutes, which may depend on the sample, starting concentrations and nature of amplification used.

[0502] In one embodiment, the amplifying of the target molecules and the detection of the target molecules can be performed in a single reaction, for example, a ‘one-pot’ method. General guidance for use of a single-pot approach can be as described in Gootenberg, et al., *Science* 2018 Apr 27: 360(6387) 439-444 (using Cas13, Cas12a and Csm6 generally, detecting multiple targets in a single reaction, and specifically performing DNA extraction in a sample and using as input for direct detection at Figure S33); and Ding et al., “All-in-One Dual CRISPR-Cas12a (AIOD-CRISPR) Assay: A Case for Rapid, Ultrasensitive and Visual Detection of Novel Coronavirus SARS-CoV-2 and HIV Virus,” doi:10.1101/2020.03.19.998724, *bioRxiv* preprint (utilizing a pair of crRNAs with dual CRISPR-Cas12a detection for a one-pot approach to target-specific nucleic acid detection).

[0503] In certain example embodiments, the RNA or DNA amplification is an isothermal amplification. In certain example embodiments, the isothermal amplification may be nucleic-acid sequenced-based amplification (NASBA), recombinase polymerase amplification (RPA), loop-mediated isothermal amplification (LAMP), strand displacement amplification (SDA), helicase-dependent amplification (HDA), or nicking enzyme amplification reaction (NEAR). In certain example embodiments, non-isothermal amplification methods may be used which include, but are not limited to, PCR, multiple displacement amplification (MDA), rolling circle amplification (RCA), ligase chain reaction (LCR), or ramification amplification method (RAM).

[0504] The amplifying of target molecules can be optimized by methods as detailed herein. In an aspect, the design optimizes the primers used in the amplification, In particular aspects, the isothermal amplification is used with CRISPR-Cas systems. In either approach, design considerations can follow a rational design for optimization of the reactions. Optimization of the methods as disclosed herein can include first screening primers to identify one or more sets of primers that work well for a particular target, Cas polypeptide and/or reaction. Once the primers have been screened, titration of magnesium concentration can be performed to identify an optimal magnesium concentration for higher signal to noise readout. In an example, varying additives with specific primers, target, Cas polypeptide, temperature, and other additive concentrations within the reaction can be identified. Optimization can be made with the goal of reducing the number of steps and buffer exchanges that have to occur in the reaction,

simplifying the reaction and reducing the risks of contamination at transfer steps. Similarly, optimizing the salt levels as well as the type of salt utilized can further facilitate and optimize the one-pot detections disclosed herein.

Loop-Mediated Isothermal Amplification

[0505] In certain example embodiments, a loop-mediated isothermal amplification (LAMP) reaction may be used to target nucleic acids, which encompasses both LAMP and RT-LAMP reactions. LAMP can be performed with a four-primer system for isothermal nucleic acid amplification in conjunction with a polymerase. Notomi et al., *Nucleic Acids Res.* 2000, 28, 12, Nagamine et al., *Molecular and Cellular Probes* (2002) 16, 223-229, doi: 10.1006/mcpr.2002.0415. When performing LAMP with a 4-primer system, two loop-forming inner primers, denoted as FIP and BIP, are provided with two outer primers, F3 and B3. The inner primers each contain two distinct sequences, one for priming in the first stage of the amplification and the other sequence for self-priming in subsequent amplification states. The two outer primers initiate strand displacement of nucleic acid strands initiated from the FIP and BIP primers, thereby generating formation of loops and strand displacement nucleic acid synthesis utilizing the provided polymerase. LAMP can be conducted with two to six primers, ranging from only the two loop-forming primers, up to at least the addition of 2 additional primers, LF and LB along with the two outer primers and two inner primers. LAMP technologies advantageously have high specificity and can work at a variety of pH and temperature. In a preferred aspect, the LAMP is an isothermal reaction at between about 45° C to 75° C, 55 to 70° C or 60° C to 65° C. Colorimetric LAMP (Y. Zhang et al., doi:10.1101/2020.92.26.20028373), RT-LAMP (Lamb et al., doi: 10.1101/2020.02.19.20025155; and Yang et al., doi:10.1101/2020.03.02.20030130) have been developed for detection of COVID-19, and are incorporated herein by reference in their entirety.

[0506] In one embodiment, the LAMP reagents may include Bst 2.0 + RTx or Bst 3.0 from New England Biolabs. In one embodiment, the LAMP reagents may comprise colorimetric or fluorescent detection. Detection of LAMP products can be accomplished using colorimetric tools, such as hydroxy naphthol blue (see, e.g. Goto, M., et al., *Colorimetric detection of loop-mediated isothermal amplification reaction by using hydroxy naphthol blue*. *Biotechniques*, 2009. 46(3): p. 167-72.), leuco triphenylmethane dyes (see, e.g. Miyamoto, S., et al., *Method for colorimetric detection of double-stranded nucleic acid using leuco triphenylmethane dyes*.

Anal Biochem, 2015. 473: p. 28-33) and pH-sensitive dyes (see, e.g. Tanner, N.A., Y. Zhang, and T.C. Evans, Jr., Visual detection of isothermal nucleic acid amplification using pH-sensitive dyes. Biotechniques, 2015. 58(2): p. 59-68); as well as fluorescent detection (see, e.g. Yu et al., *Clinical Chemistry*, hvaa102, doi:10.1093/clinchem/hvaa102 12 May 2020), including use of quenching probes (see, e.g. Shirato *et al.*, *J Virol Methods*. 2018 Aug;258:41-48. doi: 10.1016/j.jviromet.2018.05.006).

[0507] In an aspect, the primer sets for LAMP are designed to amplify one or more target sequences, generating amplicons that comprise the one or more target sequences. Optionally, the primers can comprise barcodes that can be designed as described elsewhere herein. Incubating to a temperature sufficient for LAMP amplification, e.g. 50° C-72° C, more preferably 55° C to 65° C, using a polymerase and, optionally a reverse transcriptase (in the event RT-LAMP is utilized). Preferably the enzymes utilized in the LAMP reaction are heat-stabilized. LAMP primer sites have been designed, see, e.g. Park *et al.*, “Development of Reverse Transcription Loop-Mediated Isothermal Amplification Assays Targeting SARS-CoV-2” J. of Mol. Diag. (2020). Optionally, a control template is further provided with the sample, which may differ from the target sequence but share primer binding sites. In an exemplary embodiment, visual read out of the detection results can be accomplished using commercially-available lateral flow substrate, e.g. a commercially available paper substrate.

NASBA

[0508] In certain example embodiments, the RNA or DNA amplification is NASBA, which is initiated with reverse transcription of target RNA by a sequence-specific reverse primer to create a RNA/DNA duplex. RNase H is then used to degrade the RNA template, allowing a forward primer containing a promoter, such as the T7 promoter, to bind and initiate elongation of the complementary strand, generating a double-stranded DNA product. The RNA polymerase promoter-mediated transcription of the DNA template then creates copies of the target RNA sequence. Importantly, each of the new target RNAs can be detected by the nucleic acid component molecules thus further enhancing the sensitivity of the assay. Binding of the target RNAs by the nucleic acid component molecules then leads to activation of the Cas polypeptide and the methods proceed as outlined above. The NASBA reaction has the additional advantage of being able to proceed under moderate isothermal conditions, for example at approximately 41°C, making it suitable for systems and devices deployed for early and direct detection in the field and far from clinical laboratories.

RPA

[0509] In certain other example embodiments, a recombinase polymerase amplification (RPA) reaction may be used to amplify the target nucleic acids. RPA reactions employ recombinases which are capable of pairing sequence-specific primers with homologous sequence in duplex DNA. If target DNA is present, DNA amplification is initiated and no other sample manipulation such as thermal cycling or chemical melting is required. The entire RPA amplification system is stable as a dried formulation and can be transported safely without refrigeration. RPA reactions may also be carried out at isothermal temperatures with an optimum reaction temperature of 37-42°C. The sequence specific primers are designed to amplify a sequence comprising the target nucleic acid sequence to be detected. In certain example embodiments, an RNA polymerase promoter, such as a T7 promoter, is added to one of the primers. This results in an amplified double-stranded DNA product comprising the target sequence and an RNA polymerase promoter. After, or during, the RPA reaction, an RNA polymerase is added that will produce RNA from the double-stranded DNA templates. The amplified target RNA can then in turn be detected by the CRISPR-Cas system. In this way target DNA can be detected using the embodiments disclosed herein. RPA reactions can also be used to amplify target RNA. The target RNA is first converted to cDNA using a reverse transcriptase, followed by second strand DNA synthesis, at which point the RPA reaction proceeds as outlined above.

Transposase Based Amplification

[0510] Embodiments disclosed herein provide systems and methods for isothermal amplification of target nucleic acid sequences by contacting oligonucleotides containing the target nucleic acid sequence with a transposon complex. The oligonucleotides may be single stranded or double stranded RNA, DNA, or RNA/DNA hybrid oligonucleotides. The transposon complex comprises a transposase and a transposon sequence comprising one or more RNA polymerase promoters. The transposase facilitates insertion of the one or more RNA polymerase promoters into the oligonucleotide. A RNA polymerase promoter can then transcribe the target nucleic acid sequence from the inserted one or more RNA polymerase promoters. One advantage of this system is that there is no need to heat or melt double-stranded DNA templates, since RNA polymerase polymerases require a double-stranded template. Such isothermal amplification is fast and simple, obviating the need for complicated and expensive

instrumentation for denaturation and cooling. In certain example embodiment the RNA polymerase promoter is a native or modified T7 RNA promoter.

[0511] The term “transposon”, as used herein, refers to a nucleic acid segment, which is recognized by a transposase or an integrase enzyme and which is an essential component of a functional nucleic acid-protein complex (i.e. a transposome) capable of transposition. The term “transposase” as used herein refers to an enzyme, which is a component of a functional nucleic acid-protein complex capable of transposition and which is mediating transposition. The term “transposase” also refers to integrases from retrotransposons or of retroviral origin. Transposon complexes form between a transposase enzyme and a fragment of double stranded DNA that contains a specific binding sequence for the enzyme, termed “transposon end”. The sequence of the transposon binding site can be modified with other bases, at certain positions, without affecting the ability for transposon complex to form a stable structure that can efficiently transpose into target DNA.

[0512] In embodiments provided herein, the transposon complex may comprise a transposase and a transposon sequence comprising one or more RNA polymerase promoters. The term “promoter” refers to a region of DNA involved in binding the RNA polymerase to initiate transcription. In specific embodiments, the RNA polymerase promoter may be a T7 RNA polymerase promoter. The T7 RNA promoter may be inserted into the double-stranded polynucleotide using the transposase. In one embodiment, insertion of the T7 RNA polymerase promoter into the oligonucleotide may be random.

[0513] The frequency of transposition is very low for most transposons, which use complex mechanisms to limit activity. Tn5 transposase, for example, utilizes a DNA binding sequence that is suboptimal, and the C-terminus of the transposase interferes with DNA binding. Mechanisms involved in Tn5 transposition have been carefully characterized by Reznikoff and colleagues. Tn5 transposes by a cut-and-paste mechanism. The transposon has two pairs of 19 bp elements that are utilized by the transposase: outside elements (OE) and inside elements (IE). One transposase monomer binds to each of the two elements that are utilized. After a monomer is bound to each end of the transposon, the two monomers dimerize, forming a synapse. Vectors with donor backbones of at least 200 bp, but less than 1000 bp, are most functional for transposition in bacteria. Transposon cleavage occurs by trans catalysis and only when monomers bound to each DNA end are in a synaptic complex. Tn5 transposes with a

relaxed target site selection and can therefore insert into target DNA with little to no target sequence specificity.

[0514] The natural downregulation of Tn5 transposition can be overcome by selection of a hyperactive transposase and by optimizing the transposase-binding elements [York et al. 1998]. A mosaic element (ME), made by modification of three bases of the wild type OE, led to a 50-fold increase in transposition events in bacteria as well as cell-free systems. The combined effect of the optimized ME and hyperactive mutant transposase is estimated to result in a 100-fold increase in transposition activity. Goryshin et al showed that preformed Tn5 transposition complexes could be functionally introduced into bacterial or yeast by electroporation [Goryshin et al. 2000]. Linearization of the DNA, to have inverted repeats precisely positioned at both ends of the transposon, allowed Goryshin and coworkers to bypass the cutting step of transposition thus enhancing transposition efficiency.

[0515] In one embodiment, the transposase may be used to tagment the oligonucleotide sequence comprising the target sequence. The term “tagmentation” refers to a step in the Assay for Transposase Accessible Chromatin using sequencing (ATAC-seq) as described. (See, Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., Greenleaf, W. J., Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* 2013; 10 (12): 1213-1218). Specifically, a hyperactive Tn5 transposase loaded in vitro with adapters for high-throughput DNA sequencing, can simultaneously fragment and tag a genome with sequencing adapters. In one embodiment the adapters are compatible with the methods described herein.

[0516] In one embodiment, the transposase may be a Tn5 transposase. In one embodiment, the transposase may be a variant of a Tn5 transposase, or an engineered transposase. Transposases may be engineered using any method known in the art. The engineered transposase may be optimized to function at a temperature ranging from 30°C to 45°C, 35°C to 40°C or any temperature in between. The engineered transposase may be optimized to release from the oligonucleotide at a faster rate compared to a wild type transposase.

[0517] In one embodiment, the transposase may be a Tn5 transposase, a Mu transposase, or a Tn7 transposase. Transposition efficiency in vitro may vary depending on the transposon system used. Generally, Tn5 and Mu transposases effect higher levels of transposition efficiency. In one embodiment, insertion may be random. In one embodiment, insertion may occur in GC rich regions of the target sequence.

[0518] In one embodiment, the transposon sequence may comprise two 19 base pair Mosaic End (ME) Tn5 transposase recognition sequences. Tn5 transposases will generally transpose any DNA sequence contained between such short 19 base pair ME Tn5 transposase recognition sequences.

[0519] In one embodiment, use of a transposase allows for separation of a double-stranded polynucleotide in the absence of heat or melting. Approaches can be adapted from those described in PCT/US2019/039195, incorporated herein by reference.

Nickase Dependent Amplification

[0520] In an embodiment of the invention may comprise nickase-based amplification. The nicking enzyme may be a Cas polypeptide. Accordingly, the introduction of nicks into dsDNA can be programmable and sequence-specific. In an embodiment of the invention, two guides can be designed to target opposite strands of a dsDNA target. According to the invention, the nickase can be Cas polypeptide, or one may use any Cas protein such as Cpf1, C2c1, Cas9, or any ortholog or CRISPR protein that cleaves or is engineered to cleave a single strand of a DNA duplex. In a particular embodiment, the Cas polypeptide is utilized in the nickase dependent amplification. The nicked strands may then be extended by a polymerase. In an embodiment, the locations of the nicks are selected such that extension of the strands by a polymerase is towards the central portion of the target duplex DNA between the nick sites. In one embodiment, primers are included in the reaction capable of hybridizing to the extended strands followed by further polymerase extension of the primers to regenerate two dsDNA pieces: a first dsDNA that includes the first strand Cas polypeptide guide site or both the first and second strand Cas polypeptide guide sites, and a second dsDNA that includes the second strand Cas polypeptide guide site or both the first and second strand Cas polypeptide guide sites. These pieces continue to be nicked and extended in a cyclic reaction that exponentially amplifies the region of the target between nicking sites. Alternatively, a CRISPR-Cas protein instead of Cas polypeptide can be used for nickase-based amplification, and such methods are known in the art.

[0521] The amplification can be isothermal and selected for temperature. In one embodiment, the amplification proceeds rapidly at 37 degrees. In other embodiments, the temperature of the isothermal amplification may be chosen by selecting a polymerase (e.g. Bsu, Bst, Phi29, klenow fragment etc.) operable at a different temperature.

[0522] Thus, whereas nicking isothermal amplification techniques use nicking enzymes with fixed sequence preference (e.g. in nicking enzyme amplification reaction or NEAR), which requires denaturing of the original dsDNA target to allow annealing and extension of primers that add the nicking substrate to the ends of the target, use of a reprogrammable nickase wherein the nicking sites can be programmed via RNA molecules means that no denaturing step is necessary, enabling the entire reaction to be truly isothermal. This also simplifies the reaction because these primers that add the nicking substrate are different than the primers that are used later in the reaction, meaning that NEAR requires two primer sets (i.e. 4 primers) while CpfI nicking amplification only requires one primer set (i.e. two primers). This makes nicking CpfI amplification much simpler and easier to operate without complicated instrumentation to perform the denaturation and then cooling to the isothermal temperature.

[0523] In an aspect, the isothermal amplification reagents may be utilized with a thermostable Cas polypeptide. The combination of thermostable protein and isothermal amplification reagents may be utilized to further improve reaction times for detection and diagnostics.

[0524] Accordingly, in certain example embodiments the systems disclosed herein may include amplification reagents. Different components or reagents useful for amplification of nucleic acids are described herein. For example, an amplification reagent as described herein may include a buffer, such as a Tris buffer. A Tris buffer may be used at any concentration appropriate for the desired application or use, for example including, but not limited to, a concentration of 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 11 mM, 12 mM, 13 mM, 14 mM, 15 mM, 25 mM, 50 mM, 75 mM, 1 M, or the like. One of skill in the art will be able to determine an appropriate concentration of a buffer such as Tris for use with the present invention.

[0525] A salt, such as magnesium chloride (MgCl₂), potassium chloride (KCl), or sodium chloride (NaCl), may be included in an amplification reaction, such as PCR, in order to improve the amplification of nucleic acid fragments. Although the salt concentration will depend on the particular reaction and application, In one embodiment, nucleic acid fragments of a particular size may produce optimum results at particular salt concentrations. Larger products may require altered salt concentrations, typically lower salt, in order to produce desired results, while amplification of smaller products may produce better results at higher salt concentrations. One of skill in the art will understand that the presence and/or concentration of a salt, along

with alteration of salt concentrations, may alter the stringency of a biological or chemical reaction, and therefore any salt may be used that provides the appropriate conditions for a reaction of the present invention and as described herein. In certain preferred embodiments, when polynucleotide extraction beads such as magnetic beads are utilized, a Plant QuickExtract solution can be used in combination with a KCl buffer in optimized detection methods according to the present disclosure.

[0526] Other components of a biological or chemical reaction may include a cell lysis component in order to break open or lyse a cell for analysis of the materials therein. A cell lysis component may include, but is not limited to, a detergent, a salt as described above, such as NaCl, KCl, ammonium sulfate $[(\text{NH}_4)_2\text{SO}_4]$, or others. Detergents that may be appropriate for the invention may include Triton X-100, sodium dodecyl sulfate (SDS), CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), ethyl trimethyl ammonium bromide, nonyl phenoxy polyethoxy ethanol (NP-40). Concentrations of detergents may depend on the particular application and may be specific to the reaction in some cases. Amplification reactions may include dNTPs and nucleic acid primers used at any concentration appropriate for the invention, such as including, but not limited to, a concentration of 100 nM, 150 nM, 200 nM, 250 nM, 300 nM, 350 nM, 400 nM, 450 nM, 500 nM, 550 nM, 600 nM, 650 nM, 700 nM, 750 nM, 800 nM, 850 nM, 900 nM, 950 nM, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM, 10 mM, 20 mM, 30 mM, 40 mM, 50 mM, 60 mM, 70 mM, 80 mM, 90 mM, 100 mM, 150 mM, 200 mM, 250 mM, 300 mM, 350 mM, 400 mM, 450 mM, 500 mM, or the like. Likewise, a polymerase useful in accordance with the invention may be any specific or general polymerase known in the art and useful for the invention, including Taq polymerase, Q5 polymerase, or the like.

[0527] In one embodiment, amplification reagents as described herein may be appropriate for use in hot-start amplification. Hot start amplification may be beneficial in one embodiment to reduce or eliminate dimerization of adaptor molecules or oligos, or to otherwise prevent unwanted amplification products or artifacts and obtain optimum amplification of the desired product. Many components described herein for use in amplification may also be used in hot-start amplification. In one embodiment, reagents or components appropriate for use with hot-start amplification may be used in place of one or more of the composition components as appropriate. For example, a polymerase or other reagent may be used that exhibits a desired activity at a particular temperature or other reaction condition. In one embodiment, reagents

may be used that are designed or optimized for use in hot-start amplification, for example, a polymerase may be activated after transposition or after reaching a particular temperature. Such polymerases may be antibody-based or aptamer-based. Polymerases as described herein are known in the art. Examples of such reagents may include, but are not limited to, hot-start polymerases, hot-start dNTPs, and photo-caged dNTPs. Such reagents are known and available in the art. One of skill in the art will be able to determine the optimum temperatures as appropriate for individual reagents.

[0528] Amplification of nucleic acids may be performed using specific thermal cycle machinery or equipment and may be performed in single reactions or in bulk, such that any desired number of reactions may be performed simultaneously. In one embodiment, amplification may be performed using microfluidic or robotic devices, or may be performed using manual alteration in temperatures to achieve the desired amplification. In one embodiment, optimization may be performed to obtain the optimum reactions conditions for the particular application or materials. One of skill in the art will understand and be able to optimize reaction conditions to obtain sufficient amplification.

[0529] In one embodiment, detection of DNA with the methods or systems of the invention requires transcription of the (amplified) DNA into RNA prior to detection.

[0530] It will be evident that detection methods of the invention can involve nucleic acid amplification and detection procedures in various combinations. The nucleic acid to be detected can be any naturally occurring or synthetic nucleic acid, including but not limited to DNA and RNA, which may be amplified by any suitable method to provide an intermediate product that can be detected. Detection of the intermediate product can be by any suitable method including but not limited to binding and activation of a Cas polypeptide which produces a detectable signal moiety by direct or collateral activity.

Helicase-Dependent Amplification

[0531] In helicase-dependent amplification, a helicase enzyme is used to unwind a double stranded nucleic acid to generate templates for primer hybridization and subsequent primer-extension. This process utilizes two oligonucleotide primers, each hybridizing to the 3'-end of either the sense strand containing the target sequence or the anti-sense strand containing the reverse-complementary target sequence. The HDA reaction is a general method for helicase-dependent nucleic acid amplification.

[0532] In combining this method with a CRISPR-Cas detection system, the target nucleic acid may be amplified by opening R-loops of the target nucleic acid using first and second CRISPR-Cases. The first and second strand of the target nucleic acid may thus be unwound using a helicase, allowing primers and polymerase to bind and extend the DNA under isothermal conditions.

[0533] The term “helicase” refers here to any enzyme capable of unwinding a double stranded nucleic acid enzymatically. For example, helicases are enzymes that are found in all organisms and in all processes that involve nucleic acid such as replication, recombination, repair, transcription, translation, and RNA splicing. (Kornberg and Baker, DNA Replication, W. H. Freeman and Company (2nd ed. (1992)), especially chapter 11). Any helicase that translocates along DNA or RNA in a 5' to 3' direction or in the opposite 3' to 5' direction may be used in present embodiments of the invention. This includes helicases obtained from prokaryotes, viruses, archaea, and eukaryotes or recombinant forms of naturally occurring enzymes as well as analogues or derivatives having the specified activity. Examples of naturally occurring DNA helicases, described by Kornberg and Baker in chapter 11 of their book, DNA Replication, W. H. Freeman and Company (2nd ed. (1992)), include *E. coli* helicase I, II, III, & IV, Rep, DnaB, PriA, PcrA, T4 Gp41helicase, T4 Dda helicase, T7 Gp4 helicases, SV40 Large T antigen, yeast RAD. Additional helicases that may be useful in HDA include RecQ helicase (Harmon and Kowalczykowski, *J. Biol. Chem.* 276:232–243 (2001)), thermostable UvrD helicases from *T. tengcongensis* (disclosed in this invention, Example XII) and *T. thermophilus* (Collins and McCarthy, *Extremophiles.* 7:35–41. (2003)), thermostable DnaB helicase from *T. aquaticus* (Kaplan and Steitz, *J. Biol. Chem.* 274:6889–6897 (1999)), and MCM helicase from archaeal and eukaryotic organisms ((Grainge et al., *Nucleic Acids Res.* 31:4888–4898 (2003)).

[0534] A traditional definition of a helicase is an enzyme that catalyzes the reaction of separating/unzipping/unwinding the helical structure of nucleic acid duplexes (DNA, RNA, or hybrids) into single-stranded components, using nucleoside triphosphate (NTP) hydrolysis as the energy source (such as ATP). However, it should be noted that not all helicases fit this definition anymore. A more general definition is that they are motor proteins that move along the single-stranded or double stranded nucleic acids (usually in a certain direction, 3' to 5' or 5' to 3', or both), i.e. translocases, that can or cannot unwind the duplexed nucleic acid

encountered. In addition, some helicases simply bind and “melt” the duplexed nucleic acid structure without an apparent translocase activity.

[0535] Helicases exist in all living organisms and function in all aspects of nucleic acid metabolism. Helicases are classified based on the amino acid sequences, directionality, oligomerization state and nucleic-acid type and structure preferences. The most common classification method was developed based on the presence of certain amino acid sequences, called motifs. According to this classification helicases are divided into 6 super families: SF1, SF2, SF3, SF4, SF5 and SF6. SF1 and SF2 helicases do not form a ring structure around the nucleic acid, whereas SF3 to SF6 do. Superfamily classification is not dependent on the classical taxonomy.

[0536] DNA helicases are responsible for catalyzing the unwinding of double-stranded DNA (dsDNA) molecules to their respective single-stranded nucleic acid (ssDNA) forms. Although structural and biochemical studies have shown how various helicases can translocate on ssDNA directionally, consuming one ATP per nucleotide, the mechanism of nucleic acid unwinding and how the unwinding activity is regulated remains unclear and controversial (T. M. Lohman, E. J. Tomko, C. G. Wu, “Non-hexameric DNA helicases and translocases: mechanisms and regulation,” *Nat Rev Mol Cell Biol* 9:391-401 (2008)). Since helicases can potentially unwind all nucleic acids encountered, understanding how their unwinding activities are regulated can lead to harnessing helicase functions for biotechnology applications.

[0537] The term “HDA” refers to Helicase Dependent Amplification, which is an *in vitro* method for amplifying nucleic acids by using a helicase preparation for unwinding a double stranded nucleic acid to generate templates for primer hybridization and subsequent primer-extension. This process utilizes two oligonucleotide primers, each hybridizing to the 3'-end of either the sense strand containing the target sequence or the anti-sense strand containing the reverse-complementary target sequence. The HDA reaction is a general method for helicase-dependent nucleic acid amplification.

[0538] The invention comprises use of any suitable helicase known in the art. These include, but are not necessarily limited to, UvrD helicase, CRISPR-Cas3 helicase, *E. coli* helicase I, *E. coli* helicase II, *E. coli* helicase III, *E. coli* helicase IV, Rep helicase, DnaB helicase, PriA helicase, PcrA helicase, T4 Gp41 helicase, T4 Dda helicase, SV40 Large T antigen, yeast RAD helicase, RecD helicase, RecQ helicase, thermostable *T. tengcongensis* UvrD helicase, thermostable *T. thermophilus* UvrD helicase, thermostable *T. aquaticus* DnaB

helicase, Dda helicase, papilloma virus E1 helicase, archaeal MCM helicase, eukaryotic MCM helicase, and T7 Gp4 helicase.

[0539] In particularly preferred embodiments, the helicase comprises a super mutation. In particular embodiments, although the *E. coli* mutation has been described, the mutations were generated by sequence alignment (e.g. D409A/D410A for TteUvr_d) and result in thermophilic enzymes working at lower temperatures like 37°C, which is advantageous for amplification methods and systems described herein. In one embodiment, the super mutations are an aspartate to alanine mutation, with position based on sequence alignment. In one embodiment, the super mutant helicase is selected from WP_003870487.1 *Thermoanaerobacter ethanolicus* 403/404, WP_049660019.1 *Bacillus* sp. FJAT-27231 407/408, WP_034654680.1 *Bacillus megaterium* 415/416, WP_095390358.1 *Bacillus simplex* 407/408, and WP_055343022.1 *Paenibacillus sordellii* 402/403.

Incubating

[0540] Methods of detection and/or extraction using the systems disclosed herein can comprise incubating the sample or set of samples under conditions sufficient to allow binding of the nucleic acid component molecules to one or more target molecules. Extraction can comprise incubating the sample under conditions sufficient to allow release of viral RNA present in the sample, which may comprise incubating at 22°C to 60 °C for 30 to 70 minutes or at 90°C -100°C for about 10 minutes.

[0541] In certain example embodiments, the incubation time of the amplifying and detecting in the present invention may be shortened. The assay may be performed in a period of time required for an enzymatic reaction to occur. One skilled in the art can perform biochemical reactions in 5 minutes (e.g., 5 minute ligation). Incubating may occur at one or more temperatures over timeframes between about 10 minutes and 90 minutes, preferably less than 90 minutes, 75 minutes, 60 minutes, 45 minutes, 30 minutes, 25 minutes, 20 minutes, 15 minutes, or 10 minutes depending on sample, reagents and components of the system. In one embodiment, incubating for the amplification is performed at one or more temperatures between about 20° C and 80° C, In one embodiment, about 37° C. In one embodiment, incubating for the amplification is performed at one or more temperatures between about 55° C and 65° C, between about 59° C and 61° C, In one embodiment, about 60° C.

Activating

[0542] In certain example embodiment, activating of the Cas polypeptide occurs via binding of the CRISPR-Cas via the nucleic acid component molecule to the one or more target molecules, wherein activating the Cas polypeptide results in modification of the detection construct such that a detectable signal is generated.

Detecting a Signal

[0543] Detecting may comprise visual observance of a positive signal relative to a control. Detecting may comprise a loss of signal or presence of signal at one or more capture regions, for example colorimetric detection, or fluorescent detection. In certain example embodiments, further modifications may be introduced that further amplify the detectable positive signal. For example, activated Cas polypeptide collateral activation may be used to generate a secondary target or additional nucleic acid component molecule sequence, or both. In one example embodiment, the reaction solution would contain a secondary target that is spiked in at high concentration. The secondary target may be distinct from the primary target (i.e. the target for which the assay is designed to detect) and in certain instances may be common across all reaction volumes. A secondary nucleic acid component molecule sequence for the secondary target may be protected, e.g. by a secondary structural feature such as a hairpin with an RNA loop, and unable to bind the second target or the Cas polypeptide. Cleavage of the protecting group by an activated Cas r protein (i.e. after activation by formation of complex with the primary target(s) in solution) and formation of a complex with free Cas polypeptide in solution and activation from the spiked in secondary target. In certain other example embodiments, a similar concept is used with free nucleic acid component molecule sequence to a secondary target and protected secondary target. Cleavage of a protecting group off the secondary target would allow additional Cas polypeptide, nucleic acid component sequence, secondary target sequence to form. In yet another example embodiment, activation of Cas polypeptide by the primary target(s) may be used to cleave a protected or circularized primer, which would then be released to perform an isothermal amplification reaction, such as those disclosed herein, on a template for either secondary nucleic acid component sequence, secondary target, or both. Subsequent transcription of this amplified template would produce more secondary nucleic acid component molecule sequence and/or secondary target sequence, followed by additional Cas polypeptide collateral activation.

Quantifying

[0544] In particular methods, comparing the intensity of the one or more signals to a control is performed to quantify the nucleic acid in the sample. The term “control” refers to any reference standard suitable to provide a comparison to the expression products in the test sample. In one embodiment, the control comprises obtaining a “control sample” from which expression product levels are detected and compared to the expression product levels from the test sample. Such a control sample may comprise any suitable sample, including but not limited to a sample from a control patient (can be stored sample or previous sample measurement) with a known outcome; normal tissue, fluid, or cells isolated from a subject, such as a normal patient or the patient having a condition of interest.

[0545] The intensity of a signal is “significantly” higher or lower than the normal intensity if the signal is greater or less, respectively, than the normal or control level by an amount greater than the standard error of the assay employed to assess amount, and preferably at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 150%, 200%, 300%, 350%, 400%, 500%, 600%, 700%, 800%, 900%, 1000% or than that amount. Alternatively, the signal can be considered “significantly” higher or lower than the normal and/or control signal if the amount is at least about two, and preferably at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 100%, 105%, 110%, 115%, 120%, 125%, 130%, 135%, 140%, 145%, 150%, 155%, 160%, 165%, 170%, 175%, 180%, 185%, 190%, 195%, two times, three times, four times, five times, or more, or any range in between, such as 5%-100%, higher or lower, respectively, than the normal and/or control signal. Such significant modulation values can be applied to any metric described herein, such as altered level of expression, altered activity, changes in biomarker inhibition, changes in test agent binding, and the like.

[0546] In one embodiment, the detectable positive signal may be a loss of fluorescent signal or colorimetric relative to a control, as described herein. In one embodiment, the detectable positive signal may be detected on a lateral flow device, as described herein.

Applications of Detection Methods

[0547] Systems and methods can be designed for the detection and diagnosis of microbes, including bacterial, fungi and viral microbes. In an aspect, the systems may comprise multiplex detection of multiple variants of viral infections, including coronavirus, different viruses which may be related coronaviruses or respiratory viruses, or a combination thereof. In embodiments,

assays can be performed for a variety of viruses and viral infections, including acute respiratory infections using the disclosure detailed herein. The systems can comprise two or more CRISPR-Cas systems to multiplex, as described elsewhere herein, to detect a plurality of respiratory infections or viral infections, including coronavirus. The coronavirus is a positive-sense single stranded RNA family of viruses, infecting a variety of animals and humans. SARS-CoV is one type of coronavirus infection, as well as MERS-CoV. Detection of one or more coronaviruses are envisioned, including the 2019-nCoV detected in Wuhan City. Sequences of the 2019-nCoV are available at GISAID accession no. EPI_ISL_402124 and EPI_ISL_402127-402130, and described in DOI: 10.1101/2020.01.22.914952. Further deposits of the SARS-CoV-2 deposited in the GISAID platform include EP_ISL_402119-402121 and EP_ISL 402123-402124; see also GenBank Accession No. MN908947.3.

[0548] Target molecule detection can comprise two or more detection systems utilizing Cas polypeptides. The Cas polypeptide may preferably be thermostable, with multiplexing designed such that different Cas polypeptides with different sequence specificities, operable temperatures, or cutting preferences can be used.

[0549] A multiplex embodiment can be designed to track one or more variants of coronavirus or one or more variants of coronavirus, including SARS-CoV-2, in combination with other viruses, for example, Human respiratory syncytial virus, Middle East respiratory syndrome (MERS) coronavirus, Severe acute respiratory syndrome-related (SARS) coronavirus, and influenza. In embodiments, assays can be done in multiplex to detect multiple variants of coronavirus, different viruses which may be related coronaviruses or respiratory viruses, or a combination thereof. In an aspect, each assay can take place in an individual discrete volume. An “individual discrete volume” is a discrete volume or discrete space, such as a container, receptacle, or other defined volume or space that can be defined by properties that prevent and/or inhibit migration of nucleic acids and reagents necessary to carry out the methods disclosed herein, for example a volume or space defined by physical properties such as walls, for example the walls of a well, tube, or a surface of a droplet, which may be impermeable or semipermeable, or as defined by other means such as chemical, diffusion rate limited, electro-magnetic, or light illumination, or any combination thereof. By “diffusion rate limited” (for example diffusion defined volumes) is meant spaces that are only accessible to certain molecules or reactions because diffusion constraints effectively defining a space or volume as would be the case for two parallel laminar streams where diffusion will limit the

migration of a target molecule from one stream to the other. By “chemical” defined volume or space is meant spaces where only certain target molecules can exist because of their chemical or molecular properties, such as size, where for example gel beads may exclude certain species from entering the beads but not others, such as by surface charge, matrix size or other physical property of the bead that can allow selection of species that may enter the interior of the bead. By “electro-magnetically” defined volume or space is meant spaces where the electro-magnetic properties of the target molecules or their supports such as charge or magnetic properties can be used to define certain regions in a space such as capturing magnetic particles within a magnetic field or directly on magnets. By “optically” defined volume is meant any region of space that may be defined by illuminating it with visible, ultraviolet, infrared, or other wavelengths of light such that only target molecules within the defined space or volume may be labeled. One advantage to the use of non-walled, or semipermeable is that some reagents, such as buffers, chemical activators, or other agents may be passed in Applicants’ through the discrete volume, while other material, such as target molecules, may be maintained in the discrete volume or space. Typically, a discrete volume will include a fluid medium, (for example, an aqueous solution, an oil, a buffer, and/or a media capable of supporting cell growth) suitable for labeling of the target molecule with the indexable nucleic acid identifier under conditions that permit labeling. Exemplary discrete volumes or spaces useful in the disclosed methods include droplets (for example, microfluidic droplets and/or emulsion droplets), hydrogel beads or other polymer structures (for example poly-ethylene glycol diacrylate beads or agarose beads), tissue slides (for example, fixed formalin paraffin embedded tissue slides with particular regions, volumes, or spaces defined by chemical, optical, or physical means), microscope slides with regions defined by depositing reagents in ordered arrays or random patterns, tubes (such as, centrifuge tubes, microcentrifuge tubes, test tubes, cuvettes, conical tubes, and the like), bottles (such as glass bottles, plastic bottles, ceramic bottles, Erlenmeyer flasks, scintillation vials and the like), wells (such as wells in a plate), plates, pipettes, or pipette tips among others. In certain example embodiments, the individual discrete volumes are the wells of a microplate. In certain example embodiments, the microplate is a 96 well, a 384 well, or a 1536 well microplate.

[0550] In certain example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting the presence of one or more microbial agents in a sample, such as a biological sample obtained from a subject. In certain example embodiments, the microbe

may be a bacterium, a fungus, a yeast, a protozoan, a parasite, or a virus. Accordingly, the methods disclosed herein can be adapted for use in other methods (or in combination) with other methods that require quick identification of microbe species, monitoring the presence of microbial proteins (antigens), antibodies, antibody genes, detection of certain phenotypes (e.g. bacterial resistance), monitoring of disease progression and/or outbreak, and antibiotic screening. Because of the rapid and sensitive diagnostic capabilities of the embodiments disclosed here, detection of microbe species type, down to a single nucleotide difference, and the ability to be deployed as a POC device, the embodiments disclosed herein may be used as guide therapeutic regimens, such as a selection of the appropriate antibiotic or antiviral. The embodiments disclosed herein may also be used to screen environmental samples (air, water, surfaces, food etc.) for the presence of microbial contamination.

[0551] Disclosed is a method to identify microbial species, such as bacterial, viral, fungal, yeast, or parasitic species, or the like. Particular embodiments disclosed herein describe methods and systems that will identify and distinguish microbial species within a single sample, or across multiple samples, allowing for recognition of many different microbes. The present methods allow the detection of pathogens and distinguishing between two or more species of one or more organisms, e.g., bacteria, viruses, yeast, protozoa, and fungi or a combination thereof, in a biological or environmental sample, by detecting the presence of a target nucleic acid sequence in the sample. A positive signal obtained from the sample indicates the presence of the microbe. Multiple microbes can be identified simultaneously using the methods and systems of the invention, by employing the use of more than one effector protein, wherein each effector protein targets a specific microbial target sequence. In this way, a multi-level analysis can be performed for a particular subject in which any number of microbes can be detected at once, for example, a subject with unknown respiratory infection, having symptoms of coronavirus, or an individual at risk or having been exposed to coronavirus. In one embodiment, simultaneous detection of multiple microbes may be performed using a set of probes that can identify one or more microbial species.

Microbe Detection

[0552] In one embodiment, a method for detecting microbes in samples is provided comprising distributing a sample or set of samples into one or more individual discrete volumes, the individual discrete volumes comprising a CRISPR-Cas system as described herein; incubating the sample or set of samples under conditions sufficient to allow binding of

the one or more nucleic acid component molecules to one or more microbe-specific targets; activating the Cas polypeptide via binding of the one or more nucleic acid component molecules to the one or more target molecules, wherein activating the Cas polypeptide results in modification of the RNA-based masking construct such that a detectable positive signal is generated; and detecting the detectable positive signal, wherein detection of the detectable positive signal indicates a presence of one or more target molecules in the sample. The one or more target molecules may be mRNA, gDNA (coding or non-coding), tRNA, or rRNA comprising a target nucleotide sequence that may be used to distinguish two or more microbial species/strains from one another. The nucleic acid component molecules may be designed to detect target sequences. The embodiments disclosed herein may also utilize certain steps to improve hybridization between nucleic acid component molecule and target RNA sequences. Methods for enhancing ribonucleic acid hybridization are disclosed in WO 2015/085194, entitled "Enhanced Methods of Ribonucleic Acid Hybridization" which is incorporated herein by reference. The microbe-specific target may be RNA or DNA or a protein. If DNA method may further comprise the use of DNA primers that introduce an RNA polymerase promoter as described herein. If the target is a protein, then the method will utilize aptamers and steps specific to protein detection described herein.

Detection of Single Nucleotide Variants

[0553] In one embodiment, one or more identified target sequences may be detected using nucleic acid component molecules that are specific for and bind to the target sequence as described herein. The systems and methods of the present invention can distinguish even between single nucleotide polymorphisms present among different microbial species and therefore, use of multiple nucleic acid component molecules in accordance with the invention may further expand on or improve the number of target sequences that may be used to distinguish between species. For example, In one embodiment, the one or more nucleic acid component molecules may distinguish between microbes at the species, genus, family, order, class, phylum, kingdom, or phenotype, or a combination thereof.

Detection Based on rRNA Sequences

[0554] In certain example embodiments, the devices, systems, and methods disclosed herein may be used to distinguish multiple microbial species in a sample. In certain example embodiments, identification may be based on ribosomal RNA sequences, including the 16S, 23S, and 5S subunits. Methods for identifying relevant rRNA sequences are disclosed in U.S.

Patent Application Publication No. 2017/0029872. In certain example embodiments, a set of nucleic acid component molecule may be designed to distinguish each species by a variable region that is unique to each species or strain. nucleic acid component molecules may also be designed to target RNA genes that distinguish microbes at the genus, family, order, class, phylum, kingdom levels, or a combination thereof. In certain example embodiments where amplification is used, a set of amplification primers may be designed to flank constant regions of the ribosomal RNA sequence and a nucleic acid component molecule designed to distinguish each species by a variable internal region. In certain example embodiments, the primers and nucleic acid component molecules may be designed to conserved and variable regions in the 16S subunit respectfully. Other genes or genomic regions that uniquely variable across species or a subset of species such as the RecA gene family, RNA polymerase β subunit, may be used as well. Other suitable phylogenetic markers, and methods for identifying the same, are discussed for example in Wu et al. arXiv:1307.8690 [q-bio.GN].

[0555] In certain example embodiments, a method or diagnostic is designed to screen microbes across multiple phylogenetic and/or phenotypic levels at the same time. For example, the method or diagnostic may comprise the use of multiple CRISPR-Cas systems with different nucleic acid component molecules. A first set of nucleic acid component molecules may distinguish, for example, between mycobacteria, gram positive, and gram-negative bacteria. These general classes can be even further subdivided. For example, nucleic acid components could be designed and used in the method or diagnostic that distinguish enteric and non-enteric within gram negative bacteria. A second set of nucleic acid component molecules can be designed to distinguish microbes at the genus or species level. Thus a matrix may be produced identifying all mycobacteria, gram positive, gram negative (further divided into enteric and non-enteric) with each genus of species of bacteria identified in a given sample that fall within one of those classes. The foregoing is for example purposes only. Other means for classifying other microbe types are also contemplated and would follow the general structure described above.

Screening for Drug Resistance

[0556] In certain example embodiments, the devices, systems, and methods disclosed herein may be used to screen for microbial genes of interest, for example antibiotic and/or antiviral resistance genes. nucleic acid component molecules may be designed to distinguish between known genes of interest. Samples, including clinical samples, may then be screened

using the embodiments disclosed herein for detection of such genes. The ability to screen for drug resistance at POC would have tremendous benefit in selecting an appropriate treatment regime. In certain example embodiments, the antibiotic resistance genes are carbapenemases including KPC, NDM1, CTX-M15, OXA-48. Other antibiotic resistance genes are known and may be found for example in the Comprehensive Antibiotic Resistance Database (Jia *et al.* “CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database.” *Nucleic Acids Research*, 45, D566-573).

[0557] Ribavirin is an effective antiviral that hits a number of RNA viruses. Several clinically important viruses have evolved ribavirin resistance including Foot and Mouth Disease Virus doi:10.1128/JVI.03594-13; polio virus (Pfeifer and Kirkegaard. *PNAS*, 100(12):7289-7294, 2003); and hepatitis C virus (Pfeiffer and Kirkegaard, *J. Virol.* 79(4):2346-2355, 2005). A number of other persistent RNA viruses, such as hepatitis and HIV, have evolved resistance to existing antiviral drugs: hepatitis B virus (lamivudine, tenofovir, entecavir) doi:10.1002/hep22900; hepatitis C virus (telaprevir, BILN2061, ITMN-191, SCH6, boceprevir, AG-021541, ACH-806) doi:10.1002/hep.22549; and HIV (many drug resistance mutations) hivb.stanford.edu. The embodiments disclosed herein may be used to detect such variants among others.

[0558] Aside from drug resistance, there are a number of clinically relevant mutations that could be detected with the embodiments disclosed herein, such as persistent versus acute infection in LCMV (doi:10.1073/pnas.1019304108), and increased infectivity of Ebola (Diehl *et al.* *Cell*. 2016, 167(4):1088-1098.

[0559] As described herein elsewhere, closely related microbial species (e.g. having only a single nucleotide difference in a given target sequence) may be distinguished by introduction of a synthetic mismatch in the nucleic acid component molecule.

Monitoring Microbe Outbreaks

[0560] In one embodiment, a CRISPR-Cas system or methods of use thereof as described herein may be used to determine the evolution of a pathogen outbreak. The method may comprise detecting one or more target sequences from a plurality of samples from one or more subjects, wherein the target sequence is a sequence from a microbe causing the outbreaks. Such a method may further comprise determining a pattern of pathogen transmission, or a mechanism involved in a disease outbreak caused by a pathogen.

[0561] The pattern of pathogen transmission may comprise continued new transmissions from the natural reservoir of the pathogen or subject-to-subject transmissions (*e.g.* human-to-human transmission) following a single transmission from the natural reservoir or a mixture of both. In one embodiment, the pathogen transmission may be bacterial or viral transmission, in such case, the target sequence is preferably a microbial genome or fragments thereof. In one embodiment, the pattern of the pathogen transmission is the early pattern of the pathogen transmission, *i.e.* at the beginning of the pathogen outbreak. Determining the pattern of the pathogen transmission at the beginning of the outbreak increases likelihood of stopping the outbreak at the earliest possible time thereby reducing the possibility of local and international dissemination.

[0562] Determining the pattern of the pathogen transmission may comprise detecting a pathogen sequence according to the methods described herein. Determining the pattern of the pathogen transmission may further comprise detecting shared intra-host variations of the pathogen sequence between the subjects and determining whether the shared intra-host variations show temporal patterns. Patterns in observed intra-host and inter-host variation provide important insight about transmission and epidemiology (Gire, et al., 2014).

[0563] Detection of shared intra-host variations between the subjects that show temporal patterns is an indication of transmission links between subject (in particular between humans) because it can be explained by subject infection from multiple sources (superinfection), sample contamination recurring mutations (with or without balancing selection to reinforce mutations), or co-transmission of slightly divergent viruses that arose by mutation earlier in the transmission chain (Park, et al., Cell 161(7):1516–1526, 2015). Detection of shared intra-host variations between subjects may comprise detection of intra-host variants located at common single nucleotide polymorphism (SNP) positions. Positive detection of intra-host variants located at common (SNP) positions is indicative of superinfection and contamination as primary explanations for the intra-host variants. Superinfection and contamination can be parted on the basis of SNP frequency appearing as inter-host variants (Park, et al., 2015). Otherwise superinfection and contamination can be ruled out. In this latter case, detection of shared intra-host variations between subjects may further comprise assessing the frequencies of synonymous and nonsynonymous variants and comparing the frequency of synonymous and nonsynonymous variants to one another. A nonsynonymous mutation is a mutation that alters the amino acid of the protein, likely resulting in a biological change in the microbe that is

subject to natural selection. Synonymous substitution does not alter an amino acid sequence. Equal frequency of synonymous and nonsynonymous variants is indicative of the intra-host variants evolving neutrally. If frequencies of synonymous and nonsynonymous variants are divergent, the intra-host variants are likely to be maintained by balancing selection. If frequencies of synonymous and nonsynonymous variants are low, this is indicative of recurrent mutation. If frequencies of synonymous and nonsynonymous variants are high, this is indicative of co-transmission (Park, et al., 2015).

[0564] Like Ebola virus, Lassa virus (LASV) can cause hemorrhagic fever with high case fatality rates. Andersen et al. generated a genomic catalog of almost 200 LASV sequences from clinical and rodent reservoir samples (Andersen, et al., Cell Volume 162, Issue 4, p 738–750, 13 August 2015). Andersen et al. show that whereas the 2013–2015 EVD epidemic is fueled by human-to-human transmissions, LASV infections mainly result from reservoir-to-human infections. Andersen et al. elucidated the spread of LASV across West Africa and show that this migration was accompanied by changes in LASV genome abundance, fatality rates, codon adaptation, and translational efficiency. The method may further comprise phylogenetically comparing a first pathogen sequence to a second pathogen sequence and determining whether there is a phylogenetic link between the first and second pathogen sequences. The second pathogen sequence may be an earlier reference sequence. If there is a phylogenetic link, the method may further comprise rooting the phylogeny of the first pathogen sequence to the second pathogen sequence. Thus, it is possible to construct the lineage of the first pathogen sequence. (Park, et al., 2015).

[0565] The method may further comprise determining whether the mutations are deleterious or adaptive. Deleterious mutations are indicative of transmission-impaired viruses and dead-end infections, thus normally only present in an individual subject. Mutations unique to one individual subject are those that occur on the external branches of the phylogenetic tree, whereas internal branch mutations are those present in multiple samples (i.e. in multiple subjects). Higher rate of nonsynonymous substitution is a characteristic of external branches of the phylogenetic tree (Park, et al., 2015).

[0566] In internal branches of the phylogenetic tree, selection has had more opportunity to filter out deleterious mutants. Internal branches, by definition, have produced multiple descendent lineages and are thus less likely to include mutations with fitness costs. Thus, lower rate of nonsynonymous substitution is indicative of internal branches (Park, et al., 2015).

[0567] Synonymous mutations, which likely have less impact on fitness, occurred at more comparable frequencies on internal and external branches (Park, et al., 2015).

[0568] By analyzing the sequenced target sequence, such as viral genomes, it is possible to discover the mechanisms responsible for the severity of the epidemic episode such as during the 2014 Ebola outbreak. For example, Gire et al. made a phylogenetic comparison of the genomes of the 2014 outbreak to all 20 genomes from earlier outbreaks suggests that the 2014 West African virus likely spread from central Africa within the past decade. Rooting the phylogeny using divergence from other ebolavirus genomes was problematic (6, 13). However, rooting the tree on the oldest outbreak revealed a strong correlation between sample date and root-to-tip distance, with a substitution rate of 8×10^{-4} per site per year (13). This suggests that the lineages of the three most recent outbreaks all diverged from a common ancestor at roughly the same time, around 2004, which supports the hypothesis that each outbreak represents an independent zoonotic event from the same genetically diverse viral population in its natural reservoir. They also found out that the 2014 EBOV outbreak might be caused by a single transmission from the natural reservoir, followed by human-to-human transmission during the outbreak. Their results also suggested that the epidemic episode in Sierra Leon might stem from the introduction of two genetically distinct viruses from Guinea around the same time (Gire, et al., 2014).

[0569] It has been also possible to determine how the Lassa virus spread out from its origin point, in particular thanks to human-to-human transmission and even retrace the history of this spread 400 years back (Andersen, et al., Cell 162(4):738–50, 2015).

[0570] In relation to the work needed during the 2013–2015 EBOV outbreak and the difficulties encountered by the medical staff at the site of the outbreak, and more generally, the method of the invention makes it possible to carry out sequencing using fewer selected probes such that sequencing can be accelerated, thus shortening the time needed from sample taking to results procurement. Further, kits and systems can be designed to be usable on the field so that diagnostics of a patient can be readily performed without need to send or ship samples to another part of the country or the world.

[0571] In any method described above, sequencing the target sequence or fragment thereof may be used any of the sequencing processes described above. Further, sequencing the target sequence or fragment thereof may be a near-real-time sequencing. Sequencing the target sequence or fragment thereof may be carried out according to previously described methods

(Experimental Procedures: Matranga et al., 2014; and Gire, et al., 2014). Sequencing the target sequence or fragment thereof may comprise parallel sequencing of a plurality of target sequences. Sequencing the target sequence or fragment thereof may comprise Illumina sequencing.

[0572] Analyzing the target sequence or fragment thereof that hybridizes to one or more of the selected probes may be an identifying analysis, wherein hybridization of a selected probe to the target sequence or a fragment thereof indicates the presence of the target sequence within the sample.

[0573] Currently, primary diagnostics are based on the symptoms a patient has. However, various diseases may share identical symptoms so that diagnostics rely much on statistics. For example, malaria triggers flu-like symptoms: headache, fever, shivering, joint pain, vomiting, hemolytic anemia, jaundice, hemoglobin in the urine, retinal damage, and convulsions. These symptoms are also common for septicemia, gastroenteritis, and viral diseases. Amongst the latter, Ebola hemorrhagic fever has the following symptoms fever, sore throat, muscular pain, headaches, vomiting, diarrhea, rash, decreased function of the liver and kidneys, internal and external hemorrhage.

[0574] When a patient is presented to a medical unit, for example in tropical Africa, basic diagnostics will conclude to malaria because statistically, malaria is the most probable disease within that region of Africa. The patient is consequently treated for malaria although the patient might not actually have contracted the disease and the patient ends up not being correctly treated. This lack of correct treatment can be life-threatening especially when the disease the patient contracted presents a rapid evolution. It might be too late before the medical staff realizes that the treatment given to the patient is ineffective and comes to the correct diagnostics and administers the adequate treatment to the patient.

[0575] The method of the invention provides a solution to this situation. Indeed, because the number of nucleic acid component molecules can be dramatically reduced, this makes it possible to provide on a single chip selected probes divided into groups, each group being specific to one disease, such that a plurality of diseases, e.g. viral infection, can be diagnosed at the same time. Thanks to the invention, more than 3 diseases can be diagnosed on a single chip, preferably more than 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 diseases at the same time, preferably the diseases that most commonly occur within the population of a given geographical area. Since each group of selected probes is specific to one of the diagnosed

diseases, a more accurate diagnosis can be performed, thus diminishing the risk of administering the wrong treatment to the patient.

[0576] In other cases, a disease such as a viral infection may occur without any symptoms or had caused symptoms but dissipated before the patient is presented to the medical staff. In such cases, either the patient does not seek any medical assistance, or the diagnostics is complicated due to the absence of symptoms on the day of the presentation.

[0577] The present invention may also be used in concert with other methods of diagnosing disease, identifying pathogens, and optimizing treatment based upon detection of nucleic acids, such as mRNA in crude, non-purified samples.

[0578] The method of the invention also provides a powerful tool to address this situation. Indeed, since a plurality of groups of selected nucleic acid component molecules, each group being specific to one of the most common diseases that occur within the population of the given area, are comprised within a single diagnostic, the medical staff only need to contact a biological sample taken from the patient with the chip. Reading the chip reveals the diseases the patient has contracted.

[0579] In some cases, the patient is presented to the medical staff for diagnostics of particular symptoms. The method of the invention makes it possible not only to identify which disease causes these symptoms but at the same time determine whether the patient suffers from another disease he was not aware of.

[0580] This information might be of utmost importance when searching for the mechanisms of an outbreak. Indeed, groups of patients with identical viruses also show temporal patterns suggesting a subject-to-subject transmission links.

Example Microbes

[0581] The embodiment disclosed herein may be used to detect a number of different microbes. The term microbe as used herein includes bacteria, fungus, protozoa, parasites, and viruses.

Bacteria

[0582] The following provides an example list of the types of microbes that might be detected using the embodiments disclosed herein. In certain example embodiments, the microbe is a bacterium. Examples of bacteria that can be detected in accordance with the disclosed methods include without limitation any one or more of (or any combination of) *Acinetobacter baumannii*, *Actinobacillus sp.*, *Actinomyces*, *Actinomyces sp.* (such as

Actinomyces israelii and *Actinomyces naeshundii*), *Aeromonas* sp. (such as *Aeromonas hydrophila*, *Aeromonas veronii* biovar *sobria* (*Aeromonas sobria*), and *Aeromonas caviae*), *Anaplasma phagocytophilum*, *Anaplasma marginale*, *Alcaligenes xylosoxidans*, *Acinetobacter baumannii*, *Actinobacillus actinomycesemcomitans*, *Bacillus* sp. (such as *Bacillus anthracis*, *Bacillus cereus*, *Bacillus subtilis*, *Bacillus thuringiensis*, and *Bacillus stearothermophilus*), *Bacteroides* sp. (such as *Bacteroides fragilis*), *Bartonella* sp. (such as *Bartonella bacilliformis* and *Bartonella henselae*, *Bifidobacterium* sp., *Bordetella* sp. (such as *Bordetella pertussis*, *Bordetella parapertussis*, and *Bordetella bronchiseptica*), *Borrelia* sp. (such as *Borrelia recurrentis*, and *Borrelia burgdorferi*), *Brucella* sp. (such as *Brucella abortus*, *Brucella canis*, *Brucella melintensis* and *Brucella suis*), *Burkholderia* sp. (such as *Burkholderia pseudomallei* and *Burkholderia cepacia*), *Campylobacter* sp. (such as *Campylobacter jejuni*, *Campylobacter coli*, *Campylobacter lari* and *Campylobacter fetus*), *Capnocytophaga* sp., *Cardiobacterium hominis*, *Chlamydia trachomatis*, *Chlamydophila pneumoniae*, *Chlamydophila psittaci*, *Citrobacter* sp., *Coxiella burnetii*, *Corynebacterium* sp. (such as, *Corynebacterium diphtheriae*, *Corynebacterium jeikeum* and *Corynebacterium*), *Clostridium* sp. (such as *Clostridium perfringens*, *Clostridium difficile*, *Clostridium botulinum* and *Clostridium tetani*), *Eikenella corrodens*, *Enterobacter* sp. (such as *Enterobacter aerogenes*, *Enterobacter agglomerans*, *Enterobacter cloacae* and *Escherichia coli*, including opportunistic *Escherichia coli*, such as enterotoxigenic *E. coli*, enteroinvasive *E. coli*, enteropathogenic *E. coli*, enterohemorrhagic *E. coli*, enteroaggregative *E. coli* and uropathogenic *E. coli*) *Enterococcus* sp. (such as *Enterococcus faecalis* and *Enterococcus faecium*) *Ehrlichia* sp. (such as *Ehrlichia chaffeensis* and *Ehrlichia canis*), *Epidermophyton floccosum*, *Erysipelothrix rhusiopathiae*, *Eubacterium* sp., *Francisella tularensis*, *Fusobacterium nucleatum*, *Gardnerella vaginalis*, *Gemella morbillorum*, *Haemophilus* sp. (such as *Haemophilus influenzae*, *Haemophilus ducreyi*, *Haemophilus aegyptius*, *Haemophilus parainfluenzae*, *Haemophilus haemolyticus* and *Haemophilus parahaemolyticus*, *Helicobacter* sp. (such as *Helicobacter pylori*, *Helicobacter cinaedi* and *Helicobacter fennelliae*), *Kingella kingii*, *Klebsiella* sp. (such as *Klebsiella pneumoniae*, *Klebsiella granulomatis* and *Klebsiella oxytoca*), *Lactobacillus* sp., *Listeria monocytogenes*, *Leptospira interrogans*, *Legionella pneumophila*, *Leptospira interrogans*, *Peptostreptococcus* sp., *Mannheimia hemolytica*, *Microsporium canis*, *Moraxella catarrhalis*, *Morganella* sp., *Mobiluncus* sp., *Micrococcus* sp., *Mycobacterium* sp. (such as *Mycobacterium leprae*, *Mycobacterium tuberculosis*, *Mycobacterium paratuberculosis*, *Mycobacterium*

intracellulare, *Mycobacterium avium*, *Mycobacterium bovis*, and *Mycobacterium marinum*), *Mycoplasma* sp. (such as *Mycoplasma pneumoniae*, *Mycoplasma hominis*, and *Mycoplasma genitalium*), *Nocardia* sp. (such as *Nocardia asteroides*, *Nocardia cyriacigeorgica* and *Nocardia brasiliensis*), *Neisseria* sp. (such as *Neisseria gonorrhoeae* and *Neisseria meningitidis*), *Pasteurella multocida*, *Pityrosporum orbiculare* (*Malassezia furfur*), *Plesiomonas shigelloides*, *Prevotella* sp., *Porphyromonas* sp., *Prevotella melaninogenica*, *Proteus* sp. (such as *Proteus vulgaris* and *Proteus mirabilis*), *Providencia* sp. (such as *Providencia alcalifaciens*, *Providencia rettgeri* and *Providencia stuartii*), *Pseudomonas aeruginosa*, *Propionibacterium acnes*, *Rhodococcus equi*, *Rickettsia* sp. (such as *Rickettsia rickettsii*, *Rickettsia akari* and *Rickettsia prowazekii*, *Orientia tsutsugamushi* (formerly: *Rickettsia tsutsugamushi*) and *Rickettsia typhi*), *Rhodococcus* sp., *Serratia marcescens*, *Stenotrophomonas maltophilia*, *Salmonella* sp. (such as *Salmonella enterica*, *Salmonella typhi*, *Salmonella paratyphi*, *Salmonella enteritidis*, *Salmonella choleraesuis* and *Salmonella typhimurium*), *Serratia* sp. (such as *Serratia marcescens* and *Serratia liquifaciens*), *Shigella* sp. (such as *Shigella dysenteriae*, *Shigella flexneri*, *Shigella boydii* and *Shigella sonnei*), *Staphylococcus* sp. (such as *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus hemolyticus*, *Staphylococcus saprophyticus*), *Streptococcus* sp. (such as *Streptococcus pneumoniae* (for example chloramphenicol-resistant serotype 4 *Streptococcus pneumoniae*, spectinomycin-resistant serotype 6B *Streptococcus pneumoniae*, streptomycin-resistant serotype 9V *Streptococcus pneumoniae*, erythromycin-resistant serotype 14 *Streptococcus pneumoniae*, optochin-resistant serotype 14 *Streptococcus pneumoniae*, rifampicin-resistant serotype 18C *Streptococcus pneumoniae*, tetracycline-resistant serotype 19F *Streptococcus pneumoniae*, penicillin-resistant serotype 19F *Streptococcus pneumoniae*, and trimethoprim-resistant serotype 23F *Streptococcus pneumoniae*, chloramphenicol-resistant serotype 4 *Streptococcus pneumoniae*, spectinomycin-resistant serotype 6B *Streptococcus pneumoniae*, streptomycin-resistant serotype 9V *Streptococcus pneumoniae*, optochin-resistant serotype 14 *Streptococcus pneumoniae*, rifampicin-resistant serotype 18C *Streptococcus pneumoniae*, penicillin-resistant serotype 19F *Streptococcus pneumoniae*, or trimethoprim-resistant serotype 23F *Streptococcus pneumoniae*), *Streptococcus agalactiae*, *Streptococcus mutans*, *Streptococcus pyogenes*, Group A streptococci, *Streptococcus pyogenes*, Group B streptococci, *Streptococcus agalactiae*, Group C streptococci, *Streptococcus anginosus*, *Streptococcus equismilis*, Group D streptococci, *Streptococcus*

bovis, Group F streptococci, and *Streptococcus anginosus* Group G streptococci), *Spirillum minus*, *Streptobacillus moniliformi*, *Treponema* sp. (such as *Treponema carateum*, *Treponema petenue*, *Treponema pallidum* and *Treponema endemicum*, *Trichophyton rubrum*, *T. mentagrophytes*, *Tropheryma whippelii*, *Ureaplasma urealyticum*, *Veillonella* sp., *Vibrio* sp. (such as *Vibrio cholerae*, *Vibrio parahemolyticus*, *Vibrio vulnificus*, *Vibrio parahaemolyticus*, *Vibrio vulnificus*, *Vibrio alginolyticus*, *Vibrio mimicus*, *Vibrio hollisae*, *Vibrio fluvialis*, *Vibrio metchnikovii*, *Vibrio damsela* and *Vibrio furnisii*), *Yersinia* sp. (such as *Yersinia enterocolitica*, *Yersinia pestis*, and *Yersinia pseudotuberculosis*) and *Xanthomonas maltophilia* among others.

Fungi

[0583] In certain example embodiments, the microbe is a fungus or a fungal species. Examples of fungi that can be detected in accordance with the disclosed methods include without limitation any one or more of (or any combination of), *Aspergillus*, *Blastomyces*, *Candidiasis*, *Coccidioidomycosis*, *Cryptococcus neoformans*, *Cryptococcus gatti*, sp. *Histoplasma* sp. (such as *Histoplasma capsulatum*), *Pneumocystis* sp. (such as *Pneumocystis jirovecii*), *Stachybotrys* (such as *Stachybotrys chartarum*), *Mucroymcosis*, *Sporothrix*, fungal eye infections ringworm, *Exserohilum*, *Cladosporium*.

[0584] In certain example embodiments, the fungus is a yeast. Examples of yeast that can be detected in accordance with disclosed methods include without limitation one or more of (or any combination of), *Aspergillus* species (such as *Aspergillus fumigatus*, *Aspergillus flavus* and *Aspergillus clavatus*), *Cryptococcus* sp. (such as *Cryptococcus neoformans*, *Cryptococcus gattii*, *Cryptococcus laurentii* and *Cryptococcus albidus*), a *Geotrichum* species, a *Saccharomyces* species, a *Hansenula* species, a *Candida* species (such as *Candida albicans*), a *Kluyveromyces* species, a *Debaryomyces* species, a *Pichia* species, or combination thereof. In certain example embodiments, the fungi are a mold. Example molds include, but are not limited to, a *Penicillium* species, a *Cladosporium* species, a *Byssoschlamys* species, or a combination thereof.

Protozoa

[0585] In certain example embodiments, the microbe is a protozoa. Examples of protozoa that can be detected in accordance with the disclosed methods and devices include without limitation any one or more of (or any combination of), Euglenozoa, Heterolobosea, Diplomonadida, Amoebozoa, Blastocystic, and Apicomplexa. Example Euglenozoa include, but are not limited to, *Trypanosoma cruzi* (Chagas disease), *T. brucei gambiense*, *T. brucei*

rhodesiense, *Leishmania braziliensis*, *L. infantum*, *L. mexicana*, *L. major*, *L. tropica*, and *L. donovani*. Example Heterolobosea include, but are not limited to, *Naegleria fowleri*. Example Diplomonadids include, but are not limited to, *Giardia intestinalis* (*G. lamblia*, *G. duodenalis*). Example Amoebozoa include, but are not limited to, *Acanthamoeba castellanii*, *Balamuthia mandrillaris*, *Entamoeba histolytica*. Example Blastocysts include, but are not limited to, *Blastocystis hominis*. Example Apicomplexa include, but are not limited to, *Babesia microti*, *Cryptosporidium parvum*, *Cyclospora cayetanensis*, *Plasmodium falciparum*, *P. vivax*, *P. ovale*, *P. malariae*, and *Toxoplasma gondii*.

Parasites

[0586] In example embodiments, the microbe is a parasite. Examples of parasites that can be detected in accordance with disclosed methods include without limitation one or more of (or any combination of), an *Onchocerca* species and a *Plasmodium* species.

Viruses

[0587] In example embodiments, the systems, devices, and methods, disclosed herein are directed to detecting viruses in a sample. The embodiments disclosed herein may be used to detect viral infection (e.g. of a subject or plant), or determination of a viral strain, including viral strains that differ by a single nucleotide polymorphism. The virus may be a DNA virus, a RNA virus, or a retrovirus. Non-limiting example of viruses useful with the present invention include, but are not limited to Ebola, measles, SARS, Chikungunya, hepatitis, Marburg, yellow fever, MERS, Dengue, Lassa, influenza, rhabdovirus or HIV. A hepatitis virus may include hepatitis A, hepatitis B, or hepatitis C. An influenza virus may include, for example, influenza A or influenza B. An HIV may include HIV 1 or HIV 2. In certain example embodiments, the viral sequence may be a human respiratory syncytial virus, Sudan ebola virus, Bundibugyo virus, Tai Forest ebola virus, Reston ebola virus, Achimota, Aedes flavivirus, Aguacate virus, Akabane virus, Alethinophid reptarenavirus, Allpahuayo mammarenavirus, Amapari mmarenavirus, Andes virus, Apoi virus, Aravan virus, Aroa virus, Arumwot virus, Atlantic salmon paramyxovirus, Australian bat lyssavirus, Avian bornavirus, Avian metapneumovirus, Avian paramyxoviruses, penguin or Falkland Islandsvirus, BK polyomavirus, Bagaza virus, Banna virus, Bat herpesvirus, Bat sapovirus, Bear Canon mammarenavirus, Beilong virus, Betacoronavirus, Betapapillomavirus 1-6, Bhanja virus, Bokeloh bat lyssavirus, Borna disease virus, Bourbon virus, Bovine hepacivirus, Bovine parainfluenza virus 3, Bovine respiratory syncytial virus, Brazoran virus, Bunyamwera virus, Caliciviridae virus. California encephalitis

virus, Candiru virus, Canine distemper virus, Canine pneumovirus, Cedar virus, Cell fusing agent virus, Cetacean morbillivirus, Chandipura virus, Chaoyang virus, Chapare mammarenavirus, Chikungunya virus, Colobus monkey papillomavirus, Colorado tick fever virus, Cowpox virus, Crimean-Congo hemorrhagic fever virus, Culex flavivirus, Cupixi mammarenavirus, Dengue virus, Dobrava-Belgrade virus, Donggang virus, Dugbe virus, Duvenhage virus, Eastern equine encephalitis virus, Entebbe bat virus, Enterovirus A-D, European bat lyssavirus 1-2, Eyach virus, Feline morbillivirus, Fer-de-Lance paramyxovirus, Fitzroy River virus, Flaviviridae virus, Flexal mammarenavirus, GB virus C, Gairo virus, Gemycircularvirus, Goose paramyxovirus SF02, Great Island virus, Guanarito mammarenavirus, Hantaan virus, Hantavirus Z10, Heartland virus, Hendra virus, Hepatitis A/B/C/E, Hepatitis delta virus, Human bocavirus, Human coronavirus, Human endogenous retrovirus K, Human enteric coronavirus, Human genital-associated circular DNA virus-1, Human herpesvirus 1-8, Human immunodeficiency virus 1/2, Human mastadenovirus A-G, Human papillomavirus, Human parainfluenza virus 1-4, Human paraechovirus, Human picornavirus, Human smacovirus, Ikoma lyssavirus, Ilheus virus, Influenza A-C, Ippy mammarenavirus, Irkut virus, J-virus, JC polyomavirus, Japanese encephalitis virus, Junin mammarenavirus, KI polyomavirus, Kadipiro virus, Kamiti River virus, Kedougou virus, Khujand virus, Kokobera virus, Kyasanur forest disease virus, Lagos bat virus, Langat virus, Lassa mammarenavirus, Latino mammarenavirus, Leopards Hill virus, Liao ning virus, Ljungan virus, Lloviu virus, Louping ill virus, Lujo mammarenavirus, Luna mammarenavirus, Lunk virus, Lymphocytic choriomeningitis mammarenavirus, Lyssavirus Ozernoe, MSSI2\225 virus, Machupo mammarenavirus, Mamastrovirus 1, Manzanilla virus, Mapuera virus, Marburg virus, Mayaro virus, Measles virus, Menangle virus, Mercadeo virus, Merkel cell polyomavirus, Middle East respiratory syndrome coronavirus, Mobala mammarenavirus, Modoc virus, Moijang virus, Mokolo virus, Monkeypox virus, Montana myotis leukoencephalitis virus, Mopeia lassa virus reassortant 29, Mopeia mammarenavirus, Morogoro virus, Mossman virus, Mumps virus, Murine pneumonia virus, Murray Valley encephalitis virus, Nariva virus, Newcastle disease virus, Nipah virus, Norwalk virus, Norway rat hepacivirus, Ntaya virus, O'nyong-nyong virus, Oliveros mammarenavirus, Omsk hemorrhagic fever virus, Oropouche virus, Parainfluenza virus 5, Parana mammarenavirus, Parramatta River virus, Peste-des-petits-ruminants virus, Pichande mammarenavirus, Picornaviridae virus, Pirital mammarenavirus, Piscihepevirus A, Porcine parainfluenza virus 1, porcine rubulavirus, Powassan virus, Primate

T-lymphotropic virus 1-2, Primate erythrovirus 1, Punta Toro virus, Puumala virus, Quang Binh virus, Rabies virus, Razdan virus, Reptile bornavirus 1, Rhinovirus A-B, Rift Valley fever virus, Rinderpest virus, Rio Bravo virus, Rodent Torque Teno virus, Rodent hepacivirus, Ross River virus, Rotavirus A-I, Royal Farm virus, Rubella virus, Sabia mammarenavirus, Salem virus, Sandfly fever Naples virus, Sandfly fever Sicilian virus, Sapporo virus, Sathuperi virus, Seal anellovirus, Semliki Forest virus, Sendai virus, Seoul virus, Sepik virus, Severe acute respiratory syndrome-related coronavirus, Severe fever with thrombocytopenia syndrome virus, Shamonda virus, Shimoni bat virus, Shuni virus, Simbu virus, Simian torque teno virus, Simian virus 40-41, Sin Nombre virus, Sindbis virus, Small anellovirus, Sosuga virus, Spanish goat encephalitis virus, Spondweni virus, St. Louis encephalitis virus, Sunshine virus, TTV-like mini virus, Tacaribe mammarenavirus, Taila virus, Tamana bat virus, Tamiami mammarenavirus, Tembusu virus, Thogoto virus, Thottapalayam virus, Tick-borne encephalitis virus, Tioman virus, Togaviridae virus, Torque teno canis virus, Torque teno douroucouli virus, Torque teno felis virus, Torque teno midi virus, Torque teno sus virus, Torque teno tamarin virus, Torque teno virus, Torque teno zalophus virus, Tuhoko virus, Tula virus, Tupaia paramyxovirus, Usutu virus, Uukuniemi virus, Vaccinia virus, Variola virus, Venezuelan equine encephalitis virus, Vesicular stomatitis Indiana virus, WU Polyomavirus, Wesselsbron virus, West Caucasian bat virus, West Nile virus, Western equine encephalitis virus, Whitewater Arroyo mammarenavirus, Yellow fever virus, Yokose virus, Yug Bogdanovac virus, Zaire ebolavirus, Zika virus, or *Zygosaccharomyces bailii* virus Z viral sequence. Examples of RNA viruses that may be detected include one or more of (or any combination of) Coronaviridae virus, a Picornaviridae virus, a Caliciviridae virus, a Flaviviridae virus, a Togaviridae virus, a Bornaviridae, a Filoviridae, a Paramyxoviridae, a Pneumoviridae, a Rhabdoviridae, an Arenaviridae, a Bunyaviridae, an Orthomyxoviridae, or a Deltavirus. In certain example embodiments, the virus is Coronavirus, SARS, Poliovirus, Rhinovirus, Hepatitis A, Norwalk virus, Yellow fever virus, West Nile virus, Hepatitis C virus, Dengue fever virus, Zika virus, Rubella virus, Ross River virus, Sindbis virus, Chikungunya virus, Borna disease virus, Ebola virus, Marburg virus, Measles virus, Mumps virus, Nipah virus, Hendra virus, Newcastle disease virus, Human respiratory syncytial virus, Rabies virus, Lassa virus, Hantavirus, Crimean-Congo hemorrhagic fever virus, Influenza, or Hepatitis D virus.

[0588] In certain example embodiments, the virus may be a plant virus selected from the group comprising Tobacco mosaic virus (TMV), Tomato spotted wilt virus (TSWV), Cucumber mosaic virus (CMV), Potato virus Y (PVY), the RT virus Cauliflower mosaic virus (CaMV), Plum pox virus (PPV), Brome mosaic virus (BMV), Potato virus X (PVX), Citrus tristeza virus (CTV), Barley yellow dwarf virus (BYDV), Potato leafroll virus (PLRV), Tomato bushy stunt virus (TBSV), rice tungro spherical virus (RTSV), rice yellow mottle virus (RYMV), rice hoja blanca virus (RHBV), maize rayado fino virus (MRFV), maize dwarf mosaic virus (MDMV), sugarcane mosaic virus (SCMV), Sweet potato feathery mottle virus (SPFMV), sweet potato sunken vein closterovirus (SPSVV), Grapevine fanleaf virus (GFLV), Grapevine virus A (GVA), Grapevine virus B (GVB), Grapevine fleck virus (GFkV), Grapevine leafroll-associated virus-1, -2, and -3, (GLRaV-1, -2, and -3), Arabis mosaic virus (ArMV), or Rupestris stem pitting-associated virus (RSPaV). In a preferred embodiment, the target RNA molecule is part of said pathogen or transcribed from a DNA molecule of said pathogen. For example, the target sequence may be comprised in the genome of an RNA virus. It is further preferred that Cas polypeptide hydrolyzes said target RNA molecule of said pathogen in said plant if said pathogen infects or has infected said plant. It is thus preferred that the CRISPR-Cas system is capable of cleaving the target RNA molecule from the plant pathogen both when the CRISPR-Cas system (or parts needed for its completion) is applied therapeutically, i.e. after infection has occurred or prophylactically, i.e. before infection has occurred.

[0589] In certain example embodiments, the virus may be a retrovirus. Example retroviruses that may be detected using the embodiments disclosed herein include one or more of or any combination of viruses of the Genus Alpharetrovirus, Betaretrovirus, Gammaretrovirus, Deltaretrovirus, Epsilonretrovirus, Lentivirus, Spumavirus, or the Family Metaviridae, Pseudoviridae, and Retroviridae (including HIV), Hepadnaviridae (including Hepatitis B virus), and Caulimoviridae (including Cauliflower mosaic virus).

[0590] In certain example embodiments, the virus is a DNA virus. Example DNA viruses that may be detected using the embodiments disclosed herein include one or more of (or any combination of) viruses from the Family Myoviridae, Podoviridae, Siphoviridae, Alloherpesviridae, Herpesviridae (including human herpes virus, and Varicella Zoster virus), Malcoherpesviridae, Lipothrixviridae, Rudiviridae, Adenoviridae, Ampullaviridae, Ascoviridae, Asfarviridae (including African swine fever virus), Baculoviridae,

Cicaudaviridae, Clavaviridae, Corticoviridae, Fuselloviridae, Globuloviridae, Guttaviridae, Hytrosaviridae, Iridoviridae, Marseilleviridae, Mimiviridae, Nudiviridae, Nimaviridae, Pandoraviridae, Papillomaviridae, Phycodnaviridae, Plasmaviridae, Polydnaviruses, Polyomaviridae (including Simian virus 40, JC virus, BK virus), Poxviridae (including Cowpox and smallpox), Sphaerolipoviridae, Tectiviridae, Turriviridae, Dinodnavirus, Salterprovirus, Rhizidovirus, among others. In one embodiment, a method of diagnosing a species-specific bacterial infection in a subject suspected of having a bacterial infection is described as obtaining a sample comprising bacterial ribosomal ribonucleic acid from the subject; contacting the sample with one or more of the probes described, and detecting hybridization between the bacterial ribosomal ribonucleic acid sequence present in the sample and the probe, wherein the detection of hybridization indicates that the subject is infected with *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*, *Acinetobacter baumannii*, *Candida albicans*, *Enterobacter cloacae*, *Enterococcus faecalis*, *Enterococcus faecium*, *Proteus mirabilis*, *Staphylococcus agalactiae*, or *Staphylococcus maltophilia* or a combination thereof.

Coronavirus

[0591] Systems and methods of the presently disclosed invention are designed to detect coronavirus, in an aspect, the target sequence is the 2019-nCoV, also referred to herein as SARS-CoV-2, which causes COVID-19. The coronavirus is a positive-sense single stranded RNA family of viruses, infecting a variety of animals and humans. SARS-CoV is one type of coronavirus infection, as well as MERS-CoV. Detection of one or more coronaviruses are envisioned, including the SARS-CoV-2 detected in Wuhan City. Sequences of the SARS-CoV-2 are available at GISAID accession no. EPI_ISL_402124 and EPI_ISL_402127-402130 and described in DOI: 10.1101/2020.01.22.914952. Further deposits of the SARS-CoV2 are deposited in the GISAID platform include EPI_ISL_402119-402121 and EPI_ISL_402123-402124; see also GenBank Accession No. MN908947.3. In an aspect, one may use known SARS and SARS-related coronaviruses or other viruses from one or more hosts to generate a non-redundant alignment. Related viruses can be found, for example in bats.

[0592] In one embodiment, the systems are designed to comprise at least one highly active nucleic acid component polynucleotide which is designed according to the methods disclosed herein. In a preferred embodiment, the nucleic acid component polynucleotide binds to at least one target sequence that is a unique coronavirus genomic sequence, thereby identifying the

presence of coronavirus to the exclusion of other viruses. The systems and methods can be designed to detect a plurality of respiratory infections or viral infections, including coronavirus.

[0593] In an aspect the at least one nucleic acid component polynucleotide binds to a coronavirus sequence encoding a polypeptide that is immunostimulatory to a host immune system. Immunostimulatory polypeptides have the ability to enhance, stimulate, or increase response of the immune system, typically by inducing the activation or activity of a components of the immune system (e.g. an immune cell). In embodiments, the immunostimulatory polypeptide contributes to immune-mediated disease in the host. In an aspect, the host is a mammal, for example, a human, a bat, or a pangolin, that may be infected by a coronavirus. Cyranoski, D. Did pangolins spread the China coronavirus to people? *Nature*, 7 Feb. 2020. In one embodiment, the nucleic acid component polynucleotide can be designed to detect SARS-CoV-2 or a variant thereof in meat, live animals and humans so that testing can be performed, for example at markets and other public places where sources of contamination can arise.

[0594] Gene targets may comprise ORF1ab, N protein, RNA-dependent RNA polymerase (RdRP), E protein, ORF1b-nsp14, Spike glycoprotein (S), or pancorona targets. Molecular assays have been under development and can be used as a starting point to develop nucleic acid component molecules for the methods and systems described herein. See, “Diagnostic detection of 2019-nCoV by real-time RT-PCR” Charité, Berlin Germany (17 January 2020); Detection of 2019 novel coronavirus (2019-nCoV) in suspected human cases by RT-PCR – Hong Kong University (23 January 2020); PCR and sequencing protocol for 2019-nCoV - Department of Medical Sciences, Ministry of Public Health, Thailand (updated 28 January 2020); PCR and sequencing protocols for 2019-nCoV- National Institute of Infectious Diseases Japan (24 January 2020); US CDC panel primer and probes– U.S. CDC, USAV – U.S. CDC, USA (28 January 2020); China CDC Primers and probes for detection 2019-nCoV (24 January 2020), incorporated in their entirety by reference. Further, the nucleic acid component molecule design may exploit differences or similarities with SARS-CoV. Researchers have recently identified similarities and differences between 2019-nCoV and SARS-CoV. “Coronavirus Genome Annotation Reveals Amino Acid Differences with Other SARS Viruses,” genomeweb, February 10, 2020. For example, nucleic acid component molecules based on the 8a protein, which was present in SARS-CoV but absent in SARS-CoV-2, can be utilized to differentiate between the viruses. Similarly, the 8b and 3b proteins have different lengths in SARS -CoV and SARS-CoV-2 and can be utilized to design nucleic acid component

molecules to detect non-overlapping proteins of nucleotides encoding in the two viruses. Wu et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China, Cell Host & Microbe (2020), DOI: 10.1016/j.chom.2020.02.001, incorporated herein by reference, including all supplemental information, in particular Table S1. Mutations may also be detected, with nucleic acid component and/or primers designed specifically to detect, for example, changes in the SARS-CoV-2 virus. In an embodiment, the nucleic acid component or primer can be designed to detect the D614G mutation in the SARS-CoV-2 spike protein. See, Korber et al., Cell 182, 812-827 (2020); doi: 10.1016/j.cell.2020.06.043. Other mutations in the spike protein can be designed utilizing the COVID-19 viral genome analysis pipeline available at cov.lanl.gov. Further resources to design primers and nucleic acid components to detect coronavirus or coronavirus mutations can be found at Starr, et al., , “Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints in Folding and ACE2 Binding,” Cell, 182, 1-16 (2020); doi: 10.1016/j.cell.2020.08.012.

[0595] The systems and methods of detection can be used to identify single nucleotide variants, detection based on rRNA sequences, screening for drug resistance, monitoring microbe outbreaks, genetic perturbations, and screening of environmental samples, as described in PCT/US2018/054472 filed October 22, 2018 at [0183] – [0327], incorporated herein by reference.

[0596] In certain example embodiments, the systems, devices, and methods disclosed herein may be used for biomarker detection. For example, the systems, devices, and method disclosed herein may be used for SNP detection and/or genotyping. The systems, devices and methods disclosed herein may be also used for the detection of any disease state or disorder characterized by aberrant gene expression. Aberrant gene expression includes aberration in the gene expressed, location of expression and level of expression. Multiple transcripts or protein markers related to cardiovascular, immune disorders, and cancer among other diseases may be detected. In certain example embodiments, the embodiments disclosed herein may be used for cell free DNA detection of diseases that involve lysis. In certain example embodiments, the embodiments could be utilized for faster and more portable detection for pre-natal testing of cell-free DNA. The embodiments disclosed herein may be used for screening panels of different SNPs associated with, among others, different coronaviruses, evolving SARS-CoV2, and other related respiratory viral infections. As described herein elsewhere, closely related

genotypes/alleles or biomarkers (e.g. having only a single nucleotide difference in a given target sequence) may be distinguished by introduction of a synthetic mismatch in the nucleic acid component molecule.

[0597] In an aspect, the invention relates to a method for detecting target nucleic acids in samples, comprising: distributing a sample or set of samples into one or more individual discrete volumes, the individual discrete volumes comprising a CRISPR-Cas system according to the invention as described herein; incubating the sample or set of samples under conditions sufficient to allow binding of the one or more nucleic acid component molecules to one or more target molecules; activating the Cas polypeptide via binding of the one or more nucleic acid component molecules to the one or more target molecules, wherein activating the Cas polypeptide results in modification of the RNA-based masking construct such that a detectable positive signal is generated; and detecting the detectable positive signal, wherein detection of the detectable positive signal indicates a presence of one or more target molecules in the sample.

[0598] The sensitivity of the assays described herein are well suited for detection of target nucleic acids in a wide variety of biological sample types, including sample types in which the target nucleic acid is dilute or for which sample material is limited. Methods for field deployable and rapid diagnostic assays can be optimized for the type of sample material utilized and can be adapted from approaches used for other assays known in the art. See, e.g. Myhrvold et al., 2018. Biomarker screening may be carried out on a number of sample types including, but not limited to, saliva, urine, blood, feces, sputum, and cerebrospinal fluid. The embodiments disclosed herein may also be used to detect up- and/or down-regulation of genes. For example, a sample may be serially diluted such that only over-expressed genes remain above the detection limit threshold of the assay.

[0599] In one embodiment, the present invention provides steps of obtaining a sample of biological fluid (e.g., urine, blood plasma or serum, sputum, cerebral spinal fluid), and extracting the DNA or RNA. The mutant nucleotide sequence to be detected, may be a fraction of a larger molecule or can be present initially as a discrete molecule.

[0600] In embodiments, DNA is isolated from plasma/serum of a cancer patient. For comparison, DNA samples isolated from neoplastic tissue and a second sample may be isolated from non-neoplastic tissue from the same patient (control), for example, lymphocytes. The non-neoplastic tissue can be of the same type as the neoplastic tissue or from a different organ

source. In one embodiment, blood samples are collected, and plasma immediately separated from the blood cells by centrifugation. Serum may be filtered and stored frozen until DNA/RNA extraction.

[0601] In an aspect, sample preparation can comprise methods as disclosed herein to circumvent other RNA extraction methods and can be used with standard amplification techniques such as RT-PCR as well as the CRISPR-Cas detection methods disclosed herein. In an aspect, the method may comprise a one-step extraction-free RNA preparation method that can be used with samples tested for coronavirus, which may be, in an aspect, a RT-qPCR testing method, a lateral flow detection method or other CRISPR-Cas detection method disclosed herein. Advantageously, the RNA extraction method can be utilized directly with other testing protocols. In an aspect, the method comprises use of a nasopharyngeal swab, nasal saline lavage, or other nasal sample (e.g., anterior nasal swab) with Quick Extract™ DNA Extraction Solution (QE09050), Lucigen, or QuickExtract Plant DNA Extraction Solution, Lucigen. In an aspect, the sample is diluted 2:1, 1:1 or 1:2 sample:DNA extraction solution. The sample:extraction mix is incubated at about 90 °C to about 98 °C, preferably about 95 °C. In another aspect, incubation is performed at between about 20°C to about 90°C, about 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85 86, 87, 88, 89 or 90 °C. The incubation period can be about 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, or 30 minutes, preferably about 4 to 6 minutes, or about 5 minutes. Incubation time and temperature may vary depending on sample size and quality, and incubation time may increase if using lower temperature. Current CDC Real-Time RT-PCR Diagnostic Panel are as described at [fda.gov/media/134922/download](https://www.fda.gov/media/134922/download), “CDC 2019-Novel Coronavirus (2019-nCoV) Real-Time RT-PCR Diagnostic Panel.” In one embodiment, the DNA extraction solution can remain with the sample subsequent to incubation and be utilized in the next steps of detection methods. In an aspect, the detection method is an RT-qPCR reaction, and the extraction solution is kept at a concentration of less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3% of the reaction mixture, where the reaction mixture comprises the detection reaction reagents, sample, and extraction solution.

[0602] In one embodiment, a bead is utilized with particular embodiments of the invention and may be included with the extraction solution. The bead may be used to capture, concentrate,

or otherwise enrich for particular material. The bead may be magnetic and may be provided to capture nucleic acid material. In another aspect, the bead is a silica bead. Beads may be utilized in an extraction step of the methods disclosed herein. Beads can be optionally used with the methods described herein, including with the one-pot methods that allow for concentration of viral nucleic acids from large volume samples, such as saliva or swab samples to allow for a single one-pot reaction method. Concentration of desired target molecules can be increased by about 10-fold, 50-fold, 100-fold, 200-fold, 500-fold, 800-fold, 1000-fold, 1500-fold, 2000-fold, 2500-fold, 3000-fold, or more.

[0603] Magnetic beads in a PEG and salt solution are preferred in an aspect, and in embodiments bind to viral RNA and/or DNA which allows for concentration and lysis concurrently. Silica beads can be used in another aspect. Capture moieties such as oligonucleotide functionalized beads are envisioned for use. The beads may be used with the extraction reagents, allowed to incubate with a sample and the lysis/extraction buffer, thereby concentrating target molecules on the beads. Extraction can be performed as described elsewhere herein, at 22 °C-60°C, with subsequent isothermal amplification and/or CRISPR-Cas detection performed under conditions as described elsewhere herein. When used with a cartridge device detailed elsewhere herein, a magnet can be activated and the beads collected, with optional flushing of the extraction buffer and one or more washes performed. Advantageously, the beads can be used in the one-pot methods and systems without additional washings of the beads, allowing for a more efficient process without increased risks of contamination in multi-step processes. Beads can be utilized with the isothermal amplifications detailed herein, and the beads can flow into an amplification chamber of the cartridge or be maintained in the pot for the amplification step. Upon heating, nucleic acid can be released off the beads.

[0604] In certain example embodiments, target nucleic acids are detected directly from a crude or unprocessed sample, such as blood, serum, saliva, cerebrospinal fluid, sputum, or urine. In certain example embodiments, the target nucleic acid is cell free DNA.

POLYNUCLEOTIDES ENCODING CRISPR-CAS SYSTEMS AND VECTORS

[0605] The systems herein may comprise one or more polynucleotides. The polynucleotide(s) may comprise coding sequences of components of the systems herein, e.g.,

Cas12b polypeptide guide sequence RNA(s), functional domain(s), donor polynucleotide(s), and/or other components in the systems. The present disclosure further provides vectors or vector systems comprising one or more polynucleotides herein. The vectors or vector systems include those described in the delivery sections herein.

[0606] 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 97/03211; 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. “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, Hoogstein 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.

[0607] In an embodiment, the polynucleotide sequence is recombinant DNA. In further embodiments, the polynucleotide sequence further comprises additional sequences as described elsewhere herein. In an embodiment, the nucleic acid sequence is synthesized *in vitro*.

[0608] The present disclosure provides polynucleotide molecules that encode one or more components of the system or Cas12b polypeptide nuclease as referred to in any embodiment herein. In an embodiment, the polynucleotide molecules may comprise further regulatory sequences. By means of guidance and not limitation, the polynucleotide sequence can be part of an expression plasmid, a minicircle, a lentiviral vector, a retroviral vector, an adenoviral or adeno-associated viral vector, a piggyback vector, or a tol2 vector. In an embodiment, the polynucleotide sequence may be a bicistronic expression construct. In further embodiments, the isolated polynucleotide sequence may be incorporated in a cellular genome. In yet further embodiments, the isolated polynucleotide sequence may be part of a cellular genome. In further embodiments, the isolated polynucleotide sequence may be comprised in an artificial chromosome. In an embodiment, the 5' and/or 3' end of the isolated polynucleotide sequence may be modified to improve the stability of the sequence or actively avoid degradation. In an embodiment, the isolated polynucleotide sequence may be comprised in a bacteriophage. In other embodiments, the isolated polynucleotide sequence may be contained in agrobacterium species. In an embodiment, the isolated polynucleotide sequence is lyophilized.

[0609] Aspects of the invention relate to polynucleotide molecules that encode one or more components of one or more systems as described in any of the embodiments herein, wherein at least one or more regions of the polynucleotide molecule may be codon optimized for expression in eukaryotic cells. In an embodiment, the polynucleotide molecules that encode one or more components of one or more systems as described in any of the embodiments herein are optimized for expression in a mammalian cell or a plant cell.

[0610] 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. In one embodiment, an enzyme coding sequence encoding a DNA/RNA-targeting Cas12b polypeptide 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 plant or 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 one embodiment, 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 one embodiment, 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 Cas12b polypeptide nuclease corresponds to the most frequently used codon for a particular amino acid.

DELIVERY SYSTEMS

[0611] The present disclosure also provides delivery systems for introducing components of the systems and compositions herein to cells, tissues, organs, or organisms. A delivery system may comprise one or more delivery vehicles and/or cargos. Exemplary delivery systems

and methods include those described in paragraphs [00117] to [00278] of Feng Zhang et al., (WO2016106236A1), and pages 1241-1251 and Table 1 of Lino CA et al., *Delivering CRISPR: a review of the challenges and approaches*, DRUG DELIVERY, 2018, VOL. 25, NO. 1, 1234–1257, which are incorporated by reference herein in their entireties and can be adapted for use with the Cas12b proteins disclosed herein.

[0612] In one embodiment, the delivery systems may be used to introduce the components of the systems and compositions to plant cells. For example, the components may be delivered to plant using electroporation, microinjection, aerosol beam injection of plant cell protoplasts, biolistic methods, DNA particle bombardment, and/or Agrobacterium-mediated transformation. Examples of methods and delivery systems for plants include those described in Fu et al., *Transgenic Res.* 2000 Feb;9(1):11-9; Klein RM, et al., *Biotechnology*. 1992;24:384-6; Casas AM et al., *Proc Natl Acad Sci U S A.* 1993 Dec 1; 90(23): 11212–11216; and U.S. Pat. No. 5,563,055, Davey MR et al., *Plant Mol Biol.* 1989 Sep;13(3):273-85, which are incorporated by reference herein in their entireties.

[0613] The example delivery compositions, systems, and methods described herein related to composition or Cas12b polypeptide also apply to functional domains and other components (e.g., other proteins and polynucleotides related to the Cas12b polypeptide such as reverse transcriptase, nucleotide deaminase, retrotransposon, donor polynucleotide, etc.).

Cargos

[0614] The delivery systems may comprise one or more cargos. The cargos may comprise one or more components of the systems and compositions herein. A cargo may comprise one or more of the following: i) a plasmid encoding one or more proteins components in the compositions and systems such as the Cas12b polypeptide and/or functional domains; ii) a plasmid encoding one or more guide sequences, iii) mRNA of one or more one or more proteins components in the compositions and systems such as the Cas12b polypeptide and/or functional domains; iv) one or more guide sequence RNAs; v) one or more proteins components in the compositions and systems such as the Cas12b polypeptide and/or functional domains; vi) any combination thereof. The one or more protein components may include the nucleic acid-guided nuclease (e.g., Cas), reverse transcriptase, nucleotide deaminase, retrotransposon protein, other functional domain, or any combination thereof.

[0615] In some examples, a cargo may comprise a plasmid encoding one or more proteins components in the compositions and systems such as the Cas12b polypeptide and/or functional

domains and one or more (e.g., a plurality of) guide sequence RNAs. In some cases, the plasmid may also encode a recombination template (e.g., for HDR). In one embodiment, a cargo may comprise mRNA encoding one or more protein components and one or more guide sequence RNAs.

[0616] In some examples, a cargo may comprise one or more protein components and one or more guide sequence RNAs, e.g., in the form of ribonucleoprotein complexes (RNP). The ribonucleoprotein complexes may be delivered by methods and systems herein. In some cases, the ribonucleoprotein may be delivered by way of a polypeptide-based shuttle agent. In one example, the ribonucleoprotein may be delivered using synthetic peptides comprising an endosome leakage domain (ELD) operably linked to a cell penetrating domain (CPD), to a histidine-rich domain and a CPD, e.g., as describe in WO2016161516. RNP may also be used for delivering the compositions and systems to plant cells, e.g., as described in Wu JW, et al., Nat Biotechnol. 2015 Nov;33(11):1162-4.

Physical Delivery

[0617] In one embodiment, the cargos may be introduced to cells by physical delivery methods. Examples of physical methods include microinjection, electroporation, and hydrodynamic delivery. Both nucleic acid and proteins may be delivered using such methods. For example, one or more protein components may be prepared *in vitro*, isolated, (refolded, purified if needed), and introduced to cells.

Microinjection

[0618] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In one embodiment, microinjection may be performed using a microscope and a needle (e.g., with 0.5–5.0 μm in diameter) to pierce a cell membrane and deliver the cargo directly to a target site within the cell. Microinjection may be used for *in vitro* and *ex vivo* delivery.

[0619] Plasmids comprising coding sequences for one or more protein components and/or mRNAs, and/or guide sequence RNAs, may be microinjected. In some cases, microinjection may be used i) to deliver DNA directly to a cell nucleus, and/or ii) to deliver mRNA (e.g., *in vitro* transcribed) to a cell nucleus or cytoplasm. In certain examples, microinjection may be used to delivery guide sequence directly to the nucleus and mRNA to the cytoplasm, e.g., facilitating translation and shuttling of one or more protein components to the nucleus.

[0620] Microinjection may be used to generate genetically modified animals. For example, gene editing cargos may be injected into zygotes to allow for efficient germline modification. Such approach can yield normal embryos and full-term mouse pups harboring the desired modification(s). Microinjection can also be used to provide transiently up- or down- regulate a specific gene within the genome of a cell, e.g., using Cas12b.

Electroporation

[0621] In one embodiment, the cargos and/or delivery vehicles may be delivered by electroporation. Electroporation may use pulsed high-voltage electrical currents to transiently open nanometer-sized pores within the cellular membrane of cells suspended in buffer, allowing for components with hydrodynamic diameters of tens of nanometers to flow into the cell. In some cases, electroporation may be used on various cell types and efficiently transfer cargo into cells. Electroporation may be used for *in vitro* and *ex vivo* delivery.

[0622] Electroporation may also be used to deliver the cargo to into the nuclei of mammalian cells by applying specific voltage and reagents, e.g., by nucleofection. Such approaches include those described in Wu Y, et al. (2015). *Cell Res* 25:67–79; Ye L, et al. (2014). *Proc Natl Acad Sci USA* 111:9591–6; Choi PS, Meyerson M. (2014). *Nat Commun* 5:3728; Wang J, Quake SR. (2014). *Proc Natl Acad Sci* 111:13157–62. Electroporation may also be used to deliver the cargo *in vivo*, e.g., with methods described in Zuckermann M, et al. (2015). *Nat Commun* 6:7391.

Hydrodynamic delivery

[0623] Hydrodynamic delivery may also be used for delivering the cargos, e.g., for *in vivo* delivery. In some examples, hydrodynamic delivery may be performed by rapidly pushing a large volume (8–10% body weight) solution containing the gene editing cargo into the bloodstream of a subject (e.g., an animal or human), e.g., for mice, via the tail vein. As blood is incompressible, the large bolus of liquid may result in an increase in hydrodynamic pressure that temporarily enhances permeability into endothelial and parenchymal cells, allowing for cargo not normally capable of crossing a cellular membrane to pass into cells. This approach may be used for delivering naked DNA plasmids and proteins. The delivered cargos may be enriched in liver, kidney, lung, muscle, and/or heart.

Transfection

[0624] The cargos, e.g., nucleic acids, may be introduced to cells by transfection methods for introducing nucleic acids into cells. Examples of transfection methods include calcium

phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acid.

Delivery Vehicles

[0625] The delivery systems may comprise one or more delivery vehicles. The delivery vehicles may deliver the cargo into cells, tissues, organs, or organisms (e.g., animals or plants). The cargos may be packaged, carried, or otherwise associated with the delivery vehicles. The delivery vehicles may be selected based on the types of cargo to be delivered, and/or the delivery is *in vitro* and/or *in vivo*. Examples of delivery vehicles include vectors, viruses, non-viral vehicles, and other delivery reagents described herein.

[0626] The delivery vehicles in accordance with the present invention may have a greatest dimension (e.g. diameter) of less than 100 microns (μm). In one embodiment, the delivery vehicles have a greatest dimension of less than 10 μm . In one embodiment, the delivery vehicles may have a greatest dimension of less than 2000 nanometers (nm). In one embodiment, the delivery vehicles may have a greatest dimension of less than 1000 nanometers (nm). In one embodiment, the delivery vehicles may have a greatest dimension (e.g., diameter) of less than 900 nm, less than 800 nm, less than 700 nm, less than 600 nm, less than 500 nm, less than 400 nm, less than 300 nm, less than 200 nm, less than 150nm, or less than 100nm, less than 50nm. In one embodiment, the delivery vehicles may have a greatest dimension ranging between 25 nm and 200 nm.

[0627] In one embodiment, the delivery vehicles may be or comprise particles. For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension (e.g., diameter) no greater than 1000nm. The particles 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). Nanoparticles may also be used to deliver the compositions and systems to plant cells, e.g., as described in International Patent Publication No. WO 2008042156, US Publication Application No. US 20130185823, and International Patent Publication No WO 2015/089419.

Vectors

[0628] The systems, compositions, and/or delivery systems may comprise one or more vectors. The present disclosure also include vector systems. A vector system may comprise one or more vectors. In one embodiment, a vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include 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. A vector may be a plasmid, e.g., a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Certain vectors may be 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). Some 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. In certain examples, vectors may be expression vectors, e.g., capable of directing the expression of genes to which they are operatively-linked. In some cases, the expression vectors may be for expression in eukaryotic cells. Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0629] Examples of vectors include pGEX, pMAL, pRIT5, E. coli expression vectors (e.g., pTrc, pET 11d, yeast expression vectors (e.g., pYepSec1, pMFa, pJRY88, pYES2, and picZ, Baculovirus vectors (e.g., for expression in insect cells such as SF9 cells) (e.g., pAc series and the pVL series), mammalian expression vectors (e.g., pCDM8 and pMT2PC.

[0630] A vector may comprise i) one or more protein components encoding sequence(s), and/or ii) a single, or at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 14, at least 16, at least 32, at least 48, at least 50 guide RNA(s) encoding sequences. In a single vector there can be a promoter for each RNA coding sequence. Alternatively or additionally, in a single vector, there may be a promoter controlling (e.g., driving transcription and/or expression) multiple RNA encoding sequences.

[0631] Furthermore, that compositions or systems may be delivered via a vector, e.g., a separate vector or the same vector that is encoding the complex. When provided by a separate vector, the RNA that targets Cas12b polypeptide expression can be administered sequentially or simultaneously. When administered sequentially, the RNA that targets Cas12b polypeptide

expression is to be delivered after the RNA that is intended for e.g. gene editing or gene engineering. This period may be a period of minutes (e.g. 5 minutes, 10 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes). This period may be a period of hours (e.g. 2 hours, 4 hours, 6 hours, 8 hours, 12 hours, 24 hours). This period may be a period of days (e.g. 2 days, 3 days, 4 days, 7 days). This period may be a period of weeks (e.g. 2 weeks, 3 weeks, 4 weeks). This period may be a period of months (e.g. 2 months, 4 months, 8 months, 12 months). This period may be a period of years (2 years, 3 years, 4 years). In this fashion, the Cas12b polypeptide nuclease associates with a first guide sequence RNA capable of hybridizing to a first target, such as a genomic locus or loci of interest and undertakes the function(s) desired of the system (e.g., gene engineering); and subsequently the Cas12b polypeptide may then associate with the second guide sequence RNA capable of hybridizing to the sequence comprising at least part of the Cas12b polypeptide. Where the guide RNA targets the sequences encoding expression of the Cas12b polypeptide, the enzyme becomes impeded, and the system becomes self-inactivating. In the same manner, RNA that targets Cas12b polypeptide expression applied via, for example liposome, lipofection, particles, microvesicles as explained herein, may be administered sequentially or simultaneously. Similarly, self-inactivation may be used for inactivation of one or more guide sequence RNA used to target one or more targets.

Viral Vectors

[0632] The cargos may be delivered by viruses. In one embodiment, viral vectors are used. A viral vector may comprise virally-derived DNA or RNA sequences 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. Viruses and viral vectors may be used for *in vitro*, *ex vivo*, and/or *in vivo* deliveries.

Vector Regulatory Elements

[0633] A vector may comprise one or more regulatory elements. The regulatory element(s) may be operably linked to coding sequences of Cas12b polypeptide, accessory proteins, guide and/or scaffold sequence RNA or combination thereof. The term “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). In certain examples, a vector may comprise: a first regulatory element operably linked to

a nucleotide sequence encoding a Cas12b polypeptide, and a second regulatory element operably linked to a nucleotide sequence encoding a guide sequence RNA.

[0634] Examples of regulatory elements 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.

[0635] Examples of promoters include 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 H1 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), the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter.

Adeno-Associated Viruses (AAV)

[0636] The systems and compositions herein may be delivered by adeno associated virus (AAV). AAV vectors may be used for such delivery. AAV, of the Dependovirus genus and Parvoviridae family, is a single stranded DNA virus. In one embodiment, AAV may provide a persistent source of the provided DNA, as AAV delivered genomic material can exist indefinitely in cells, e.g., either as exogenous DNA or, with some modification, be directly integrated into the host DNA. In one embodiment, AAV do not cause or relate with any diseases

in humans. The virus itself is able to efficiently infect cells while provoking little to no innate or adaptive immune response or associated toxicity.

[0637] Examples of AAV that can be used herein include AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, and AAV-9. The type of AAV may be selected 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. AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium. Examples of cell types targeted by AAV are described in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)), and shown in Table 1 as follows:

Table 1. Examples of cell types targeted by AAV.

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

[0638] The AAV particles may be created in HEK 293 T cells. Once particles with specific tropism have been created, they are used to infect the target cell line much in the same way that native viral particles do. This may allow for persistent presence of the components in the infected cell type, and what makes this version of delivery particularly suited to cases where long-term expression is desirable. Examples of doses and formulations for AAV that can be used include those describe in US Patent Nos. 8,454,972 and 8,404,658.

[0639] Various strategies may be used for delivery the systems and compositions herein with AAVs. In some examples, coding sequences of Cas12b polypeptide nuclease and ω RNA may be packaged directly onto one DNA plasmid vector and delivered via one AAV particle. In some examples, AAVs may be used to deliver guide sequence RNA into cells that have been previously engineered to express Cas12b polypeptide. In some examples, coding sequences of Cas12b polypeptide and guide sequence RNA may be made into two separate AAV particles, which are used for co-transfection of target cells. In some examples, markers, tags, and other sequences may be packaged in the same AAV particles as coding sequences of Cas12b polypeptide and/or guide sequence RNAs.

Lentiviruses

[0640] The systems and compositions herein may be delivered by lentiviruses. Lentiviral vectors may be used for such delivery. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells.

[0641] Examples of lentiviruses include human immunodeficiency virus (HIV), which may use its envelope glycoproteins of other viruses to target a broad range of cell types; minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV), which may be used for ocular therapies. In an 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 herein.

[0642] Lentiviruses may be pseudo-typed with other viral proteins, such as the G protein of vesicular stomatitis virus. In doing so, the cellular tropism of the lentiviruses can be altered to be as broad or narrow as desired. In some cases, to improve safety, second- and third-generation lentiviral systems may split essential genes across three plasmids, which may reduce the likelihood of accidental reconstitution of viable viral particles within cells.

[0643] In some examples, leveraging the integration ability, lentiviruses may be used to create libraries of cells comprising various genetic modifications, e.g., for screening and/or studying genes and signaling pathways.

Adenoviruses

[0644] The systems and compositions herein may be delivered by adenoviruses. Adenoviral vectors may be used for such delivery. Adenoviruses include nonenveloped viruses with an icosahedral nucleocapsid containing a double stranded DNA genome. Adenoviruses

may infect dividing and non-dividing cells. In one embodiment, adenoviruses do not integrate into the genome of host cells, which may be used for limiting off-target effects of systems in gene editing applications.

Viral vehicles for delivery into plants

[0645] The systems and compositions may be delivered to plant cells using viral vehicles. In one embodiment, the compositions and systems may be introduced in the plant cells using a plant viral vector (e.g., as described in Scholthof et al. 1996, *Annu Rev Phytopathol.* 1996;34:299-323). Such viral vector may be a vector from a DNA virus, e.g., 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). The viral vector may be a vector from an RNA virus, e.g., tobnavirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordevirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses may be non-integrative vectors.

Non-viral vehicles

[0646] The delivery vehicles may comprise non-viral vehicles. In general, methods and vehicles capable of delivering nucleic acids and/or proteins may be used for delivering the systems compositions herein. Examples of non-viral vehicles include lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, gold nanoparticles, streptolysin O, multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Lipid particles

[0647] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes.

Lipid nanoparticles (LNPs)

[0648] LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes), and may be delivered to cells with relative ease. In some examples, lipid nanoparticles do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro*, *ex vivo*, and *in vivo* deliveries. Lipid particles may be used for various scales of cell populations.

[0649] In some examples, LNPs may be used for delivering DNA molecules (e.g., those comprising coding sequences of Cas12b polypeptide and/or guide sequence RNA) and/or RNA

molecules (e.g., mRNA of Cas12b polypeptide guide sequence RNAs). In certain cases, LNPs may be used for delivering RNP complexes of Cas12b polypeptide or guide sequence RNA.

[0650] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoley-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxylpropyl-3-amine (PEG-C-DMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011).

Liposomes

[0651] In one embodiment, a lipid particle may be liposome. 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. In one embodiment, liposomes 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).

[0652] Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as 1,2-distearoyl-sn-glycero-3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

[0653] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin, and/or 1,2-dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

Exosomes

[0654] The delivery vehicles may comprise exosomes. Exosomes include membrane bound extracellular vesicles, which can be used to contain and delivery various types of biomolecules, such as proteins, carbohydrates, lipids, and nucleic acids, and complexes thereof (e.g., RNPs). Examples of exosomes include those described in Schroeder A, et al., J Intern Med. 2010 Jan;267(1):9-21; El-Andaloussi S, et al., Nat Protoc. 2012 Dec;7(12):2112-26; Uno

Y, et al., Hum Gene Ther. 2011 Jun;22(6):711-9; Zou W, et al., Hum Gene Ther. 2011 Apr;22(4):465-75.

[0655] In some examples, the exosome may form a complex (e.g., by binding directly or indirectly) to one or more components of the cargo. In certain examples, a molecule of an exosome may be fused with first adapter protein and a component of the cargo may be fused with a second adapter protein. The first and the second adapter protein may specifically bind each other, thus associating the cargo with the exosome. Examples of such exosomes include those described in Ye Y, et al., Biomater Sci. 2020 Apr 28. doi: 10.1039/d0bm00427h.

Genetically-Modified Cells and Organisms

[0656] The present disclosure further provides cells comprising one or more components of the compositions and systems herein, e.g., the Cas12b polypeptide and/or guide sequence RNAs. Also provided include cells modified by the systems and methods herein, and cell cultures, tissues, organs, organism comprising such cells or progeny thereof. In one embodiment, the present disclosure provides 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 be a therapeutic T cell or antibody-producing B-cell. 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.

[0657] In one embodiment, one or more polynucleotide molecules, vectors, or vector systems driving expression of one or more elements of the compositions, systems, or delivery systems comprising one or more elements of the 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. In an embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0658] In one embodiment, the host cell is a cell of a cell line. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In one embodiment, 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 one embodiment, a cell transiently transfected with the components of a 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 complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In one embodiment, 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.

[0659] Further intended are isolated human cells or tissues, plants or non-human animals comprising one or more of the polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein. In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including (isolated) stem cells, and progeny thereof.

[0660] In an embodiment, the plants or non-human animals comprise at least one of the system components, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein at least one tissue type of the plant or non-human animal. In an embodiment, non-human animals comprise at least one of the system components, polynucleotide molecules, vectors, vector systems, or cells described in any of the embodiments herein in at least one tissue type. In an embodiment, the presence of the system components is transient, in that they are degraded over time. In an embodiment, expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is limited to certain tissue types or regions in the plant or non-human animal. In an embodiment, the expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent of a physiological cue. In an embodiment, expression of the components of the systems and compositions described in any of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells may be triggered by an exogenous molecule. In an embodiment, expression of the components of the systems and compositions described in any

of the embodiments comprised in polynucleotide molecules, vectors, vector systems, or cells is dependent on the expression of a non-cas molecule in the plant or non-human animal.

[0661] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLES

Example 1 – Identification of mini-Cas12 polypeptides and orthologs thereof

[0662] An *in silico* database search was performed using Cas12 (e.g., Cpf1, C2c1) polypeptide sequences, which lack an HNH domain, with the aim of identifying smaller Cas12 polypeptides.

[0663] Fig. 1 illustrates the phylogenetic relationships of the Cas12b orthologs identified in the database search. The genomic architecture of Cas12b from *Planctomycetes bacterium* RBG 13-46-10 and *Phycisphaera bacterium* ST-NAGAB-D1 are shown. The rectangle surrounds the accession numbers of the two Cas12b orthologs used in this study for *Planctomycetes bacterium* RBG 13-46-10 and *Phycisphaera bacterium* ST-NAGAB-D1.

[0664] Fig. 2 illustrates the ancestral Cas12b loci for *Planctomycetes bacterium* RBG 13-46-10 and *Phycisphaera bacterium* ST-NAGAB-D1. The Cas12b amino acid lengths are 743 amino acids for *Planctomycetes bacterium* RBG 13-46-10 and 747 amino acids *Phycisphaera bacterium* ST-NAGAB-D1. The small RNA-seq data of the tracrRNA region and array for *Phycisphaera bacterium* ST-NAGAB-D1 Cas12b is shown.

[0665] Fig. 3A-B illustrates the PAM depletion vs. rank for *Phycisphaera bacterium* ST-NAGAB-D1 Cas12b and a Weblogo at >0.9 depletion (43 PAMs) and >0.8 depletion (195 PAMs)(Fig. 3A) and PAM depletion vs. rank for *Planctomycetes bacterium* RBG 13 46 10 Cas12b and a Weblogo graph at >0.9 depletion and >0.8 depletion (Fig. 3B). It was determined that the PAM preference is YANTTN for both species, where Y is T or C and N is any nucleotide.

[0666] Fig. 4A-B shows that the *Phycisphaera bacterium* ST-NAGAB-D1 (Fig. 4A) and *Planctomycetes bacterium* RBG 13-46-10 (Fig. 4B) Cas12b systems are functional in an *E. coli* expression system.

[0667] FIG. 5 shows *in vitro* activity of the RNP *Phycisphaera bacterium* ST-NAGAB-D1 complex with pulldown of the RNA complex using different buffer systems. The results show the expected pulldown products at 320 bp and 250 bp.

[0668] FIG. 6 shows that the *Phycisphaera bacterium*/RNA binary complex is more temperature stable than the apo-protein (Cas12b only) based on fluorescence absorption of denatured complexes at 580 nm. The binary complex showed a temperature maximum at about 48°C whereas the apo-protein showed a temperature maximum at about 34°C. The RNase added control gave a temperature maximum mid-way (~39°C) between the two experimental samples.

[0669] FIG. 7 illustrates an RNA-seq and mapping analysis of *Phycisphaera bacterium* tracrRNA. It was experimentally determined by generating a deletion series of tracrRNA that a long functional tracrRNA (~190 nt) and sgRNA is required for activity in *E. coli*.

[0670] FIG. 8 shows *Phycisphaera bacterium in vitro* loading and activity of Cas12b (C2c1~90 kDa; at left) and that dsDNA cleavage activity is dependent on Cas12b, that a sgRNA consisting of a 194 nt sgRNA without spacer is required and has a preference at 48°C (middle). Sequencing revealed a TTTA PAM, a 25 nt protospacer and a potential 13 nt staggered overhang (at right). A predominantly nicked product was observed at 37°C (possible non-target strand cutting) (middle, right-most column).

[0671] FIG. 9 illustrates that *Phycisphaera bacterium* Cas12b lacks activity in HEK293T cells targeting the DNMT-1 and VEGFA loci.

[0672] FIG. 10 illustrates the *Planctomycetes bacterium* RBG 13 46 10 tracrRNA region with accompanying direct repeat and Fn Spacer regions (top). The *Phycisphaera bacterium* sgRNA is aligned with the *Planctomycetes bacterium* sgRNA (bottom) for comparison.

[0673] FIG. 11A-11B – illustrates *Planctomycetes bacterium in vitro* loading and activity. Fig. 11A shows very weak *in vitro* activity for *Planctomycetes bacterium* Cas12b and also shows that it requires a 189 nt tracrRNA in addition to crRNA. Fig. 11B shows no RNP pulldown at 48°C and slight pulldown activity at 37°C, suggesting that the *Planctomycetes bacterium* Cas12b might prefer lower temperatures for increased activity.

[0674] 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 hereinbefore set forth.

CLAIMS

What is claimed is:

1. A non-naturally occurring or engineered composition comprising:
 - (a) a Cas polypeptide that comprises a RuvC-I, -II, -III domain but does not comprise an HNH domain and is less than 850 amino acids in size; and
 - (b) one or more nucleic acid components comprising a guide molecule capable of forming a complex with the Cas polypeptide and directing sequence-specific binding of the complex to bind to a target sequence on a target polynucleotide.
2. The composition of claim 1, wherein the guide molecule comprises a scaffold sequence between about 170 nt and about 210 nt in length.
3. The composition of claim 1, wherein the Cas polypeptide is derived from *Phycisphaerae* bacterium ST-NAGAB-D1 or *Planctomycetes* bacterium RBG 134610.
4. The composition of claim 1, wherein the guide molecule is derived from *Phycisphaerae* bacterium ST-NAGAB-D1 or *Planctomycetes* bacterium RBG 134610.
5. The composition of claim 1, wherein the complex recognizes a PAM sequence comprising YANTTN, where Y is C or T, and N is any nucleotide.
6. The composition of claim 3, wherein the *Phycisphaerae* bacterium ST-NAGAB-D1 complex is stable and active between about 37°C to about 60°C.
7. A vector system comprising one or more polynucleotide sequences encoding the Cas and guide molecule of anyone of the preceding claims.
8. A delivery system comprising the composition of any one of claims 1 to 7.
9. The delivery system of claim 8, wherein the delivery system comprises a ribonucleoprotein complex, one or more particles, one or more vesicles, or one or more

liposomes, nanoparticles, exosomes, microvesicles, nucleic acid nanoassemblies, a gene gun, an implantable device.

10. A cell comprising the composition of any of the preceding claims, or progeny thereof.

11. An *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising the composition of any one of claims 1 to 9.

12. A non-naturally occurring or engineered composition comprising:

- a. a Cas polypeptide that comprises a RuvC-I, -II, -III domain but does not comprise an HNH domain and is less than 850 amino acids in size, wherein the Cas protein is catalytically inactive;
- b. a nucleotide deaminase associated with or otherwise capable of forming a complex with the Cas protein; and,
- c. a guide sequence capable of forming a complex with the Cas protein and directing the complex to bind to a target sequence.

13. The composition of claim 12, wherein the nucleotide deaminase is an adenosine deaminase or a cytidine deaminase.

14. One or more polynucleotides encoding one or more components of the composition of any one of claims 12 or 13.

15. One or more vectors encoding the one or more polynucleotides of claim 14.

16. A cell or progeny thereof genetically engineered to express one or more components of the composition of any one of claims 14 or 15.

17. A method of editing nucleic acids in target polynucleotides comprising delivering the composition of claim 12 or claim 13, the one or more polynucleotides of claim 14, or one or more vectors of claim 18 to a cell or population of cells comprising the target polynucleotides.

18. The method of claim 17, wherein the target polynucleotide is genomic DNA.
19. The method of claim 17 or 18, wherein the target polynucleotide is edited at one or more bases to introduce a G→A or C→T mutation.
20. An isolated cell or progeny thereof comprising one or more base edits made using the method of any one of claims 18 or 19.
21. An engineered, non-naturally occurring composition comprising:
- a. a Cas polypeptide that comprises a RuvC-I, -II, -III domain but does not comprise a HNH domain and is less than 850 amino acids in size, wherein the Cas protein is catalytically inactive,
 - b. a reverse transcriptase associated with or otherwise capable of forming a complex with the Cas12b polypeptide, and
 - c. a scaffold sequence between about 170 nt and about 210 nt in length capable of forming a complex with the Cas12b protein and directing site-specific binding of the complex to a target sequence of a target polynucleotide, the scaffold sequence further comprising a donor template encoding a donor sequence for insertion into the target polynucleotide.
22. One or more polynucleotides encoding one or more components of the composition of claim 21.
23. One or more vectors encoding the one or more polynucleotides of claim 22.
24. A method of modifying target polynucleotides comprising:
delivering the composition of claim 21, the one or more polynucleotides of claim 22, or the one or more vectors of claim 23 to a cell, or population of cells, comprising the target polynucleotides, wherein the complex directs the reverse transcriptase to the target sequence and the reverse transcriptase facilitates insertion of a donor sequence encoded by the donor template from the scaffold sequence into the target polynucleotide.
25. The method of claim 24, wherein insertion of the donor sequence:

- a. introduces one or more base edits;
- b. corrects or introduces a premature stop codon;
- c. disrupts a splice site;
- d. inserts or restores a splice site;
- e. inserts a gene or gene fragment at one or both alleles of the target polynucleotide; or;
- f. a combination thereof.

26. An isolated cell or progeny thereof comprising the modifications made using the method of claim 24 or 25.

27. An engineered, non-naturally occurring composition comprising:
- a. a Cas polypeptide that comprises a RuvC-I, -II, -III domain but does not comprise a HNH domain and is less than 850 amino acids in size, wherein the Cas protein is catalytically inactive,
 - b. a non-LTR retrotransposon polypeptide associated with or otherwise capable of forming a complex with the TnpB polypeptide, and
 - c. a nucleic acid component capable of forming a complex with the Cas polypeptide and directing site-specific binding of the complex to a target sequence of a target polynucleotide, the nucleic acid component further comprising a donor template encoding a donor sequence for insertion into the target polynucleotide and located between two binding elements capable of forming a complex with the non-LTR retrotransposon polypeptide.

28. The composition of claim 27, wherein the Cas polypeptide is fused to the N-terminus of the non-LTR retrotransposon polypeptide.

29. The composition of claim 27 or 28, wherein the Cas polypeptide is engineered to have nickase activity.

30. The composition of claim 29, wherein the nucleic acid component directs the Cas polypeptide to a target sequence 5' of the targeted insertion site, and wherein the Cas polypeptide generates a strand break at the targeted site of insertion.

31. The composition of claim 29, wherein the nucleic acid component directs the Cas polypeptide to a target sequence 3' of the targeted insertion site, and wherein the Cas polypeptide generates a strand break at the targeted insertion site.

32. The composition of claim 29, wherein the donor polynucleotide further comprises a polymerase processing element of facilitate 3' end processing of the donor polynucleotide sequence.

33. The composition of claim 29, wherein the donor polynucleotide further comprises a homology region to the target sequence on the 5' end of the donor construct, the 3' end of the donor construct, or both.

34. The composition of claim 33, wherein the homology region is from 8 to 25 base pairs.

35. One or more polynucleotide encoding one or more component of the composition of anyone of claims 29 to 34.

36. One or more vectors comprising the one or more polynucleotides of claim 35

37. A method of modifying target polynucleotides, comprising:
delivering the composition of any one of claims 29 to 34, the one or more polynucleotides of claim 35, or one or more vectors of claim 36 to a cell or population of cells comprising the target polynucleotides, wherein the complex directs the non-LTR retrotransposon protein to the target sequence and the non-LTR retrotransposon protein facilitates insertion of the donor polynucleotide sequence from the donor construct into the target polynucleotide.

38. The method of claim 37, wherein the insertion of the donor sequence:

- a. introduces one or more base edits;
- b. corrects or introduces a premature stop codon;
- c. disrupts a splice site;

- d. inserts or restores a splice site;
- e. inserts a gene or gene fragment at one or both alleles of the target polynucleotide; or;
- f. a combination thereof.

39. An isolated cell or progeny thereof, comprising the modifications made using the methods of claims 37 or 38.

40. A composition for detecting the presence of a target polynucleotide in a sample, comprising:

one or more Cas polypeptides comprising a split RuvC nuclease domain, but no HNH nuclease domain, less than 850 amino acids in size, and possessing collateral activity;

at least one nucleic acid component comprising a sequence capable of binding a target polynucleotide and designed to form a complex with the one or more Cas polypeptides;

a detection construct comprising a polynucleotide component, wherein the Cas polypeptides exhibits collateral nuclease activity and cleaves the polynucleotide component of the detection construct once activated by the target sequence; and

optionally, isothermal amplification reagents.

41. The composition of claim 40, wherein the isothermal amplification reagents are loop-mediated isothermal amplification (LAMP) reagents.

42. The composition of claim 41, wherein the LAMP reagents comprise LAMP primers.

43. The composition of any one of the claims 40 to 42, further comprising one or more additives to increase reaction specificity or kinetics.

44. The composition of any one of claims 40 to 43, further comprising polynucleotide binding beads.

45. A method for detecting polynucleotides in a sample, the method comprising:

contacting one or more target sequences with the composition of anyone of claims 40 to 44, wherein the Cas polypeptide exhibits collateral nuclease activity and cleaves the detection construction once activated by the one or more target sequences; and

detecting a signal from cleavage of the detection construction thereby detecting the one or more target polynucleotides.

46. The method of claim 45, further comprising amplifying the target polynucleotides using isothermal amplification prior to the contacting step.

1/14

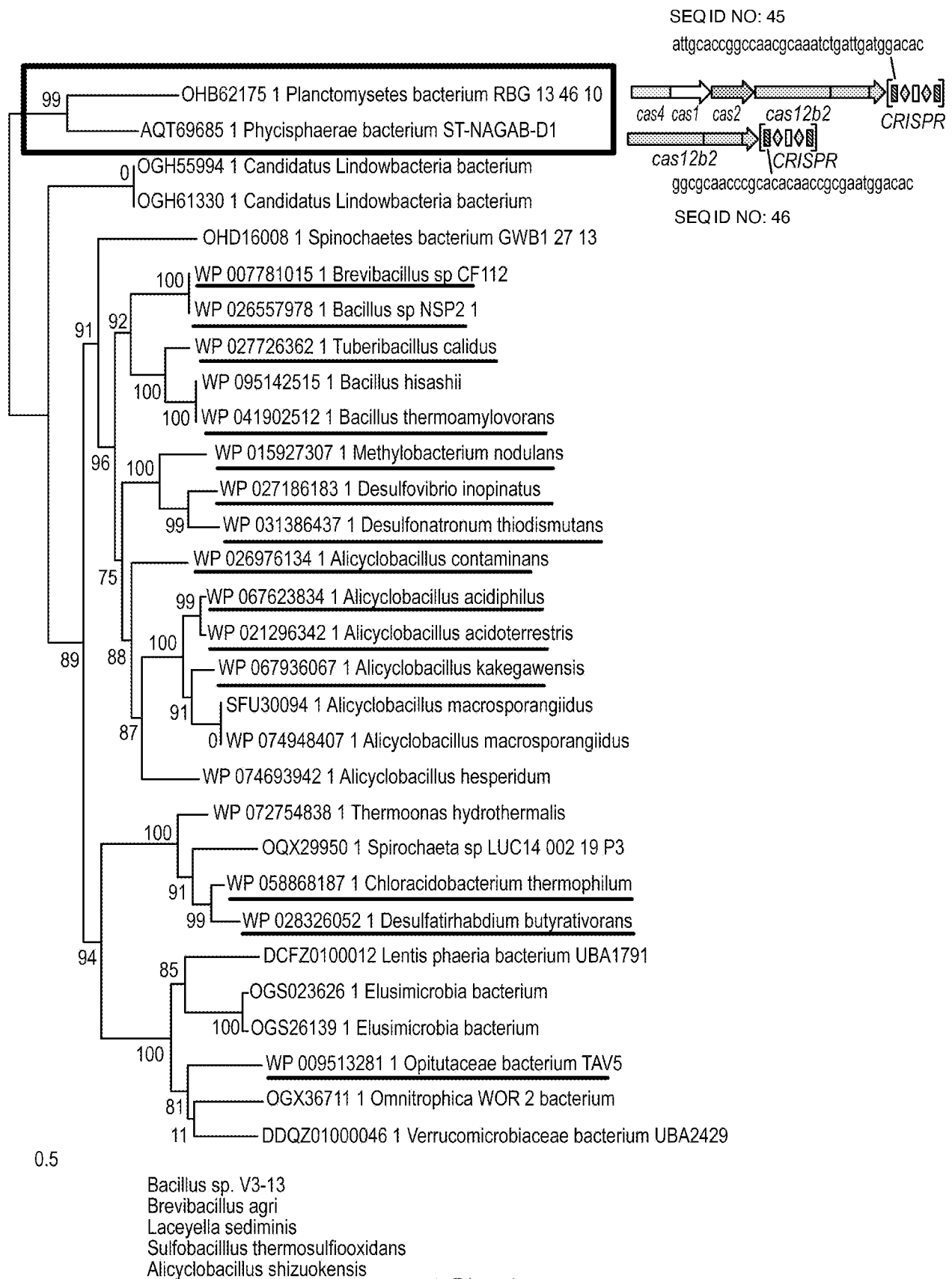


FIG. 1

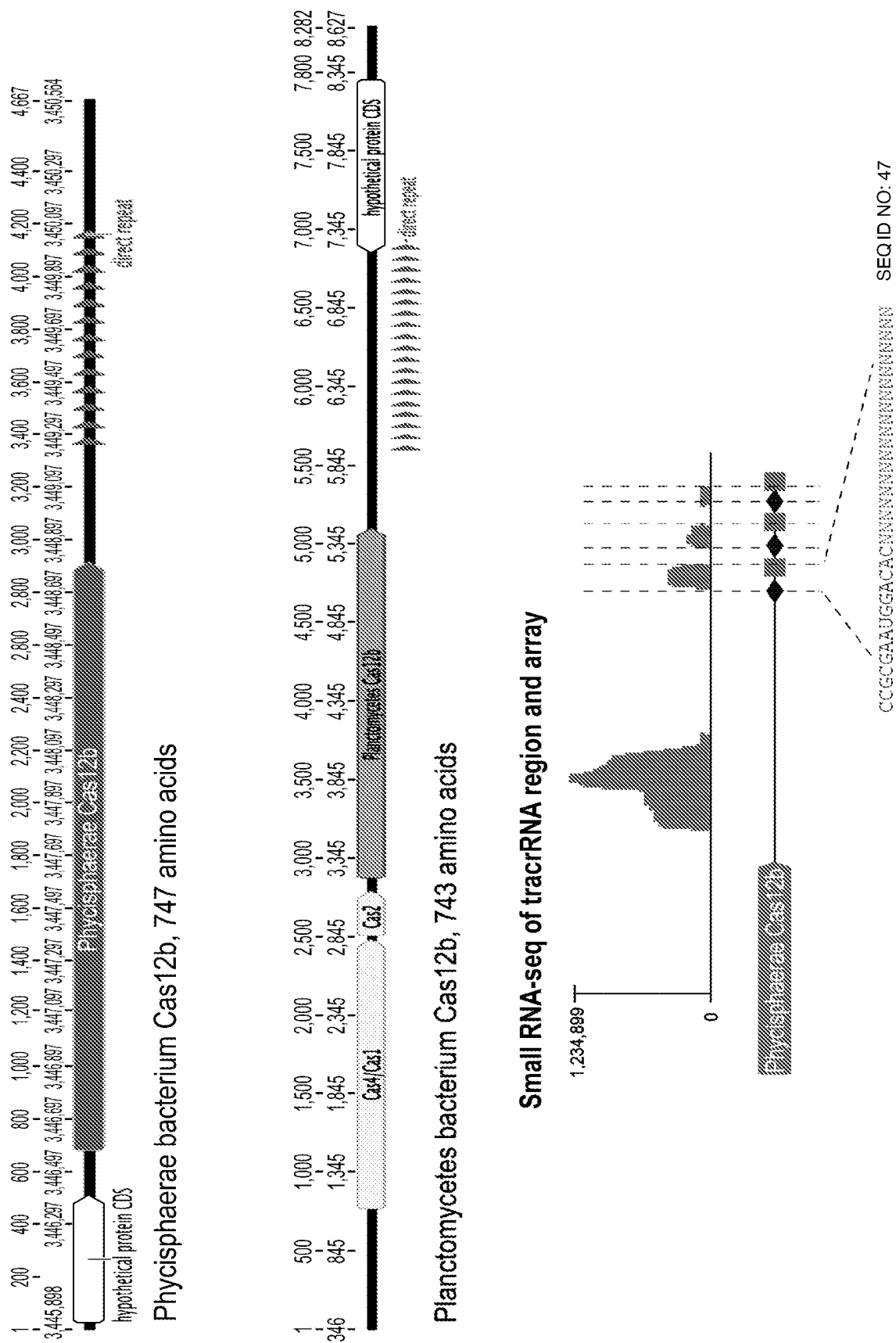


FIG. 2

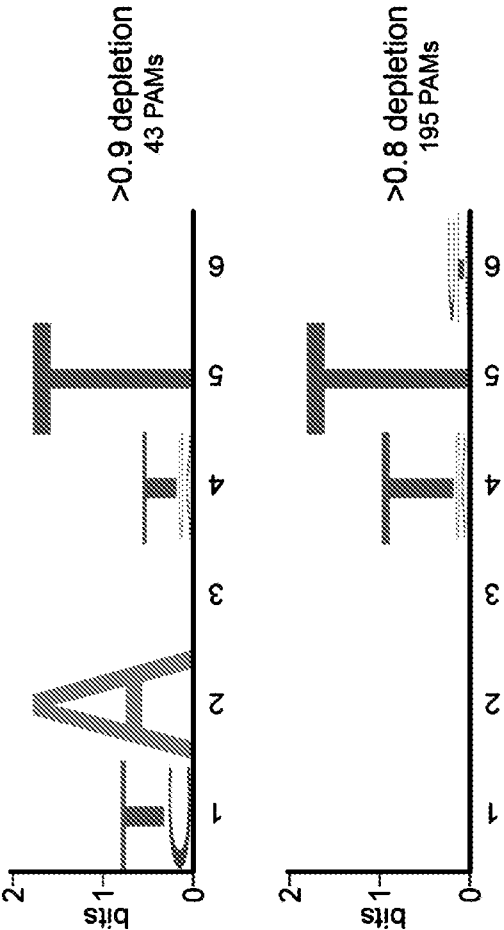


FIG. 3A

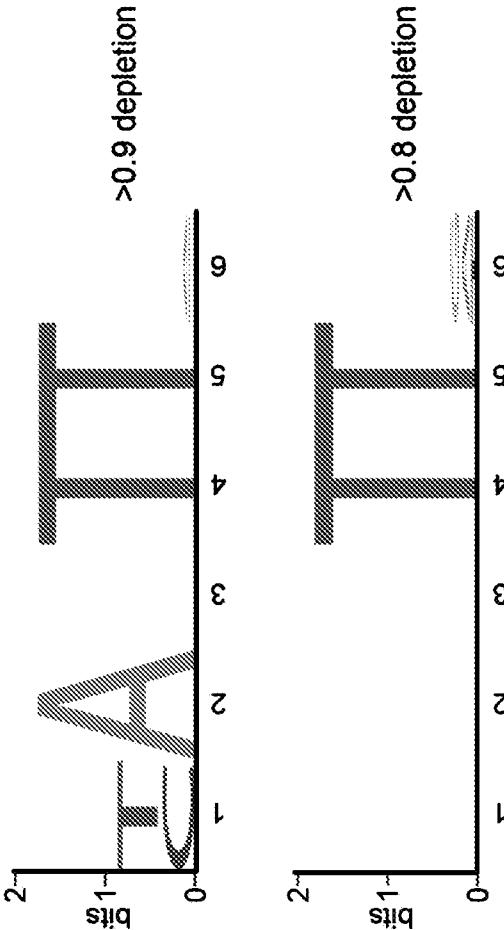


FIG. 3B

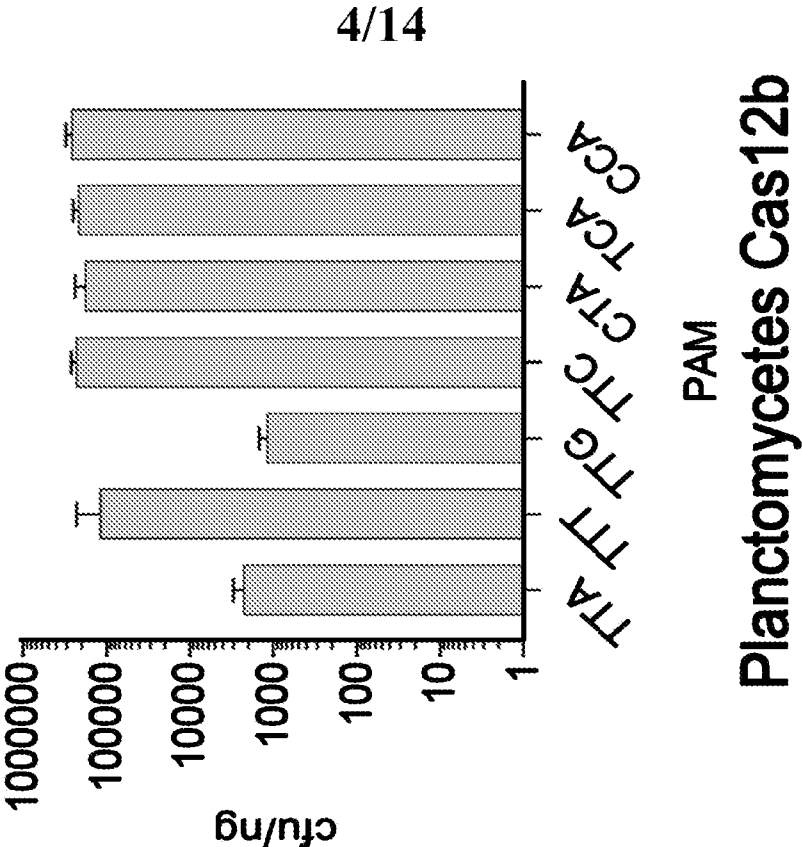


FIG. 4B

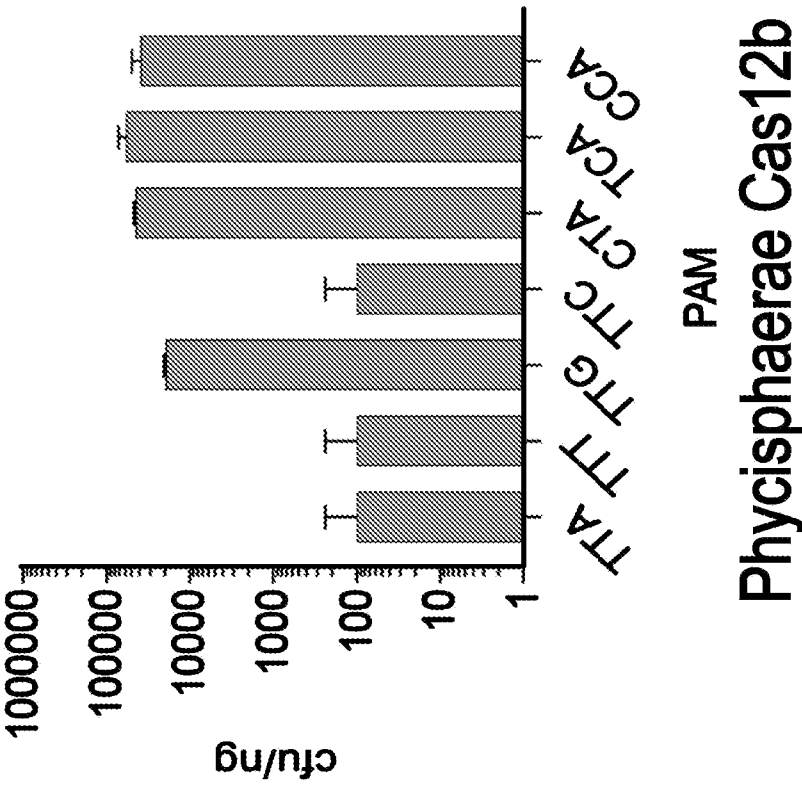


FIG. 4A

5/14

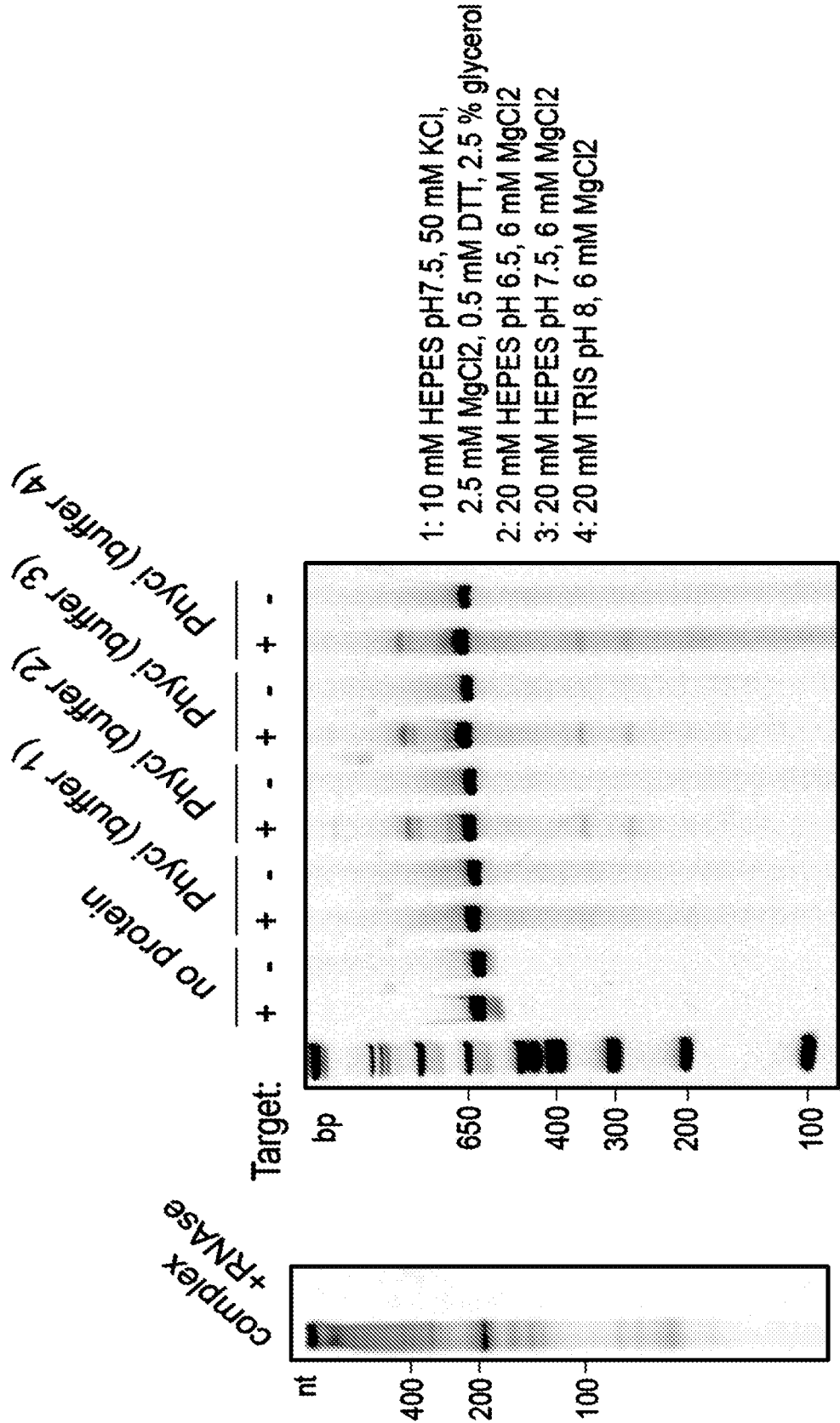


FIG. 5

6/14

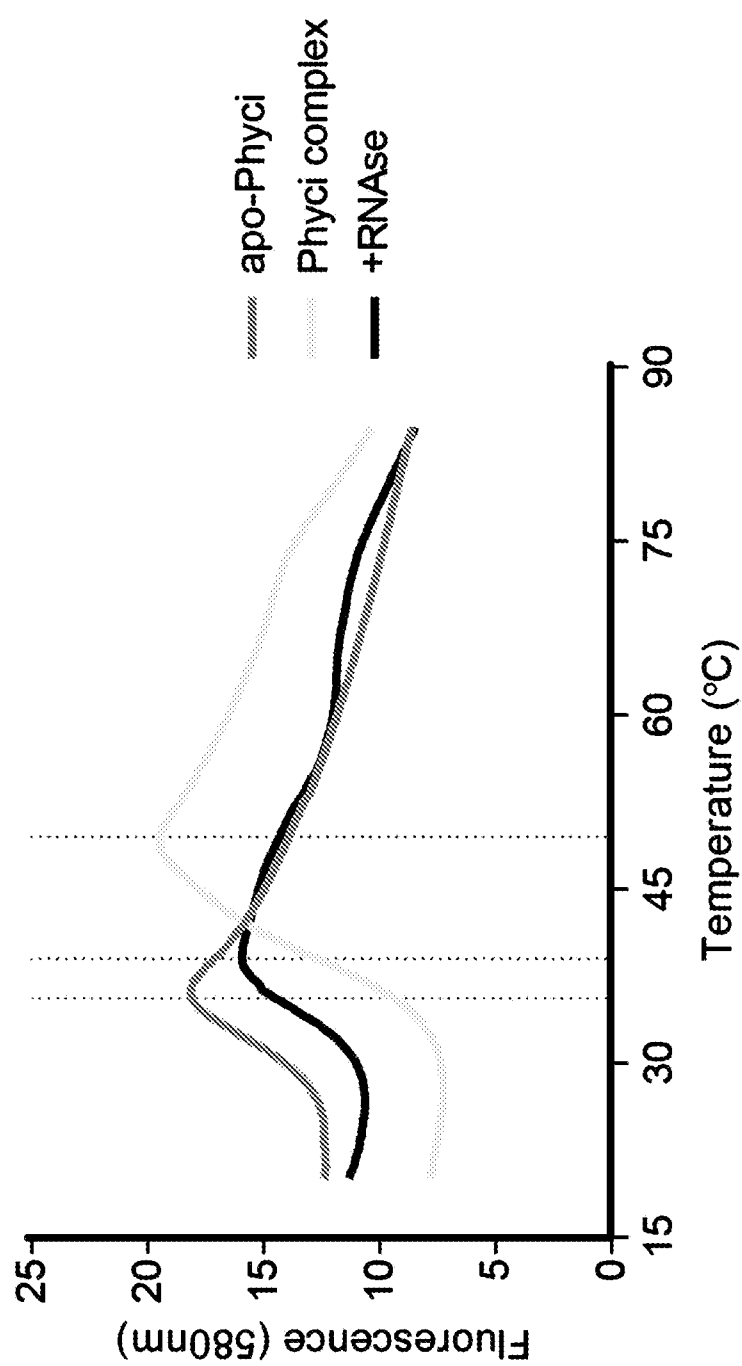


FIG. 6

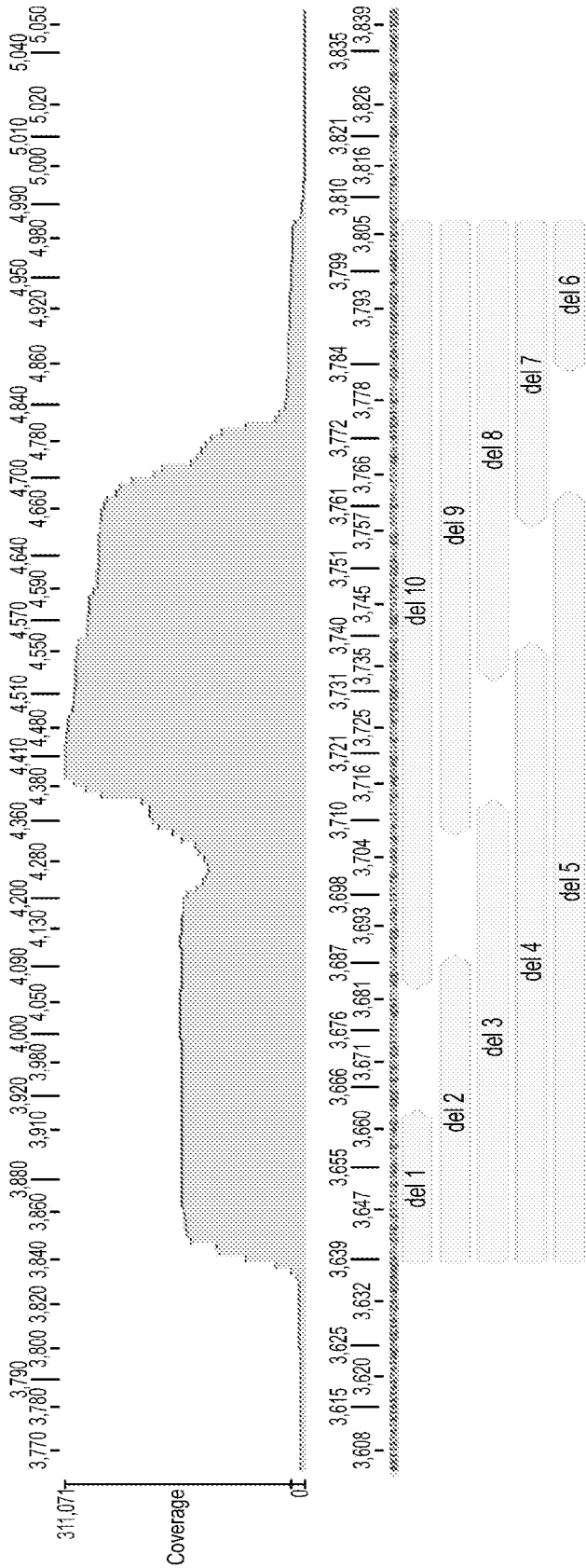


FIG. 7

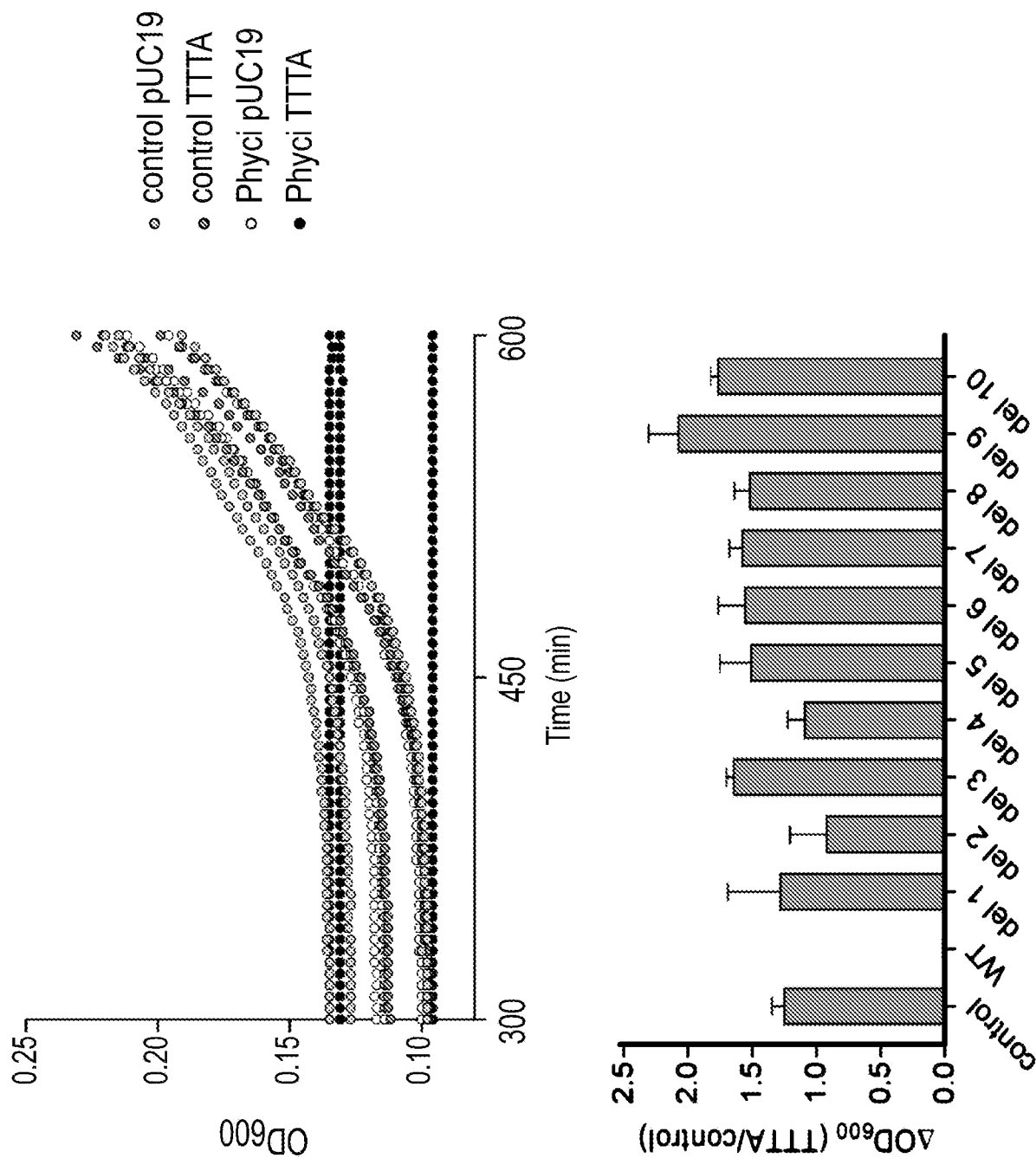


FIG. 7 Continued

9/14

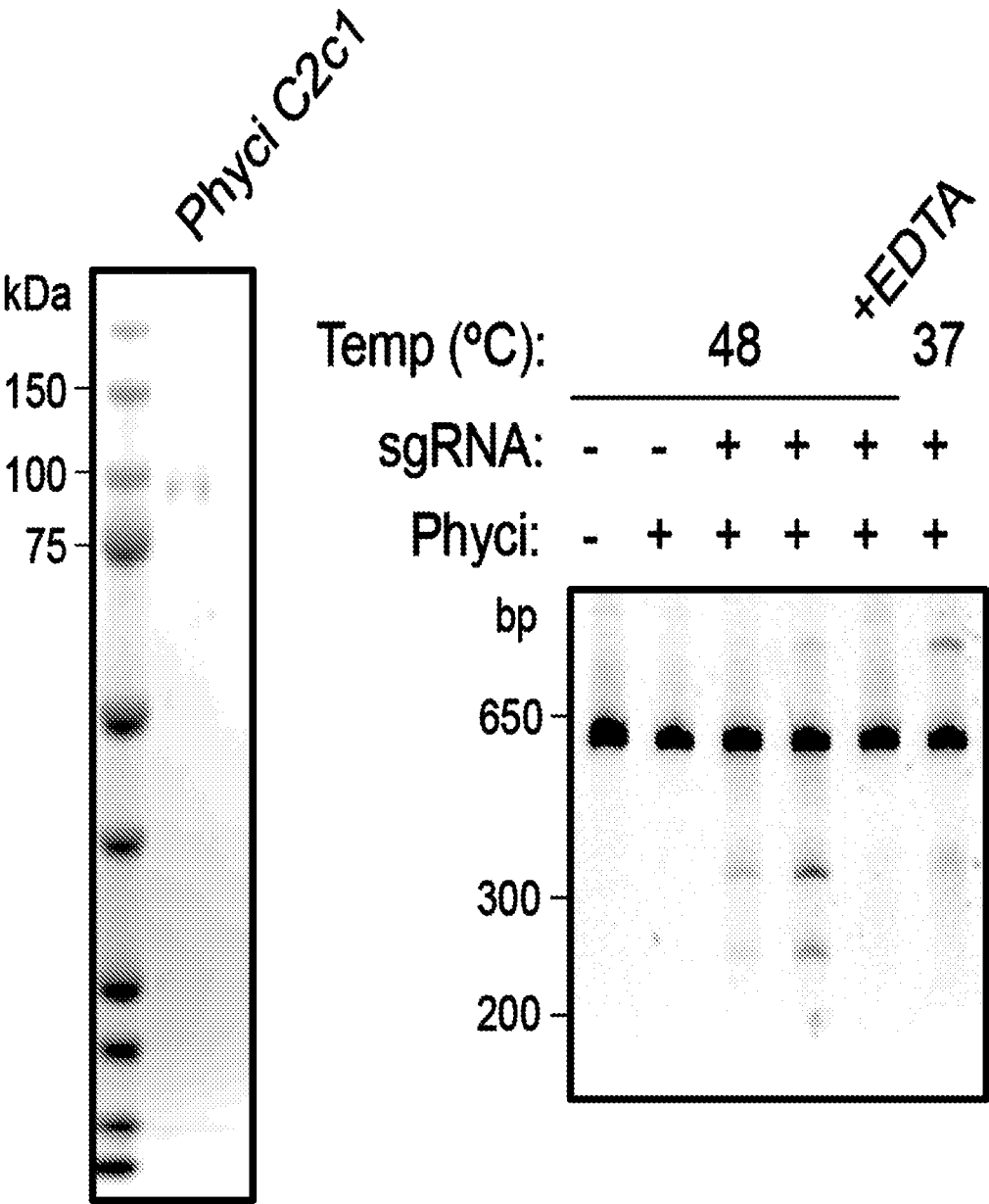


FIG. 8

PAM

25 nt protospacer

CTTAGAGAGTCATTAAATAAGGCCACTGTTAAAAGCTTGGCGTA
|||||
GAAATCTCTTCAGTAAATATTCGGTGACAATTTTCGAACCGCAT

SEQ ID NO: 48

.....

GAAATCTCTTCAGTAAATTAATCCGGTGACAATTTTGAACCGCAT

SEQID NO: 49

10/14

reverse primer

forward
primer

FIG. 8 Continued

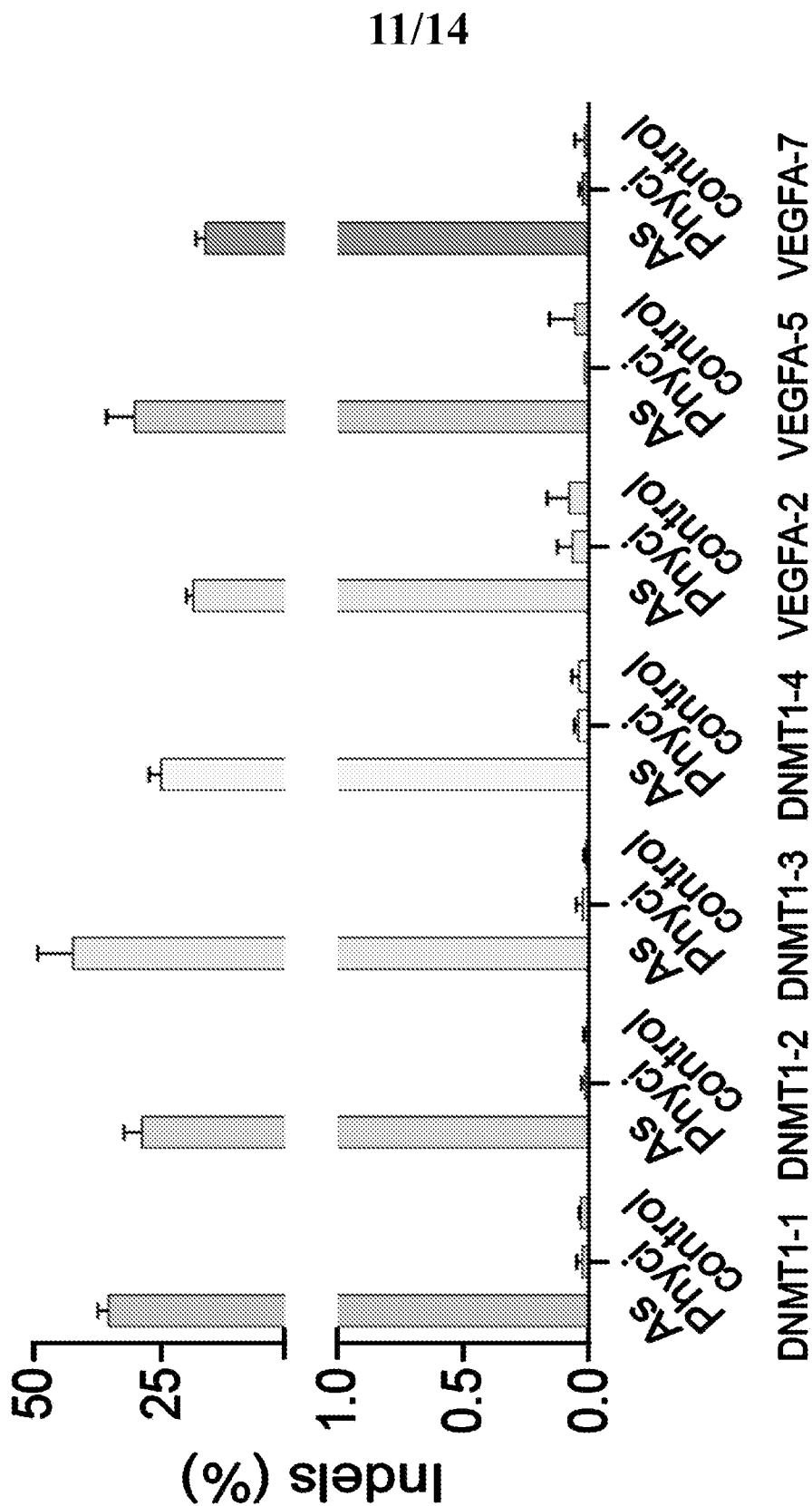


FIG. 9

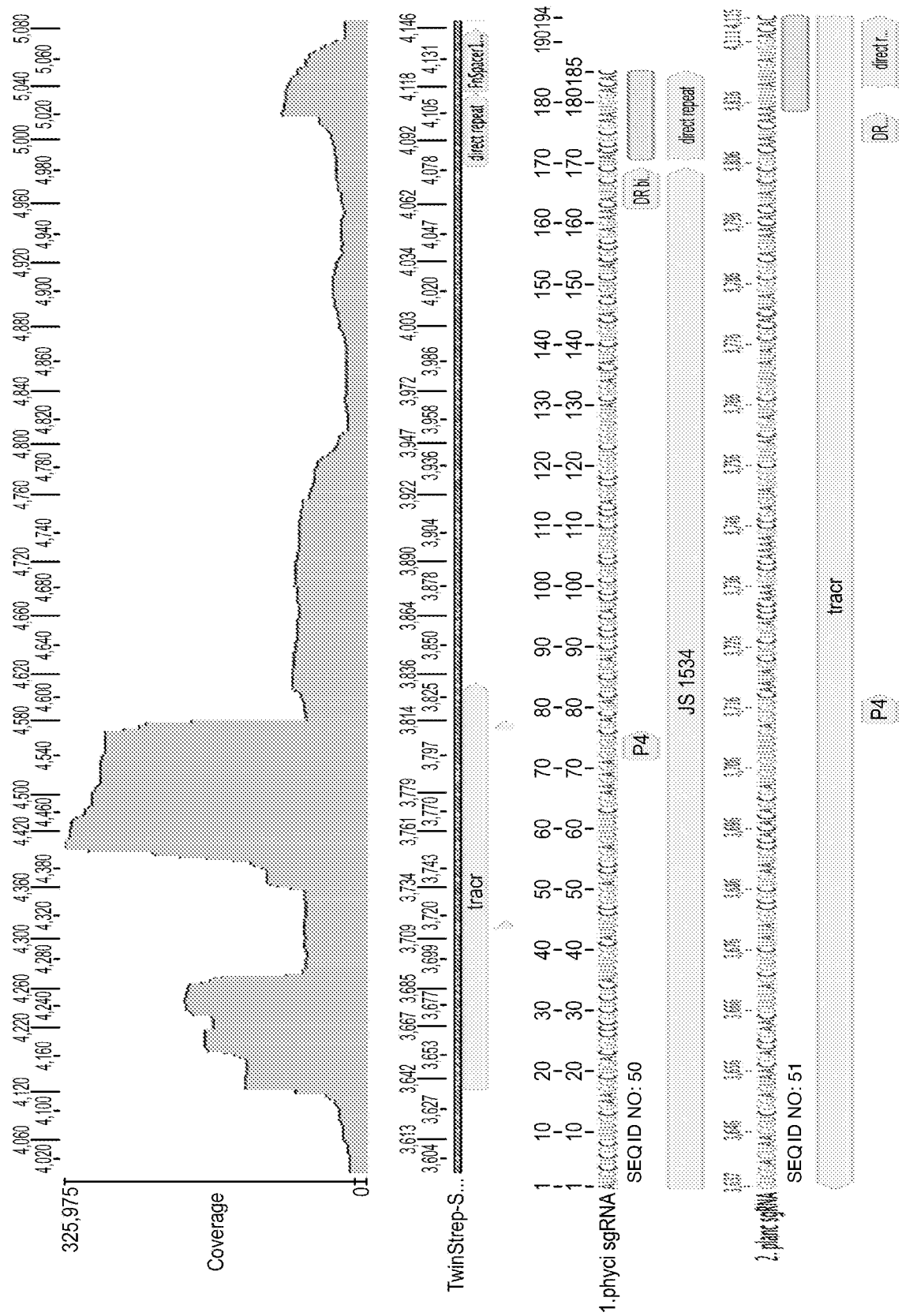


FIG. 10

13/14

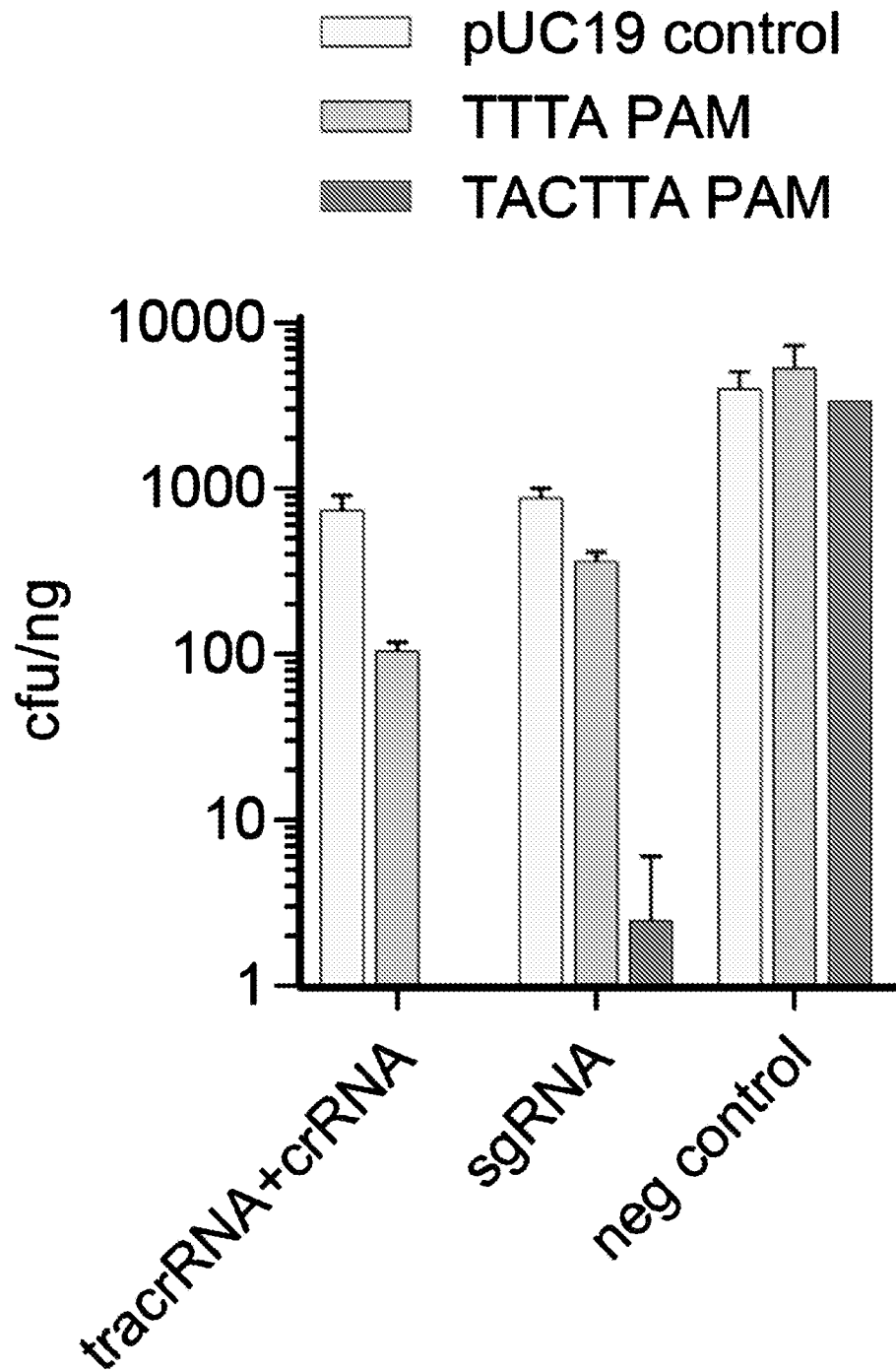
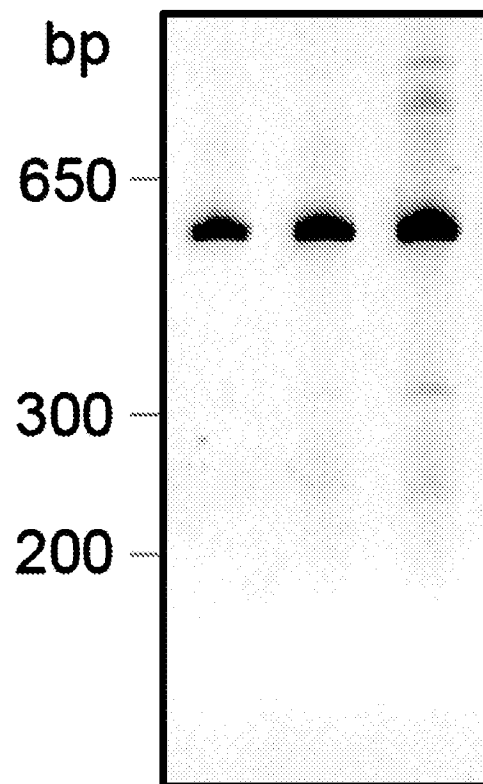


FIG. 11A

14/14

Temp (°C): 48 48 37
sgRNA: - + +
Planc: + + +

**FIG. 11B**