

(51) International Patent Classification:

A61K 48/00 (2006.01) CI2N 15/90 (2006.01)
C07K 14/195 (2006.01) CI2N 9/22 (2006.01)
CI2N 15/52 (2006.01)

(21) International Application Number:

PCT/US2024/022895

(22) International Filing Date:

03 April 2024 (03.04.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/493,942 03 April 2023 (03.04.2023) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **CYRUS BIOTECHNOLOGY** [US/US]; 3005 1st Avenue, Suite 203, Seattle, Washington 98121 (US).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **RAGHAVAN, Rumya**; c/o 77 Massachusetts Avenue, Cambridge, Massa-

chusetts 02139 (US). **MACRAE, Rhiannon**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **SONG, Yifan**; c/o 3005 1st Avenue, Suite 203, Seattle, Washington 98121 (US). **NIVON, Lucas**; c/o 3005 1st Avenue, Suite 203, Seattle, Washington 98121 (US). **KING, Indigo**; c/o 3005 1st Avenue, Suite 203, Seattle, Washington 98121 (US).

(74) Agent: **SOUTHGATE, Christopher D.** et al.; Johnson, Marcou, Isaacs & Nix LLC, PO Box 691, Hoschton, Georgia 30548 (US).

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

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

(54) Title: ENGINEERED TYPE V CAS POLYNUCLEOTIDES WITH REDUCED IMMUNOGENICITY AND USES THEREOF

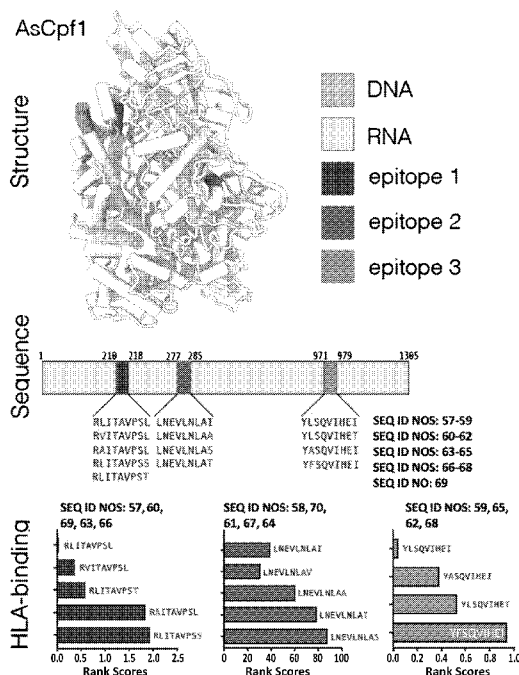


FIG. 1

(57) Abstract: Engineered Type V Cas polypeptides with reduced immunogenicity, CRISPR-Cas systems thereof, compositions thereof, delivery systems thereof, and methods of use thereof for modifying target polynucleotides, such as, for example, in cells.

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

Published:

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

ENGINEERED TYPE V CAS POLYNUCLEOTIDES WITH REDUCED IMMUNOGENICITY AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/493,942, filed on April 3, 2023, the contents of which is incorporated by reference herein in its entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The contents of the electronic sequence listing ("BROD-5750WP_ST26.xml; size is 197,553 bytes and it was created on April 3, 2024) is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0003] The subject matter disclosed herein is generally directed to engineered Type V Cas polypeptides with reduced immunogenicity, CRISPR-Cas systems thereof, compositions thereof, delivery systems thereof, and methods of use thereof for modifying target polynucleotides, such as, for example, in cells.

BACKGROUND

[0004] The development of CRISPR-Cas endonucleases for eukaryotic genome editing has sparked intense interest in the use of this technology for therapeutic applications. Extensive research has led to the identification of different technologies which can address the challenges of safety and efficacy. There are still challenges to overcome in the development of CRISPR-based therapeutics, including immunogenicity of Type V Cas polypeptides (e.g., Cas12 proteins or orthologs or variants) in order to allow the translation of these genome editing technologies to the clinic, particularly, the deimmunization Type V Cas polypeptides without significantly altering activity, including but not limited to nucleic acid binding, complex formation, target binding, and target cleavage.

[0005] 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

[0006] In an aspect, the present disclosure provides an engineered Type V Cas having reduced immunogenicity compared to a corresponding wild-type Type V Cas. In certain example embodiments, the engineered Type V Cas comprises one or more modifications of one or more epitopes of the wild-Type V Cas. In certain example embodiments, the one or more modifications of one or more epitopes comprise one or more amino acid substitutions in the one or more epitopes. In certain example embodiments, the one or more epitopes correspond to one or more MHC Class I and/or MHC Class II binding sites in the wild-type Type V Cas. In certain example embodiments, the MHC Class I and/or MHC Class II binding sites are selected from HLA Class I binding sites, HLA Class II binding sites, or any combination thereof. In certain example embodiments, the Type V Cas is a Cas12. In certain example embodiments, the Cas12 is a Cas12a (Cpf1), a Cas12b (C2c1), a Cas12c (C2c3), a Cas12d, a Cas 12e, a Cas12f (Cas14), a Cas12k, or a variant thereof. In certain example embodiments, the Cas12 is a Cas12a. In certain example embodiments, the Cas12a is a *Lachnospiraceae bacterium* Cas12a (LbCas12a), an *Acidaminococcus* sp. (AsCas12a), a *Francisella novicida* Cas12a (FnCas12a), a *Moraxella bovoculi* Cas12a (MbCas12), an *Eubacterium rectale* Cas12a (MAD7), or a variant thereof. In certain example embodiments, the Cas12 is an *Acidaminococcus* sp. (AsCas12a), or a variant thereof.

[0007] In certain example embodiments, the one or more epitopes corresponding to one or more MHC Class I and/or MHC Class II binding sites are selected from AsCas12 residues 210-218, 227-285, and/or 971-979, or residues in another Type V Cas analogous thereto. In certain example embodiments, the one or more amino acid substitutions are at AsCas12 residues 211, 218, 277, 281, 285, 971, 972, and/or 979, or at residues in another Type V Cas analogous thereto. In certain example embodiments, the one or more amino acid substitutions comprise substitution of one or more original residues selected from I, L, Y, and any combination thereof, with one or more naturally occurring residues selected from A, D, E, F, G, I, K, M, N, P, Q, R, S, T, V, W, Y, and any combination thereof. In certain example embodiments, the one or more amino acid substitutions comprise: L211V, L211A, L218A, L218D, L218E, L218F, L218G, L218I, L218K, L218M, L218N, L218P, L218Q, L218R, L218S, L218T, L218V, L218W, L218Y, L277D, L277F, L277K, L277M, L277N, L277P, L277Q, L277R, L277W, L281G, L281M, L281N, L281P, L281Q, L281T, L281W, I285A, I285S, I285T, Y971D, Y971M, Y971N, Y971P, Y971Q, Y971W, L972A, L972D, L972E, L972F, L972G, L972I, L972K,

L972M, L972N, L972P, L972Q, L972R, or any combination thereof. In certain example embodiments, the one or more amino acid substitutions comprise two or more amino acid substitutions selected from:

- a. L218S and I285A,
- b. L218S and I285S,
- c. L218S and I285T,
- d. L218S and L972A,
- e. L218S and I972F,
- f. L218T and I285A,
- g. L218T and I285S,
- h. L218T and L972A, and
- i. L218T and L972F.

[0008] In certain example embodiments, the engineered Type V Cas further comprises one or more additional modifications that increase nuclease efficiency, RNA binding efficiency, or reduce off-target nuclease activity, or any combination thereof. In certain example embodiments, the Cas is a nickase or catalytically inactive (dCas). In certain example embodiments, the Cas is further linked to or otherwise capable of associating with a heterologous functional domain. In certain example embodiments, the heterologous functional domain is a nucleotide deaminase, a transposase, a reverse transcriptase, a recombinase, a methylase, a demethylase, an acetylase, or a deacetylase.

[0009] In an aspect, the present disclosure provides a composition comprising (i) the engineered Type V Cas of any one of the previous embodiments and (ii) at least one guide molecule capable of forming a complex with the Type V Cas and directing binding of the complex to a target sequence on a target polynucleotide. In certain example embodiments, the composition further comprising a donor template.

[0010] In an aspect, the present disclosure provides a nucleic acid molecule comprising a nucleotide sequence that encodes the engineered Type V Cas polypeptide of any one of the previous embodiments.

[0011] In an aspect, the present disclosure provides a nucleic acid molecule comprising a nucleotide sequence that encodes the Type V Cas and the at least one guide molecule of the composition of any one of the previous embodiments.

[0012] In an aspect, the present disclosure provides a vector comprising the nucleic acid molecule of any one of the previous embodiments. In certain example embodiments, the vector is a viral vector.

[0013] In an aspect, the present disclosure provides a delivery particle comprising the vector of any one of the preceding embodiments.

[0014] In an aspect, the present disclosure provides a delivery particle comprising the composition of any one of the preceding embodiments.

[0015] In an aspect, the present disclosure provides a cell comprising the engineered Type V Cas, the nucleic acid molecule, the vector, the composition, or any combination thereof, of any one of the previous embodiments.

[0016] In an aspect, the present disclosure provides a method of modifying target polynucleotides comprising administering the composition, the vector, the delivery particle, or any combination thereof, of any one of the previous embodiments.

[0017] 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

[0018] 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:

[0019] **FIG. 1** – (SEQ ID NO: 57-70) Structure of AsCas12, AsCas12 epitope locations and sequences, and HLA-binding scores of selected AsCas12 single mutants (single epitope with a single amino acid substitution).

[0020] **FIG. 2A-2B** – (A) AsCas12 single mutant nuclease efficiency (single epitope with a single amino acid substitution); (B) AsCas12 double mutant nuclease efficiency (two epitopes, each with a single amino substitution).

[0021] **FIG. 3** – (SEQ ID NO: 71-75) An example sampling method for deimmunizing polypeptides.

[0022] **FIG. 4** – A hybrid approach to calculating the potential energy of polypeptide structure.

- [0023] **FIG. 5** – (SEQ ID NO: 76-108) Additional energy functions implemented by an example method for deimmunizing polypeptides.
- [0024] **FIG. 6** – (SEQ ID NO: 76-87) Example scoring of 9mers by the additional energy functions.
- [0025] **FIG. 7** – Flowchart of example process for designing deimmunized polypeptides.
- [0026] **FIG. 8** – HLA Class I supertypes based on peptide binding motif.
- [0027] **FIG. 9** – Example machine learning models trained by supertype.
- [0028] **FIG. 10** – Example of peptide binding motifs corresponding to supertypes used to train the example machine learning models.
- [0029] **FIG. 11** – Six HLA Class I alleles represented by cell line belong to 4 supertypes.
- [0030] **FIG. 12** – (SEQ ID NO: 57) AsCas12 P210 epitope sequence, supertype, RNA contact, and structure.
- [0031] **FIG. 13** – (SEQ ID NO: 57) Scoring and structure: native AsCas12 P210 epitope vs. single mutants.
- [0032] **FIG. 14** – (SEQ ID NO: 58) AsCas12 P277 epitope sequence, supertype, RNA contact, and structure.
- [0033] **FIG. 15** – (SEQ ID NO: 58) Scoring and structure: native AsCas12 P277 epitope vs. single mutants.
- [0034] **FIG. 16** – (SEQ ID NO: 59) AsCas12 P972 epitope sequence, supertype, RNA contact, and structure.
- [0035] **FIG. 17** – (SEQ ID NO: 59) Scoring and structure: native AsCas12 P972 epitope vs. single mutants.
- [0036] **FIG. 18** – (SEQ ID NO: 109-113) asCas12a p2 mutant epitopes, sequence, structure, and scoring.
- [0037] **FIG. 19** – (SEQ ID NO: 114-117) asCas12a p11 mutant epitopes, sequence, structure, and scoring.
- [0038] **FIG. 20** – (SEQ ID NO: 118-121) asCas12a p160 mutant epitopes, sequence, structure, and scoring.
- [0039] **FIG. 21** – (SEQ ID NO: 122-127) asCas12a p241 mutant epitopes, sequence, structure, and scoring.
- [0040] **FIG. 22** – (SEQ ID NO: 128-130) asCas12a p505 mutant epitopes, sequence, structure, and scoring.

[0041] **FIG. 23** – (SEQ ID NO: 131-135) asCas12a p557 mutant epitopes, sequence, structure, and scoring.

[0042] **FIG. 24** – (SEQ ID NO: 136-143) asCas12a p869 mutant epitopes, sequence, structure, and scoring.

[0043] **FIG. 25** – (SEQ ID NO: 144-150) Scoring of asCas12a predicted epitope mutants.

[0044] The figures disclosed herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0045] Unless defined otherwise, technical, and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.); Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlett, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 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).

[0046] As used herein, unless otherwise indicated, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0047] 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.

[0048] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0049] The terms “about” or “approximately” as used herein, unless otherwise indicated, 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.

[0050] As used herein, unless otherwise indicated, 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. The biological sample can be obtained from an environment (e.g., water source, soil, air, and the like). The biological sample can be obtained from a plant or algae. The biological sample can contain prokaryotic organisms. Biological samples can be obtained via any suitable collection or harvesting technique including active and passive collection/harvesting methods, including but not limited to, puncture, cutting, digging, filtering, bagging, draining, and/or the like.

[0051] 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.

[0052] “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, unless otherwise indicated, 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.

[0053] As used herein, unless otherwise indicated, “stringent conditions” for hybridization refer to conditions under which a nucleic acid having complementarity to a target polynucleotide sequence predominantly hybridizes with the target polynucleotide sequence, and substantially does not hybridize to non-target polynucleotide 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., which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. 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.

[0054] “Hybridization” refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide

residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogsteen binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the “complement” of the given sequence.

[0055] As used herein, unless otherwise indicated, 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, unless otherwise indicated, “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, unless otherwise indicated, “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, unless otherwise indicated, “expression” also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and 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.

[0056] 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, unless otherwise indicated, 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, unless otherwise indicated, 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.

[0057] Various embodiments are disclosed 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 disclosed 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.

[0058] 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

[0059] Embodiments disclosed herein provide non-natural or engineered Type V Cas polypeptides, CRISPR-Cas systems thereof, compositions thereof, and delivery systems thereof, e.g., for delivery to cells. In general, the engineered Type V Cas polypeptide (e.g., an engineered Cas12 protein) or ortholog or variant thereof, has been modified to exhibit reduced immunogenicity as compared to a wild-type counterpart Type V Cas polypeptide. In general, embodiments disclosed herein provide Type V Cas polypeptides that are deimmunized without significantly altering activity, including but not limited to nucleic acid binding, complex formation, target binding, and target cleavage. In particular, the engineered Type V Cas polypeptides comprise modifications to various epitopes, including MHC I and MHC II binding regions. The engineered Type V Cas polypeptides may comprise a single mutation in a single epitope or may comprise a combination of mutations epitopes (e.g., two or three epitopes, each with a single mutation) to further reduce immunogenicity.

[0060] In another aspect, the present disclosure provides methods of reducing immunogenicity of a Type V Cas polypeptide. The methods may comprise introducing one or more modifications in the Type V Cas polypeptide. The modifications may include mutation of one or more epitopes of immune cells (e.g., T cells). In another aspect, embodiments disclosed herein provide methods of using the engineered Type V Cas polypeptides herein, CRISPR-Cas systems herein, compositions herein, and delivery systems herein, to modify one or more target polynucleotides, such as, for example, in cells.

[0061] Other engineered Type V Cas polypeptides, CRISPR-Cas systems, compositions, delivery systems, methods, features, and advantages of the present disclosure will be or become apparent to one having ordinary skill in the art upon examination of the following drawings, detailed description, and examples. It is intended that all such additional engineered Type V Cas proteins, systems, compositions, cargos and/or delivery vehicles, delivery systems, methods, features, and advantages be included within this description, and be within the scope of the present disclosure.

ENGINEERED TYPE V CAS POLYPEPTIDES WITH REDUCED IMMUNOGENICITY

[0062] In one aspect, embodiments disclosed herein comprise engineered Type V Cas polypeptides comprising one or more modifications that result in reduced immunogenicity compared to wild-type Type V Cas polypeptide. In certain example embodiments, the engineered Type V Cas polypeptides comprise one or more amino acid modifications in one or more epitopes that result in reduced immunogenicity compared to wild-type Type V Cas polypeptide. 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 used herein, unless otherwise indicated, the term “engineered” indicates the involvement of the hand of man and can include one or more modifications or mutations to the wild-type Type V Cas polypeptide. As used herein, unless otherwise indicated, the terms “modification” and “modified” with regard to an engineered Type V Cas polypeptide having reduced immunogenicity generally refer to a Type V Cas polypeptide having one or more amino acid insertions, deletions, substitutions relative to a base Type V Cas polypeptide from which it is derived. The base Type V Cas polypeptide may be a wild-type Type V Cas polypeptide sequence, or the base Type V Cas polypeptide may be a Type V Cas Polypeptide may be a having one or more previous modifications. For example, as will be described in further detail below, a number of Type V Cas polypeptides have been modified to affect other aspects of Cas activity, for example, to increase nuclease activity, enhance DNA binding, reduce off-target effects, change PAM recognition, and the like.

[0063] As used herein, unless otherwise indicated, “reduced immunogenicity” means modifications to the Type V Cas polypeptide that reduce binding or recognition by major histocompatibility complex (MHC) molecules such as MHC Class I and MHC Class II polypeptides. T cell receptors bind to antigens when they are complexed with the major histocompatibility complex. The present application identifies epitopes in Type V Cas polypeptides likely to be recognized, bound, and displayed on the cell surface by MHC polypeptides, and provides modifications that reduce the ability of MHC polypeptides to recognize, bind, and/or display the Type V Cas polypeptide-derived epitopes. In one example embodiment, the modifications reduce recognition and/or binding of MHC I polypeptides. In one example embodiment, the modifications reduce recognition and/or binding by MHC polypeptides encoded by HLA-A, HLA-B, HLA-C, HLA-E, HLA-F, or HLA-G. In another

example embodiment, the modification reduce recognition and/or binding by MHC II polypeptides. In one example embodiment, the modification reduce recognition and/or binding by MHC polypeptides encoded by HLA-DP, HDLA-DQ, or HLA-DR genes.

Type V Cas Polypeptide

[0064] Type V Cas polypeptides (e.g., Cas12 proteins) are a functionally divergent class of RNA-guided effector proteins which share a common single RuvC nuclease domain, but otherwise have low sequence similarity, and display diverse biochemical activities. *See* Zetsche et. al., “Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System”, *Cell* (2015), 163:3, 759-771; Shmakov et al., “Discovery and functional characterization of diverse class 2 CRISPR-Cas systems.” *Mol. Cell.* (2015), 60:385–397; Yan et al. “Functionally diverse type V CRISPR-Cas systems.” *Science* (2019), 363:88–91; Strecker et. al., “RNA-guided DNA insertion with CRISPR-associated transposases”. *Science* (2019). 365:48–53; Makarova et al., “Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants”. *Nat. Rev. Microbiol.* (2020). 18:67–83; Pausch et al., “CRISPR-CasΦ from huge phages is a hypercompact genome”. *Science* (2020). 369:333–337.

[0065] Cas12 proteins differ in cleavage activity. Cas12 proteins cleave target dsDNA and/or ssDNA (cis-cleavage), and many Cas12 proteins also exhibit collateral cleavage (trans-cleavage) activity, which cleaves ssDNA indiscriminately without requiring a PAM recognition. *See, e.g.,* Zetsche, (2015); Chen et al. “CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity.” *Science* (2018), 360 436–439; Li et al. “CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA.” *Cell Res.* (2018), 28 491–493; Li Y., Li S., Wang J., Liu et al. “CRISPR/Cas systems towards next-generation biosensing.” *Trends Biotechnol.* (2019), 37 730–743; Leung et al. “CRISPR-Cas12-based nucleic acids detection systems.” *Methods.* (2022), 203:276-281. Cas12c (also known as C2c3) proteins, including Cas12c1, Cas12c2, and OspCas12c, exhibit RNA-guided dsDNA interference activity (Shmakov et al., 2015; Yan et al., 2019).

[0066] Cas12 proteins differ in the recognition of PAM sequences. Many Cas12 proteins recognize T-rich PAM sequences. Cas12a (also known as Cpf1) cleaves dsDNA targets with a TTTV (V is A, G, or C) PAM. *See* Zetsche et al., 2015. Cas 12b recognizes a TTN PAM sequence. (Shmakov et al., 2015). Cas12c recognizes short TN or TG PAMs, respectively. *See* Kurihara et al., “Structure of the type V-C CRISPR-Cas effector enzyme.” *Mol Cell.* (2022), 82(10):1865-1877.e4. Certain Cas12 proteins recognize C-rich PAM sequences. Cas12l

recognizes short CC PAM sequences (e.g., CCY and CCB PAM sequences). See Urbaitis et al. “A new family of CRISPR-type V nucleases with C-rich PAM recognition”, EMBO Reports (2022) 23:e55481.

[0067] Cas12 proteins differ in their guide RNA architectures. Cas12a (Cpf1), Cas12i, and Cas12j (CasΦ) use a crRNA, whereas Cas12b (C2c1), Cas12c, Cas12e (CasX), Cas12f (Cas14), and Cas12l use dual-guide RNAs (e.g., crRNA/tracrRNA or crRNA/scoutRNA). See Kurihara et al. (2022); Urbaitis et al., (2022); Harrington et al. “A scoutRNA Is Required for Some Type V CRISPR-Cas Systems”, Molecular Cell (2020), 79:3, 416-424.e5.

[0068] Cas12 proteins differ in size. Cas12 protein orthologs are generally large proteins. Cpf1 (Cas12a) ranges from 1200–1500 amino acids, while C2c1 (Cas12b) has about 1300 amino acids. (Shmakov et al., 2015). In certain example embodiments, a Cas12 protein is a “small Cas12 protein” (e.g., with a size less than ~1200 amino acids), such as for example, Cas12e, Cas12f, Cas12g, Cas12h, Cas12i, Cas12j, Cas12k, or Cas12l. In certain example embodiments, the small Cas12 protein is a Cas12e protein (also known as a CasX protein). See Liu et al., “CasX enzymes comprise a distinct family of RNA-guided genome editors.” Nature (2019), 566, 218–223; Tsuchida, et al., “Chimeric CRISPR-CasX enzymes and guide RNAs for improved genome editing activity.” Mol. Cell (2022), 82, 1199–1209; Burstein et al., “New CRISPR-Cas systems from uncultivated microbes.” Nature (2017), 542, 237–241. In certain example embodiments, the small Cas12 protein is a Cas12f protein (formerly named a Cas 14 protein). See Kim, et al., “Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus.” Nat. Biotechnol. (2022), 40, 94–102; Harrington, et al., “Programmed DNA destruction by miniature CRISPR-Cas14 enzymes”. Science (2018), 362, 839–842; Karvelis, et al., “PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage”. Nucleic Acids Res. (2020), 48, 5016–5023. In certain example embodiments, the small Cas12 protein is a Cas12l protein. See Urbaitis et al. (2022).

Cas12a

[0069] In some examples, the Type II Cas polypeptide is Cas12a (Cpf1). By “Cpf1 (CRISPR associated protein Cpf1)” is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to GenBank Accession No. AJI61006.1 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). “Cpf1 function” can be defined by any of a number of assays including, but

not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein. By "Cpf1 nucleic acid molecule" is meant a polynucleotide encoding a Cpf1 polypeptide or fragment thereof. An exemplary Cpf1 nucleic acid molecule sequence is provided at GenBank Accession No. CP009633, nucleotides 652838 - 656740. Cpf1(CRISPR-associated protein Cpf1, subtype PREFRAN) is a large protein (about 1300 amino acids) that contains a RuvC-like nuclease domain homologous to the corresponding domain of Cas9 along with a counterpart to the characteristic arginine-rich cluster of Cas9. However, Cpf1 lacks the HNH nuclease domain that is present in all Cas9 proteins, and the RuvC-like domain is contiguous in the Cpf1 sequence, in contrast to Cas9 where it contains long inserts including the HNH domain. Accordingly, in particular embodiments, the CRISPR-Cas enzyme comprises only a RuvC-like nuclease domain.

[0070] The Cpf1 gene is found in several diverse bacterial genomes, typically in the same locus with *cas1*, *cas2*, and *cas4* genes and a CRISPR cassette (for example, FNFX1_1431-FNFX1_1428 of *Francisella cf. novicida* Fx1). Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the Cpf1 protein contains a readily identifiable C-terminal region that is homologous to the transposon ORF-B and includes an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). However, unlike Cas9, Cpf1 is also present in several genomes without a CRISPR-Cas context and its relatively high similarity with ORF-B suggests that it might be a transposon component. It was suggested that if this was a genuine CRISPR-Cas system and Cpf1 is a functional analog of Cas9 it would be a novel CRISPR-Cas type, namely type V (See Marakova et al., "Annotation and Classification of CRISPR-Cas Systems". *Methods Mol Biol.* (2015), 1311:47-75). However, as described herein, Cpf1 is denoted to be in subtype V-A to distinguish it from C2c1p which does not have an identical domain structure and is hence denoted to be in subtype V-B.

[0071] Cpf1 proteins are single RNA-guided endonucleases. Cpf1 cleaves DNA via a staggered DNA double stranded break. Its cleavage relies on a crRNA lacking tracr RNA to recruit a guide RNA comprising a guide sequence and a direct repeat, where the guide sequence hybridizes with the target nucleotide sequence to form a DNA/RNA heteroduplex. Based on current studies, Cpf1 nuclease activity also requires relies on recognition of a PAM sequence

which may be a T-rich sequence. In some embodiments, the PAM sequence is 5' TTN 3', wherein N is any nucleotide. Cpf1 creates a staggered cut at the target locus, with a 5' overhang, or a "sticky end" at the PAM distal side of the target sequence. In some embodiments, the 5' overhang is 4 or 5-nt.

[0072] In certain example embodiments, the Cas12a protein is derived from an organism from a genus comprising *Francisella*, *Prevotella*, *Lachnospiraceae*, *Butyrivibrio*, *Peregrinibacteria*, *Parcubacteria*, *Smithella* sp., *Acidaminococcus* sp., *Candidatus*, *Eubacterium*, *Moraxella*, *Leptospira*, and *Porphyromonas*. In certain example embodiments, the Cas12a protein is derived from an organism selected from *Francisella tularensis* 1, *Prevotella albensis*, *Lachnospiraceae* bacterium MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria* bacterium GW2011_GWA2_33_10, *Parcubacteria* bacterium GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae* bacterium MA2020, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Eubacterium rectale* MAD7, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae* bacterium ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, and *Porphyromonas macacae*. In certain example embodiments, the Cas12a is a *Lachnospiraceae* bacterium Cas12a (LbCas12a), an *Acidaminococcus* sp. (AsCas12a), a *Francisella novicida* Cas12a (FnCas12a), a *Moraxella bovoculi* Cas12a (MbCas12), an *Eubacterium rectale* Cas12a (MAD7), or a variant thereof. In certain example embodiments, the Cas12 is an *Acidaminococcus* sp. (AsCas12a), or a variant thereof.

Cas12b

[0073] In some examples, the Cas protein is Cas12b (C2c1). The C2c1 gene is found in several diverse bacterial genomes, typically in the same locus with cas1, cas2, and cas4 genes and a CRISPR cassette. Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the C2c1 protein contains an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). C2c1 (Cas12b) is derived from a C2c1 locus denoted as subtype V-B. Herein such effector proteins are also referred to as "C2c1p", e.g., a C2c1 protein (and such effector protein or C2c1 protein or protein derived from a C2c1 locus is also called "CRISPR enzyme"). Presently, the subtype V-B loci encompasses cas1-Cas4 fusion, cas2, a distinct gene denoted C2c1 and a CRISPR array. C2c1 (CRISPR-associated protein C2c1) is a large protein (about 1100 - 1300 amino acids) that contains a RuvC-like nuclease domain homologous to the corresponding domain of Cas9 along

with a counterpart to the characteristic arginine-rich cluster of Cas9. However, C2c1 lacks the HNH nuclease domain that is present in all Cas9 proteins, and the RuvC-like domain is contiguous in the C2c1 sequence, in contrast to Cas9 where it contains long inserts including the HNH domain. Accordingly, in particular embodiments, the CRISPR-Cas enzyme comprises only a RuvC-like nuclease domain.

[0074] C2c1 proteins are RNA guided nucleases. Its cleavage relies on a tracr RNA to recruit a guide RNA comprising a guide sequence and a direct repeat, where the guide sequence hybridizes with the target nucleotide sequence to form a DNA/RNA heteroduplex. Based on current studies, C2c1 nuclease activity also requires relies on recognition of PAM sequence. C2c1 PAM sequences may be T-rich sequences. In some embodiments, the PAM sequence is 5' TTN 3' or 5' ATTN 3', wherein N is any nucleotide. In a particular embodiment, the PAM sequence is 5' TTC 3'. In a particular embodiment, the PAM is in the sequence of *Plasmodium falciparum*. C2c1 creates a staggered cut at the target locus, with a 5' overhang, or a “sticky end” at the PAM distal side of the target sequence. In some embodiments, the 5' overhang is 7 nt. See Lewis and Ke, *Mol Cell*. (2017), 65(3):377-379.

[0075] In certain example embodiments, the Cas12 protein is a Cas12b protein. In certain example embodiments, the Cas12b protein is derived from an organism from a genus comprising *Alicyclobacillus*, *Desulfovibrio*, *Desulfonatronum*, *Opitutaceae*, *Tuberibacillus*, *Bacillus*, *Brevibacillus*, *Candidatus*, *Desulfatirhabdium*, *Elusimicrobia*, *Citrobacter*, *Methylobacterium*, *Omniotrophica*, *Phycisphaerae*, *Planctomycetes*, *Spirochaetes*, *Verrucomicrobiaceae*, *Lentisphaeria*, and *Laceyella*.

[0076] In certain example embodiments, the Cas12b protein is derived from an organism selected from *Alicyclobacillus acidoterrestris* (AaCas12b) (e.g., ATCC 49025), *Alicyclobacillus contaminans* (e.g., DSM 17975), *Alicyclobacillus herbarius* (e.g., DSM 13609), *Alicyclobacillus kakegawensis*, *Alicyclobacillus macrosporangiidus* strain DSM 17980, *Desulfovibrio inopinatus* (e.g., DSM 10711), *Desulfonatronum thiodismutans* (e.g., strain MLF-1), *Opitutaceae* bacterium TAV5, *Tuberibacillus calidus* (e.g., DSM 17572), *Bacillus thermoamylovorans* (e.g., strain B4166), *Bacillus hisashii* strain C4, *Bacillus sp.* NSP2.1, *Bacillus sp.* V3-13, *Brevibacillus agri* (e.g., BAB-2500), *Brevibacillus sp.* CF112, *Candidatus Lindowbacteria bacterium* RIFCSPLOWO2, *Desulfatirhabdium butyrativorans* (e.g., DSM 18734), *Elusimicrobia bacterium* RIFOXYA12, *Citrobacter freundii* (e.g., ATCC 8090), *Methylobacterium nodulans* (e.g., ORS 2060), *Omniotrophica WOR_2 bacterium*

RIFCSPHIGHO2, *Phycisphaerae bacterium* ST-NAGAB-D1, *Planctomycetes bacterium* RBG_13_46_10, *Spirochaetes bacterium* GWB1_27_13, *Verrucomicrobiaceae bacterium* UBA2429, *Lentisphaeria bacterium*, *Laceyella sediminis*, *Opitutaceae bacterium*, or a variant thereof.

[0077] **Table 1** provides a set of example Cas12 that may be modified to generate the reduced immunogenicity Type V Cas polypeptide embodiments disclosed herein. Any Cas12 may be in Human Codon Optimized form (e.g., Cas12 amino acids are followed by nuclear localization signals (NLS) (*italics*), a glycine-serine (GS) linker, and 3x HA tag).

[0078] **Table 1.**

CAS12 ORTHOLOGS	
<i>C2c1 orthologs</i>	
Alicyclobacillus macrosporangioides strain DSM 17980	SEQ ID NO: 1
Bacillus hisashii strain C4	SEQ ID NO: 2
Candidatus Lindowbacteria bacterium RIFCSPLOWO2	SEQ ID NO: 3
Elusimicrobia bacterium RIFOXYA12	SEQ ID NO: 4
Omnitrophica WOR_2 bacterium RIFCSPHIGHO2	SEQ ID NO: 5
Phycisphaerae bacterium ST-NAGAB-D1	SEQ ID NO: 6
Planctomycetes bacterium RBG_13_46_10	SEQ ID NO: 7
Spirochaetes bacterium GWB1_27_13	SEQ ID NO: 8
Verrucomicrobiaceae bacterium UBA2429	SEQ ID NO: 9
<i>Alicyclobacillus kakegawensis</i>	SEQ ID NO: 10
<i>Bacillus sp. V3-13</i>	SEQ ID NO: 11
<i>Desulfatirhabdium butyrativorans</i>	SEQ ID NO: 12
<i>Desulfonatromum thiodismutans</i>	SEQ ID NO: 13
<i>Lentisphaeria bacterium</i>	SEQ ID NO: 14

<i>Laceyella sediminis</i>	SEQ ID NO: 15
<i>Methylobacterium nodulans</i> (long form)	SEQ ID NO: 16
<i>Opitutaceae bacterium</i>	SEQ ID NO: 17
<i>Cpf1 orthologs</i> (shown in optional Human Codon Optimized form: Cpf1 amino acids are followed by nuclear localization signals (NLS) (<i>italics</i>), a glycine-serine (GS) linker, and 3x HA tag)	
Francisella tularensis subsp. novicida U112 (FnCpf1)	SEQ ID NO: 18
Lachnospiraceae bacterium MC2017 (Lb3Cpf1)	SEQ ID NO: 19
Butyrivibrio proteoclasticus (BpCpf1)	SEQ ID NO: 20
Peregrinibacteria bacterium GW2011_GWA_33_10 (PeCpf1)	SEQ ID NO: 21
Parcubacteria bacterium GWC2011_GWC2_44_17 (PbCpf1)	SEQ ID NO: 22
Smithella sp. SC_K08D17 (SsCpf1)	SEQ ID NO: 23
Acidaminococcus sp. BV3L6 (AsCpf1)	SEQ ID NO: 24
Lachnospiraceae bacterium MA2020 (Lb2Cpf1)	SEQ ID NO: 25
Candidatus Methanoplasma termitum (CMtCpf1)	SEQ ID NO: 26
Eubacterium eligens (EeCpf1)	SEQ ID NO: 27
Moraxella bovoculi 237 (MbCpf1)	SEQ ID NO: 28
Leptospira inadai (LiCpf1)	SEQ ID NO: 29
Lachnospiraceae bacterium ND2006 (LbCpf1)	SEQ ID NO: 30
Porphyromonas crevioricanis (PcCpf1)	SEQ ID NO: 31
Prevotella disiens (PdCpf1)	SEQ ID NO: 32
Porphyromonas macacae (PmCpf1)	SEQ ID NO: 33
Thiomicrospira sp. XS5 (TsCpf1)	SEQ ID NO: 34
Moraxella bovoculi AAX08_00205 (Mb2Cpf1)	SEQ ID NO: 35

Moraxella bovoculi AAX11_00205 (Mb3Cpf1)	SEQ ID NO: 36
Butyrivibrio sp. NC3005 (BsCpf1)	SEQ ID NO: 37
<i>Additional Cpf1 orthologs</i>	
NCBI WP_055225123.1	SEQ ID NO: 38
NCBI WP_055237260.1	SEQ ID NO: 39
NCBI WP_055272206.1	SEQ ID NO: 40
GenBank OLA16049.1	SEQ ID NO: 41

Other Engineered Type V Variants

[0079] In addition to wild-type Type V sequences, a base Type V Cas polypeptide sequence for the reduced immunogenicity modifications disclosed herein may be a Cas12 variant which has been previously modified. In a one example embodiment, the previous modifications could be other modifications to reduce immunogenicity. In another example, the previous modifications could be for reasons other than reducing immunogenicity. For example, the modifications could be to modify the level of nuclease activity (either to increase or decrease), to increase target polynucleotide binding, to alter the PAM recognition sequence required for target binding, to reduce off-target effects, or a combination thereof. Therefore, it is contemplated with the scope of the reduced immunogenicity Type V Cas polypeptides disclosed herein that the modifications for reducing immunogenicity disclosed herein may be layered on top of other modifications directed to other aspects of the Type V Cas polypeptide's function or activity. Example modifications are disclosed in Ling et al., "Improving the efficiency of CRISPR-Cas12a-based genome editing with site-specific covalent Cas12a-crRNA conjugates", Molecular Cell (2021), 81:22, 4747-4756.e7; Kleinstiver et al., "Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing", Nat. Biotechnol. (2019), 37, 276-282; P. Liu et al., "Enhanced Cas12a editing in mammalian cells and zebrafish", Nucleic Acids Res. (2019), 47 4169-4180; Jones et al., "Massively parallel kinetic profiling of natural and engineered CRISPR nucleases", Nat. Biotechnol. (2021), 39, pp. 84-93; Li et al., "Engineering CRISPR-Cpf1 crRNAs and mRNAs to maximize genome editing efficiency", Nat. Biomed. Eng. (2017), 1, 0066; Moon et al., "Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA

with a uridynylate-rich 3'-overhang", Nat. Commun. (2018), 9, 3651; Kocak, et al., "Increasing the specificity of CRISPR systems with engineered RNA secondary structures", Nat. Biotechnol. (2019), 37, 657-666), which are incorporated by reference herein in their entireties and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0080] In certain example embodiments, Type V Cas polypeptides have been modified to change the inherent double-strand DNA cleavage activity of the Type V Cas polypeptide so that it cleaves only a single-strand of dsDNA, i.e., to generate a nickase variant of the Type V Cas polypeptide. Accordingly, it is contemplated with the scope of the embodiments disclosed herein that the reduced immunogenicity modifications disclosed herein may also be used in combination with nickase variants of Type V Cas polypeptides. Such nickase variants may also be couple with other functional domains to expand gene editing functionality. For example, Cas12 nickases, may be combined with reverse transcriptase domains to generate prime editing systems (Anzalone et al. "Search-and-replace genome editing without double-strand breaks or donor DNA" Nature, 576:149-157 (2019); WO 2020/191233; WO 2020/191234; WO 2020/191245; WO 2020/191246; WO2020/191248; WO2020/191249; WO2020/191153; WO2020/191171; Chen et al. "Enhanced prime editing systems by manipulating cellular determinants of editing outcomes" Cell, 184(22):5635-5652 (2021); WO 2022/150790; Nelson et al. "Engineered pegRNAs improve prime editing efficiency" Nature Biotechnology 40:402-410 (2022); WO 2022/067130), or with both reverse transcriptase and integrase domains (WO2021/138469; Anzalone et al. "Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing," Nature Biotechnology. 40(5):731-740 (2021); WO2021/226558, Yarnall et al. "Drag-and-drop genome insertion of large sequences without double strand DNA cleavage using CRISPR-directed integrases," Nature Biotechnology. November 24, 2022; WO 2022/087235. Accordingly, one of ordinary skill in the art may use such modifications of Cas9 nickases to modify nickase variants of example Type V Cas polypeptides provided herein in a similar fashion. Accordingly, it is contemplated within the scope of the embodiments disclosed herein, that the reduced immunogenicity modifications may be used in Type V nickase variants that are further coupled to other functional domains such as, but not limited to, reverse transcriptase and integrase domains.

[0081] In certain example embodiments, Type V Cas polypeptides have also been modified to render the Type V Cas polypeptide catalytically inactive, i.e., a dead Cas (dCas). These dCas variants retain their ability for guide molecule mediated binding of target polynucleotides. As such they are often coupled with other functional domains to expand the gene editing functionality of such systems. For example, dCas9, which have many similarities with dCas12, have been coupled with adenosine and cytidine deaminase domains to generate single nucleotide base editors (Increasing the Genome-Targeting Scope and Precision of Base Editing With Engineered Cas9-Cytidine Deaminase Fusions. Kim, et al. *Nat. Biotechnol.* 35 371-376 (2017); Improving the DNA Specificity and Applicability of Base Editing Through Protein Engineering and Protein Delivery In Vitro and In Vivo. Rees et al., *Nat. Commun.* 8, 15790 (2017); Programmable Base Editing of A*T to G*C in Genomic DNA Without DNA Cleavage. Gaudelli et al. *Nature* 551, 464-471 (2017); In Vivo Base Editing of Post-Mitotic Sensory Cells. Yeh et al. *Nat. Commun.* 9, 2184 (2018); Base Editing: Precision Chemistry on the Genome and Transcriptome of Living Cells. Rees, H. A. and Liu, D. R. *Nat. Rev. Genet.* 19, 770-788. (2018); Targeting Fidelity of Adenine and Cytosine Base Editors in Mouse Embryos. Lee et al. *Nat. Commun.* 9: 4804. 1-5 (2018); Adenine Base Editing in an Adult Mouse Model of Tyrosinemia. Song et al. *Nat. Biomed. Eng.* 36, 536-539 (2018); Simultaneous Targeting of Linked Loci in Mouse Embryos Using Base Editing. Lee et al. *Sci. Rep.* 9, 1662 (2019); Cytosine and Adenine Base Editing of the Brain, Liver, Retina, Heart and Skeletal Muscle of Mice Via Adeno-Associated Viruses. Levy et al. *Biomed. Eng.* 1, 97-110 (2020); In Vivo Base Editing Restores Sensory Transduction and Transiently Improves Auditory Function in a Mouse Model of Recessive Deafness. Yeh et al. *Sci Trans. Med.* 12(546):eaay9101 (2020); Determinants of Base Editing Outcomes from Target Library Analysis and Machine Learning Arbab et al. *Cell.* 182, 463-480 (2020); Genome Editing with CRISPR-Cas Nucleases, Base Editors, Transposases, and Prime Editors. Anzalone et al. *Nat. Biotechnol.* 38, 824-844 (2020); A Bacterial Cytidine Deaminase Toxin Enables CRISPR-Free Mitochondrial Base Editing. Mok et al. *Nature.* 583, 631-637 (2020); Restoration of Visual Function in Adult Mice with an Inherited Retinal Disease via Adenine Base Editing. Suh et al. *Nat. Biomed. Eng.* 5, 169-178 (2021); In Vivo Adenine Base Editing Corrects Hutchinson-Gilford Progeria Syndrome. Koblan et al. *Nature* 589, 608-614 (2021); Base Editing of Hematopoietic Stem Cells Rescues Sickle Cell Disease in Mice. Newby et al. *Nature* 595, 295-302 (2021); In Vivo Somatic Base Editing and Prime Editing. Newby, G. A. and Liu, D.

R. *Molecular Therapy*. 29, 3107-3124 (2021); Disruption of HIV-1 Co-Receptors CCR5 and CXCR4 in Primary Human T Cells and Hematopoietic Stem and Progenitor Cells Using Base Editing. Knipping et al. *Mol. Ther.* 30, 130-144 (2022); In Vivo Base Editing Rescues Cone Photoreceptors in a Mouse Model of Early-Onset Inherited Retinal Degeneration. Choi et al. *Nat. Commun.* 13, 1830 (2022); In Vivo Base Editing by a Single I.V. Vector Injection for Treatment of Hemoglobinopathies. Li et al. *Clin. Investig. Insight* 7 (19): e162939, non-LTR retrotransposon (WO2021/102042), epigenetic modifiers such as methylases and demethylases (Hsu et al. "DNA targeting specificity of RNA-guided Cas9 nucleases". *Nature biotechnology* (2013) 31: 827–32. Liu et al. "Editing DNA Methylation in the Mammalian Genome". *Cell* (2016) 167: 233–247 e17) acetylases and deacetylases (CHROMA), transcription initiators (CRISPRa) (Gilbert et al. "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013 Jul 18;154(2):442-51; Gilbert et al. "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation". *Cell*. 2014 Oct 23;159(3):647-61; Konermann et al. "Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex". *Nature* 517, 583–588 (2015); Zalatan et al. "Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds". *Cell*. 2015 Jan 15;160(1-2):339-50), and transcription repressors (CRISPRi) (Qi et al. "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression". *Cell*. 2013 Feb 28;152(5):1173-83; Gilbert et al. "CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013 Jul 18;154(2):442-51; Gilbert et al. "Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation". *Cell*. 2014 Oct 23;159(3):647-61). Accordingly, it is contemplated within the scope of the reduced immunogenicity Type V Cas modifications disclosed herein, that such modifications may be used in combination with dCas Type V variants that are further coupled with functional domains such as but not limited to the aforementioned functional domains.

Identification of Analogous Residues

[0082] For the sake of convenience, the following example reduced immunogenicity Type V Cas modifications will be discussed in the context of modifications made to a wild-type *Acidaminococcus* sp. (AsCas12a). However, it is contemplated with the scope of the embodiments that the same or biochemically similar modifications can be made to positions in other Type V Cas polypeptides outlined above that are analogous to the example AsCas12a modifications discussed in this section. An analogous position is an amino acid residue in

another Type V Cas polypeptide that is similarly situated within the 3D structure of the other Type V Cas polypeptide and that would otherwise be recognized as sharing similar biochemical and biophysical properties that led to selection of the example AsCas12a residues disclosed herein.

[0083] The above modifications are by way of example only, as noted above one of ordinary skill in the art will be able to apply the examples to analogous residues on other Type V Cas polypeptides. For example, one of ordinary skill in the art can identify analogous residues by their location in the 3D structure of the Type Cas V polypeptide and identify other modifications that result in substitutions similar biophysical and biochemical properties. Several methods are available in the art and routinely used to identify protein 3D structures, including Type V Cas polypeptide structures including X-ray crystallography, cryo-EM, *in silico* base prediction, protein structural searching, and epitope prediction methods.

X-ray Crystallography

[0084] The crystal structures of several Cas12 proteins (e.g., FnCas12a, AaCas12b, Cas12c2, *Ruminococcus* sp. Cas13d (UrCas13d)) have been identified. *See*, Swarts et al. “Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a.” *Mol Cell*. 2017 Apr 20;66(2):221-233.e4; Yang et al. “PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease.” *Cell*. 167(7):1814-1828.e12. (2016); Liu et al., “C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism”, *Molecular Cell* (2017) 65, 310–322; Zhang et al. “Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d.” *Nat Commun* (2019) 10, 2544, which are incorporated by reference herein in their entireties and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. Accordingly, one of ordinary skill in the art may use X-ray crystallography resolved structures of other Type V Cas polypeptides to identify residues analogous to the example AsCas12a residues provided herein which can be modified in a similar fashion to reduce immunogenicity of those other Type V Cas polypeptides.

Cryo-EM

[0085] In addition to traditional X-ray crystal structures, recent advances in Cryo-EM have allowed for determination of sharper molecule 3D structures. The advantages of Cryo-EM are that, compared to SC-XRD, the rapid freeze treatment of the protein maintains its closer-to-

native state, only a small amount of protein (about 0.1 mg) is required, the method is more forgiving on sample purity, and the protein does not need to crystalize. Advances in Cryo-EM have enabled more structures to be imaged and determined at better resolutions, at faster speeds, and at lower cost. *See* Better, Faster, Cheaper: Recent Advances in Cryo–Electron Microscopy Eugene Y.D. Chua, Joshua H. Mendez, Micah Rapp, Serban L. Ilca, Yong Zi Tan, Kashyap Maruthi, Huihui Kuang, Christina M. Zimanyi, Anchi Cheng, Edward T. Eng, Alex J. Noble, Clinton S. Potter, Bridget Carragher. *Annual Review of Biochemistry* 2022 91:1, 1–32) It is estimated that by 2024, more protein structures will be determined by cryo-EM than by X-ray crystallography indicating the technique has become a routine part of protein characterization. *See*, e.g., Ewen Callaway, “The Protein-Imaging Technique Taking Over Structural Biology” *Nature* 2020, 578:201; Assaiya et al., “An overview of the recent advances in cryo-electron microscopy for life sciences.” *Emerg Top Life Sci* 14 May 2021; 5 (1): 151–168; Shoemaker & Ando. “X-rays in the Cryo-Electron Microscopy Era: Structural Biology's Dynamic Future.” *Biochemistry*. 2018 Jan 23;57(3):277-285. Accordingly, one of ordinary skill in the art may use Cryo-EM resolved structures of other Type V Cas polypeptides to identify residues analogous to the example AsCas12a residues provided herein which can be modified in a similar fashion to reduce immunogenicity of those other Type V Cas polypeptides.

[0086] A number Cryo-EM structures of several Cas12 proteins (e.g., LbCas12a, *Eubacterium siraeum* Cas13d (EsCas13d), Cas12e (formerly CasX), UnCas12f1, Cas12g, LbCas12i, Cas12j (Cas12Φ), and Cas12k) have been identified. *See* Zhang et al. “Structural Basis for the Inhibition of CRISPR-Cas12a by Anti-CRISPR Proteins.” *Cell Host Microbe* (2019) 25: 815; Kurihara et al. “Cryo-EM structure of the Cas12c2-sgRNA-target DNA ternary complex”, *Mol Cell* (2022) 82: 1865-1877.e4Zhang et al. “Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d.” *Cell* (2018) 175: 212-223.e17; Liu et al. “CasX enzymes comprise a distinct family of RNA-guided genome editors.” *Nature* (2019) 566(7743):218-223; Renjian et al. “Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease”, *Nucleic Acids Research* (2021), 49:7, 4120–4128; Li et al. “Cryo-EM structure of the RNA-guided ribonuclease Cas12g”. *Nat Chem Biol* (2021) 17, 387–393; Zhang et al. “Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease.” *Nat Struct Mol Biol* (2020) 27: 1069-1076; Pausch et al. “DNA interference states of the hypercompact CRISPR-Cas Phi effector.” *Nat Struct Mol Biol*

(2021) 28: 652-661; Xiao et al. "Structural basis of target DNA recognition by CRISPR-Cas12k for RNA-guided DNA transposition." *Mol Cell* (2021) 81: 4457-4466.e5, which are incorporated by reference herein in their entireties and can be used to identify residues analogous to the example AsCas12a residues in other Type V Cas polypeptides.

In Silico Structure Predictions

[0087] Further, a number of *in silico* methods are also available to predict a protein's 3D structure based on the protein's primary sequence and/or other features. One example method for polypeptide structure prediction is AlphaFold. Jumper *et al.* "Highly accurate protein structure prediction with AlphaFold". *Nature*. 2021 596 (7873): 583–589. Database repositories of predicted 3D polypeptide structures have also been established. The European Bioinformatics Institute and DeepMind have jointly constructed the AlphaFold - EBI database (alphafold.ebi.ac.uk) for predicted protein structures. *See* Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. (January 2022). "AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models". *Nucleic Acids Res.* 50 (D1): D439–D444. Other example *in silico* methods for 3D protein structure prediction include RoseTTAFold (Baek *et al.* "Accurate prediction of protein structures and interactions using a three-track neural network" *Science* (2021) 373:871-876); OmegaFold (Wu *et al.* "High-resolution *de novo* structure prediction from primary sequence." *bioRxiv* 2022.07.21.500999; doi.org/10.1101/2022.07.21.500999). Accordingly, one of ordinary skill in the art may use *in silico* predicted 3D structures of other Type V Cas polypeptides to identify residues analogous to the example AsCas12a residues provided herein which can be modified in a similar fashion to reduce immunogenicity of those other Type V Cas polypeptides.

Structural Searching

[0088] In addition, given the explosion in structural data provided by the aforementioned methods, a number of powerful protein structural similarity searching methods have been developed that allow one of ordinary skill in the art to identify proteins having a similar domain or overall 3D structure. *See, e.g.,* Deng et al. "MADOKA: an ultra-fast approach for large-scale protein structure similarity searching". *BMC Bioinformatics* 2019, 20(Suppl 19), 662; Aderinwale *et al.* "Real-time structure search and structure classification for AlphaFold protein models." *Communications Biology*, 5, 316 (2022). doi.org/10.1038/s42003-022-03261; van Kempen *et al.* "Foldseek; fast and accurate protein structure search" *bioRxiv*

2022.02.07.479398; doi.org/10.1101/2022.02.07.479398. Accordingly, one of ordinary skill in the art may use a sequence of a Type V Cas polypeptide to determine its likely structure by identifying other proteins for which a 3D structure is known. The predicted protein structural similarity may then be used to identify residues in the other Type V Cas polypeptide that are analogous to the example AsCas12a residues provided herein, which can be modified in a similar fashion to reduce immunogenicity of other Type V Cas polypeptides.

Epitope Prediction

[0089] Finally, in addition to the epitope prediction method described in the Working Example section of this application, there are other epitope prediction methods available. Accordingly, one of ordinary skill in the art could use said methods on other Type V Cas polypeptides to determine likely MHC binding epitopes and compare those epitopes to the example AsCas12a epitopes provided in this application to determine if they are similarly located in that particular Type V Cas polypeptide in order to identify analogous residues. The aforementioned structural tools may also be used to resolve the relevant 3D structure around the identified epitopes to further aid in the identification of analogous residues.

[0090] Immunogenic T cell epitopes in a CRISPR polypeptide or a fragment thereof can be identified using an epitope prediction tool. *See e.g.* Sanchez-Trincado et al., “Fundamentals and Methods for T- and B-Cell Epitope Prediction,” J. Immunol. Res. 2017; 2017: 2680160. Published online 2017 Dec 28. Such methods fall into two general categories: structure-based methods that rely on modeling the peptide-MHC structure and data-driven methods that rely on peptide sequences that are known to bind to MHC molecules. Sanchez-Trincado et al., list epitope prediction tools that rely on a variety of methods including structure-based tools (SB); sequence motifs (SM); motif matrices (MM); quantitative structure-activity relationship models (QSAR); quantitative affinity matrices (QAM); support vector machines (SVM); artificial neural networks (ANN) and combinations of these methods.

[0091] Epitope prediction tools that are useful for predicting MHC Class I epitopes in CRISPR polypeptide or fragment thereof include, for example MAPP (See e.g., Hakenberg J. et al., “MAPPP: MHC class I antigenic peptide processing prediction.” Applied Bioinformatics. 2003; 2(3):155–158.); PEPVAC (See e.g., Reche et al., space “PEPVAC: a web server for multi-epitope vaccine development based on the prediction of supertypic MHC ligands.” Nucleic Acids Research. 2005; 33 (Supplement 2):W138–W142); EPISOPT (See e.g., Molero-Abraham M. et al., “Selection of conserved epitopes from hepatitis C virus for

pan-population stimulation of T-cell responses.” *Clinical and Developmental Immunology*. 2013;2013: 10); BIMAS (See e.g., Parker K. C. et al., “Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains.” *The Journal of Immunology*. 1994; 152 (1):163–175; Propred-1 (See e.g., Singh H. et al., “ProPred1: prediction of promiscuous MHC class-I binding sites.” *Bioinformatics*. 2003; 19(8):1009–1014.); EpiJen (See e.g., Doytchinova I. A. et al., “EpiJen: a server for multistep T cell epitope prediction.” *BMC Bioinformatics*. 2006;7(1):p. 131); IEDB-MHCI (See e.g., Zhang Q. et al., “Immune epitope database analysis resource (IEDB-AR)” *Nucleic Acids Research*. 2008;36(Web Server issue):W513–W518); NetMHC (See e.g., Nielsen M. et al., “Reliable prediction of T-cell epitopes using neural networks with novel sequence representations.” *Protein Science*. 2003; 12(5):1007–1017); NetMHCpan (See e.g., Nielsen M. et al., “NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence.” *PLoS One*. 2007; 2(8, article e796)); nHLApred (See e.g., Bhasin M. et al., “A hybrid approach for predicting promiscuous MHC class I restricted T cell epitopes.” *Journal of Biosciences*. 2007; 32(1):31–42); NetCTL (See e.g., Larsen M. V. et al., “An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions.” *European Journal of Immunology*. 2005; 35(8):2295–2303); and WAPP (See e.g., Donnes P. et al., “Integrated modeling of the major events in the MHC class I antigen processing pathway.” *Protein Science*. 2005; 14 (8):2132–2140).

[0092] Epitope prediction tools that are useful for predicting MHC Class II epitopes in a CRISPR polypeptide or fragment thereof include, for example, EpiDOCK (See e.g., Atanasova M. et al., “EpiDOCK: a molecular docking-based tool for MHC class II binding prediction.” *Protein Engineering, Design and Selection*. 2013;26(10):631–634); PREDIVAC (See e.g., Oyarzun P. et al., “PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity.” *BMC Bioinformatics*. 2013; 14(1):p. 52.); EpiTOP (See e.g., Dimitrov I. et al., “EpiTOP—a proteochemometric tool for MHC class II binding prediction.” *Bioinformatics*. 2010; 26(16):2066–2068); TEPITOPE (See e.g., Sturniolo T., et al. “Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices.” *Nature Biotechnology*. 1999; 17(6):555–561); Propred (See e.g., Singh H. et al., “ProPred: prediction of HLA-DR binding sites.” *Bioinformatics*. 2001; 17(12):1236–1237); IEDB-MHCII (See e.g., Zhang Q. et al., “Immune

epitope database analysis resource (IEDB-AR)” *Nucleic Acids Research*. 2008;36(Web Server issue):W513–W518); IL4pred (See e.g., Dhanda S. K. et al., “Prediction of IL4 inducing peptides.” *Clinical and Developmental Immunology*. 2013;2013:9); MHC2PRED (See e.g., Bhasin M. et al., “SVM based method for predicting HLA-DRB1*0401 binding peptides in an antigen sequence.” *Bioinformatics*. 2004;20(3):421–423); NetMHCII (See e.g., Nielsen M. et al., “Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method.” *BMC Bioinformatics*. 2007;8(1):p. 238); and NetMHCIIpan (See e.g., Nielsen M., et al. “Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan.” *PLoS Computational Biology*. 2008;4(7)).

[0093] Epitope prediction tools that are useful for predicting either MHC Class I or MHC Class II epitopes in a CRISPR polypeptide or fragment thereof include, for example, MotifScan (See e.g., [hiv.lanl.gov/content/immunology/motif_scan/motif_scan.](http://hiv.lanl.gov/content/immunology/motif_scan/motif_scan/)); Rankpep (See e.g., Reche P. A., et al., “Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles.” *Immunogenetics*. 2004;56(6):405–419.); SYFPEITHI (See e.g., Rammensee H. G. et al., “SYFPEITHI: database for MHC ligands and peptide motifs.” *Immunogenetics*. 1999;50(3-4):213–219); Vaxign (See e.g., He Y. et al., “Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development.” *Journal of Biomedicine and Biotechnology*. 2010;2010:15); MHCpred (See e.g., Guan P. et al., “MHCpred: a server for quantitative prediction of peptide–MHC binding.” *Nucleic Acids Research*. 2003;31(13):3621–3624); MULTIPRED2 (See e.g., Zhang G. L. Et al., “MULTIPRED2: a computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles.” *Journal of Immunological Methods*. 2011;374(1-2):53–61); SVMHC (See e.g., Donnes P. et al., “Prediction of MHC class I binding peptides, using SVMHC.” *BMC Bioinformatics*. 2002;3(1):p. 25); and SVRMHC (See e.g., Liu W. et al., “Quantitative prediction of mouse class I MHC peptide binding affinity using support vector machine regression (SVR) models.” *BMC Bioinformatics*. 2006;7(1):p. 182).

[0094] Alternatively, or in addition, MHC Class I binding specificity can be analyzed using allele-specific predictors, either by using simple sequence motifs e.g., xLxxxxxx(L/V) for HLA-A02:01) or Position Weight Matrices (PWM) or with machine learning frameworks such as neural networks, hidden Markov Models, support vector machines, or convolutional neural networks. The machine learning models have the capacity to consider potential correlations between different positions within HLA-I ligands. (See e.g., Gfeller, D. et al., “Predicting

Antigen Presentation—What Could We Learn From a Million Peptides?” Front Immunol. 2018; 9: 1716.)

[0095] Experimental ligands/peptides are available for only about 100 HLA-I alleles, so that the ligand specific predictors described above are useful only for a small fraction of the more than 12,000 HLA I alleles that have been identified to date. To identify additional ligands/peptides, “pan-allele predictors” can be used. For these methods, the input of the algorithm includes both the sequence of the ligand and the sequence of the HLA-I allele (or of its binding site). These algorithms can capture correlations between amino acids in the HLA-I-binding site and in the ligand. An exemplary pan-specific algorithm is the NetMHCpan tool. (See e.g., Jurtz V. et al., “NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data.” J. Immunol (2017) 199:3360–8.10.4049) which includes several features specific for HLA-I molecules, such as combining peptides of different lengths in the training and incorporating peptide length preferences.

[0096] Useful ligand predictors for both allele specific and pan-allele analysis are summarized in **Table 2** below (adapted from Gfeller, D. et al., “Predicting Antigen Presentation—What Could We Learn From a Million Peptides?” Front Immunol. 2018; 9: 1716; Sanchez-Trincado et al., “Fundamentals and Methods for T- and B-Cell Epitope Prediction,” J. Immunol. Res. 2017; 2017: 2680160.)

[0097] **Table 2.**

Name	Training data	Output	Algorithm	Allele coverage	Access
NetMHC4.0	BA	BA	NN	S	cbs.dtu.dk/services/NetMHC/
NetMHCpan4.0	BA + MS	R (BA)	NN	Pan	cbs.dtu.dk/services/NetMHCpan-4.0/
NetMHCII	BA + MS	R (BA)	NN	S	cbs.dtu.dk/services/NetMHCII/
NetMHCIIpan	BA + MS	R (BA)	NN	Pan	cbs.dtu.dk/services/NetMHCIIpan/
MixMHCpred	MS	R	PWM	S	github.com/GfellerLab/MixMHCpred
MHCflurry	BA	BA	NN	S	github.com/openvax/mhcflurry
PickPocket	BA	BA	PWM	Pan	cbs.dtu.dk/services/PickPocket/
NetMHCstabpan	BS	BS	NN	Pan	cbs.dtu.dk/services/NetMHCstabpan/
NetMHCstab	BS	BS	NN	S	cbs.dtu.dk/services/NetMHCstab/
NetMHCcons	BA	BA	C	S	cbs.dtu.dk/services/NetMHCcons/
IEDB consensus	BA	R	C	S	ools.iedb.org/mhci/

Name	Training data	Output	Algorithm	Allele coverage	Access
SMMPMBEC	BA	R	PWM	S	github.com/ykimbiology/smmpmbec
MHCnuggets	BA	BA	NN	S	github.com/KarchinLab/mhcnuggets-2.0

(Abbreviations: BA, binding affinity; BS, binding stability; MS, HLA peptidomics data; R, ranking; NN, Neural network (including deep networks); PWM, position weight matrices; C, consensus; S, allele specific; Pan, pan-class I.)

[0098] MHC Class-II binding specificity can also be analyzed with machine learning frameworks. Modeling the binding specificity of MHC Class II alleles can be more challenging than modeling of MHC Class I alleles due to a number of factors, for example: 1) MHC Class II alleles tend to be more degenerate and less specific motifs; 2) MHC Class II molecules form dimers, resulting in more diversity, particularly where both members of a dimer or polymorphic; and 3) MHC Class II molecules tend to have greater conformational flexibility, which can be difficult to predict from short peptide sequences.

[0099] Useful Allele-specific HLA-II ligand predictors can include NetMHCII (See e.g., Jensen K.K. et al., “Improved methods for predicting peptide binding affinity to MHC class II molecules.” *Immunology* (2018) 154:394-406), ProPred (See e.g., Singh H. et al., “ProPred: prediction of HLA-DR binding sites.” *Bioinformatics* (2001) 17:1236–7.), MHCpred (See e.g., Guan P. et al., “MHCpred: a server for quantitative prediction of peptide-MHC binding.” *Nucleic Acids Res* (2003) 31:3621–4); TEPITOPE (See e.g., Sturniolo T. et al., “Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices.” *Nat Biotechnol* (1999) 17:555–61.); and consensus methods (See e.g., Wang P., et al., “A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach.” *PLoS Comput Biol* (2008) 4:e1000048.). Pan-specific class II predictors typically include NetMHCIIpan (See e.g., Jensen K.K. et al., “Improved methods for predicting peptide binding affinity to MHC class II molecules.” *Immunology* (2018) 154:394-406).

Example Reduced Immunogenicity Modifications

[0100] In an aspect, the present disclosure provides an engineered Type V Cas having reduced immunogenicity compared to a corresponding wild-type Type V Cas. In certain example embodiments, the engineered Type V Cas comprises one or more modifications of one or more epitopes of the wild-Type V Cas. In certain example embodiments, the one or

more modifications of one or more epitopes comprise one or more amino acid substitutions in the one or more epitopes. In certain example embodiments, the one or more epitopes correspond to one or more MHC Class I and/or MHC Class II binding sites in the wild-type Type V Cas. In certain example embodiments, the one or more MHC Class I and/or MHC Class II binding sites are selected from one or more HLA Class I binding sites, HLA Class II binding sites, or any combination thereof. In certain example embodiments, the one or more HLA Class I binding sites are selected from one or more supertypes selected from A1, A2, A3, A24, A26, B7, B8, B27, B38, B44, B58, B62, C2, C3 (or C17), C4 (or C02), or any combination thereof. In certain example embodiments, the one or more HLA Class I binding sites are selected from one or more supertypes selected from A2 or C17, or any combination thereof.

[0101] In certain example embodiments, the one or more modifications of one or more epitopes result in fewer binding regions and/or weaker binding regions as compared to the wild-type epitopes. In certain example embodiments, the one or more epitopes have contact with RNA of a CRISPR-Cas complex. In certain example embodiments, the one or more epitopes do not have contact with RNA of a CRISPR-Cas complex. In certain example embodiments, the one or more modifications of one or more epitopes results in an engineered Type V Cas polypeptide with equivalent or greater catalytic activity as compared to the wild-type Type V Cas polypeptide. In certain example embodiments, the one or more modifications of one or more epitopes results in a Type V Cas polypeptide with reduced catalytic activity as compared to the wild-type Type V Cas polypeptide. In certain example embodiments, the one or more modifications of one or more epitopes results in a dead Type V Cas polypeptide.

[0102] In certain example embodiments, the one or more modifications are made at one or more Cas12 domains within a nuclease (NUC) lobe and/or a recognition (REC) lobe. In certain example embodiments, the one or more modifications are made at one or more Cas12 domains selected from an REC1 subdomain, an REC2 subdomain, an RuvC-II subdomain, and any combination thereof. In certain example embodiments, the one or more modifications are made at one or more residues of the helix 1 of the RuvC-II subdomain. In certain example embodiments, the Type V Cas polypeptide is a AsCas12a. In certain example embodiments, the one or more modifications are made at AsCas12a residues 210-218, 227-285, and/or 971-979, or at residues in another Type V Cas analogous thereto. In one example embodiment, the one or more modifications are made at AsCas12a 211, 218, 277, 281, 285, 971, 972, and/or 979, or at residues in another Type V Cas analogous thereto.

[0103] In certain example embodiments, the one or more amino acid substitutions comprise substitution of one or more original residues selected from I, L, Y, and any combination thereof, with one or more naturally occurring residues selected from A, D, E, F, G, I, K, M, N, P, Q, R, S, T, V, W, Y, and any combination thereof. In certain example embodiments, the one or more amino acid substitutions comprise: L211V, L211A, L218A, L218D, L218E, L218F, L218G, L218I, L218K, L218M, L218N, L218P, L218Q, L218R, L218S, L218T, L218V, L218W, L218Y, L277D, L277F, L277K, L277M, L277N, L277P, L277Q, L277R, L277W, L281G, L281M, L281N, L281P, L281Q, L281T, L281W, I285A, I285S, I285T, Y971D, Y971M, Y971N, Y971P, Y971Q, Y971W, L972A, L972D, L972E, L972F, L972G, L972I, L972K, L972M, L972N, L972P, L972Q, L972R, or any combination thereof. In certain example embodiments, the one or more amino acid substitutions comprise two or more amino acid substitutions selected from: L218S and I285A, L218S and I285S, L218S and I285T, L218S and L972A, L218S and L972F, L218T and I285A, L218T and I285S, L218T and L972A, and L218T and L972F. In certain example embodiments, the engineered Type V Cas polypeptide is catalytically dead and comprises the following substitution or substitutions: L211A, I285V, or L218T and I285T.

CRISPR-CAS COMPLEXES COMPRISING ENGINEERED TYPE V CAS POLYPEPTIDES

[0104] In another aspect, embodiments disclosed herein comprise CRISPR-Cas complexes comprising the reduced immunogenicity Type V Cas polypeptides discussed above and a guide molecule and compositions thereof. Type V Cas polypeptides form a ribonucleoprotein (RNP) complex with a guide RNA (e.g., a gRNA or a sgRNA) (also referred to herein as a CRISPR complex).

Guide Molecules

[0105] In certain example embodiment, CRISPR-Cas systems disclosed herein include one or more guide molecules. As used herein, unless otherwise indicated, the term “guide molecule” (used interchangeably herein with “guide sequence” or “guide RNA” or “gRNA” or “dual guide RNA” or “crRNA” or “tracrRNA” or “scoutRNA” (see Harrington et al., 2020) or “single guide RNA” or “sgRNA” or “chimeric RNA”) comprises any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target polynucleotide sequence and direct sequence-specific binding of the CRISPR-Cas complex between the engineered Type V Cas polypeptide and the guide molecule to the target

polynucleotide sequence. As used herein, unless otherwise indicated, the term “polynucleotide” refers to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof.

[0106] In certain example embodiments, each engineered Type V Cas polypeptide in a CRISPR-Cas system is coupled with, is configured to complex with, or is otherwise associated with its own guide RNA. In some embodiments, each engineered Type V Cas polypeptide, in a system composed of more than one engineered Type V Cas polypeptides, is associated with one or more different guide molecules compared to other engineered Type V Cas polypeptides within the same system.

[0107] The ability of a guide RNA to direct sequence-specific binding of a CRISPR-complex to a target polynucleotide sequence (e.g., DNA sequence) may be assessed by any suitable assay. For example, the components of a CRISPR-Cas system sufficient to form a polynucleotide-targeting complex, including the guide RNA to be tested, may be provided to a host cell having the corresponding target polynucleotide sequence, such as by transfection with vectors encoding the components of the CRISPR complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target double-stranded polynucleotide sequence, such as by Surveyor assay (Qui et al. 2004. *BioTechniques*. 36(4)702-707), which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas polypeptides, CRISPR-Cas systems, compositions, nucleic acid molecules, delivery vehicles, delivery systems, and methods disclosed herein.

[0108] Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target polynucleotide-targeting sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible and will occur to those skilled in the art.

[0109] A guide sequence may be selected to target any target polynucleotide sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding

RNA (ncRNA), long non-coding RNA (lncRNA), and small 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.

[0110] In some embodiments, the degree of complementarity between the guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, can be 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 examples of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), Clustal W, 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), which are incorporated by reference herein in their entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0111] In some embodiments, a guide sequence is selected to reduce the degree of secondary structure within the guide RNA. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (*See, e.g.,* A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62), which are incorporated by reference herein in their entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0112] In some embodiments, the guide RNA is configured to minimize or reduce off-target effects. Guide sequences and strategies to minimize toxicity and off-target effects can be via mutation as disclosed herein, or as in WO 2014/093622 (PCT/US2013/074667), which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0113] In certain embodiments, a guide RNA or crRNA includes or is only composed of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA includes or is only composed of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0114] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[0115] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27 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.

[0116] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, 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 some embodiments, 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 some embodiments, the tracr sequence and crRNA sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin.

[0117] In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and tracr RNA can be 30 or 50 nucleotides in length. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it being advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0118] In some embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e., an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0119] Many modifications to guide sequences are known in the art and are further contemplated within the context of this invention. Various modifications may be used to increase the specificity of binding to the target sequence and/or increase the activity of the Cas protein and/or reduce off-target effects. Example guide sequence modifications are described in International Patent Application No. PCT US2019/045582, specifically paragraphs [0178]-[0333], which is incorporated herein by reference in its entirety and can be adapted for use with the engineered Type V Cas proteins disclosed herein.

Donor Polynucleotide

[0120] A CRISPR-Cas system herein can further integrate a donor polynucleotide (also referred to herein as a “donor template” or a “recombination template polynucleotide” or “template sequence” or “template polynucleotide”) or a portion thereof into a target polynucleotide. As such, in some embodiments, the CRISPR-Cas system includes, in some embodiments, one or more donor polynucleotides. The terms donor oligodeoxynucleotide (ODN) (which encompasses both single stranded (ss) and double stranded (ds) polynucleotides and sequences) is used in some instances herein interchangeably with “donor polynucleotide”. In some embodiments, the donor/insert polynucleotide is a double stranded (ds) polynucleotide. In some embodiments, the donor/insert polynucleotide is a dsDNA, dsRNA, or a DNA hybrid (e.g., a dsDNA/RNA hybrid). In some embodiments, the donor/insert polynucleotide is a single stranded (ss) polynucleotide. In some embodiments, the donor/insert polynucleotide is a ssDNA or ssRNA. In some embodiments, the donor/insert polynucleotide is protected from degradation with chemical modifications. Suitable chemical modifications for protecting DNA and/or RNA from degradation are generally known in the art.

[0121] In some embodiments, the donor/insert polynucleotide is configured to introduce one or more mutations to the target polynucleotides, polypeptides, and/or other gene product, introduce or correct a premature stop codon in the target polynucleotides, polypeptides, and/or other gene product, disrupt a splicing site, restore a splicing site, or insert a gene or gene fragment at one or multiple copies of the target polypeptide, or any combination thereof. In some embodiments the donor/insert polynucleotide contains a marker, barcode, or other identifier. In some embodiments, such marker, barcode, or other identifier can facilitate downstream screening for e.g., confirmation of insertion. Suitable markers, barcodes, or other identifiers are described in greater detail elsewhere herein and are generally known in the art.

[0122] In some embodiments, a double stranded donor/insert polynucleotide has one or more overhanging ends. In some embodiments, a double stranded donor/insert polynucleotide has a 5', a 3', or both a 5' and a 3' overhanging end(s). In some embodiments the overhanging ends can be composed of 1 to/or 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 or more nucleotides. In some embodiments the overhangs are in whole or at least in part complimentary to a splint or bridge polynucleotide, one or more overhangs produced by a double stranded break or nicking of a target and/or non-target strand

in a target polynucleotide, and/or a “flap” in a non-target or non-target strand of a target polynucleotide.

[0123] In some embodiments, a recombination template is designed 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.

[0124] In an embodiment, the template nucleic acid alters the sequence of the target position. In an embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

[0125] The template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by a Cas mediated cleavage event. In an embodiment, the template nucleic acid may include sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas mediated event, and a second site on the target sequence that is cleaved in a second Cas mediated event.

[0126] In certain embodiments, the 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 certain embodiments, the 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.

[0127] A 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 template sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The 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.

[0128] The 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.

[0129] A 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 an embodiment, the template nucleic acid may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 110+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 180+/- 10, 190+/- 10, 200+/- 10, 210+/-10, or 220+/- 10 nucleotides in length. In an embodiment, the template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/- 20, 80+/-20, 90+/-20, 100+/-20, 110+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, or 220+/-20 nucleotides in length. In an embodiment, the 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.

[0130] In some embodiments, the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a 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 some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the 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.

[0131] The exogenous polynucleotide template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides 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.

[0132] 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.

[0133] 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 pb.

[0134] In certain embodiments, 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 some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[0135] In some methods, the exogenous polynucleotide template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

[0136] In certain embodiments, 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.

[0137] Suzuki et al. describe in vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration (2016, Nature 540:144–149). Li et al. describe how Cas12a mediates efficient and precise endogenous gene tagging via MITI: microhomology-dependent targeted integrations (Cell Mol Life Sci. (2020) 77(19):3875-3884).

Attachment of Donor Polynucleotide(s) to a Cas Protein

[0138] In some embodiments, the donor/insert polynucleotide is directly attached to or coupled to via a linker to a Cas of the CRISPR-Cas system (including but not limited to a Cas-associated ligase). As used herein, unless otherwise indicated, “attached” refers to covalent or non-covalent interaction between two or more molecules. Non-covalent interactions can include ionic bonds, electrostatic interactions, van der Waals forces, dipole-dipole interactions, dipole-induced-dipole interactions, London dispersion forces, hydrogen bonding, halogen

bonding, electromagnetic interactions, π - π interactions, cation- π interactions, anion- π interactions, polar π -interactions, and hydrophobic effects. In some embodiments, the attachment is a covalent attachment. In some embodiments, the attachment is a non-covalent attachment. In some embodiments, the donor/insert polynucleotide can be attached via chemical linker such as any of those described in e.g., International Application Publication WO 2019135816, which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. In some embodiments, a linker or other tether can be used to couple the donor polynucleotide to a Cas protein or other CRISPR-Cas system component. In some embodiments, attachment (direct or via a linker or other tether) occurs at one or more sites in the Cas protein, such as any of those expressed in or homologous to those FIG. 15A of International Application Publication WO 2019135816. In some embodiments, attachment (direct or via a linker or other tether) of the donor polynucleotide is at any one or more residues E1207, S1154, S1116, S355, E471, E1068, E945, E1026, Q674, E532, K558, S204, Q826, D435, S867 relative to a Cas9 or a homologue thereof in another Cas protein (e.g., analogous residues in a Type V Cas protein).

Attachment through an HUH Endonuclease

[0139] In some embodiments, donor/insert polynucleotides, e.g., single-stranded oligodeoxynucleotide (ssODN) donor sequences or double-stranded oligodeoxynucleotide (dsODN) donor sequences can be conjugated or linked or attached to a Cas protein via a covalent link to HUH endonucleases which is/are fused to the Cas protein. It has recently been shown that HUH endonucleases can form robust covalent bonds with specific sequences of unmodified single-stranded DNA (ssDNA) and can function in fusion tags with diverse protein partners, including Cas9 and Cas12. *See, e.g.,* Aird et al. Communications Biology. 1 (1): 54; and Lovendahl, Klaus N.; Hayward, Amanda N.; Gordon, Wendy R. (2017-05-24). "Sequence-Directed Covalent Protein–DNA Linkages in a Single Step Using HUH-Tags". Journal of the American Chemical Society. 139 (20): 7030–7035; Nagy et al. "Site-directed integration of exogenous DNA into the soybean genome by LbCas12a fused to a plant viral HUH endonuclease." Plant J. (2022) 111(3):905-916, which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. Formation of a phosphotyrosine bond between ssDNA and HUH

endonucleases occurs within minutes at room temperature. Tethering the donor DNA template to Cas12, or other Cas protein utilizing an HUH endonuclease can, without being bound by theory, create a stable covalent RNP-donor (e.g., ssODN) complex without the need for chemical modification of the donor polynucleotide (e.g., ssODN), alteration of the sgRNA, or additional proteins. In the present invention, dsOND and/or ssODN donor sequences can be covalently-tethered via HUH-Cas (e.g., HUH-Cas12, or the like). In some embodiments, the donor polynucleotide is covalently tethered to an HUH-Cas-associated ligase.

[0140] In some embodiments, the HUH endonuclease fused to, coupled to, or otherwise associated with a Cas protein is a PCV2 rep protein (*See, e.g.,* Aird et al. Communications Biology. 1 (1): 54), MobA relaxase (Zdechlik, et al. Bioconjugate Chemistry. 31 (4): 1093–1106), TrwC, TraI (Guo et al., nanotechnology. 31(5):255102 or a combination thereof). These preceding references are incorporated by reference herein in their entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0141] An exemplary construct design for a PCV based approach is as follows. In some embodiments, a Cas protein can be amplified and inserted in a plasmid containing a sequence encoding for Porcine Circovirus 2 (PCV) Rep protein. For example, a Cas12 protein (e.g., a *Lachnospiraceae bacterium* Cas12a (LbCas12a), an *Acidaminococcus* sp. (AsCas12a), a *Francisella novicida* Cas12a (FnCas12a), a *Moraxella bovoculi* Cas12a (MbCas12), or an *Eubacterium rectale* Cas12a (MAD7), AsCas12a or LbCas12a can be amplified and inserted in a plasmid containing sequence encoding for Porcine Circovirus 2 (PCV) Rep protein. An exemplary plasmid is pTD68_SUMO-PCV2. Other plasmids that containing a PCV2 coding sequencing can also be used for this purpose. In some embodiments, the PCV2 sequence is at the C-terminal of a Cas protein to create Cas-PCV fusion protein. In some embodiments, the PCV2 sequence is at the N-terminal of a Cas protein to create PCV-Cas fusion protein. Catalytically dead Cas protein, for example, Cas12-PCV can be created by Quik-Change II site directed mutagenesis kit (Agilent Technologies).

[0142] Exemplary covalent attachment of a donor polynucleotide to a PCV-Cas protein is as follows. In some embodiments, covalent DNA attachment to Cas-PCV can be achieved by adding equimolar amounts of Cas-PCV and the sequence specific dsODN or ssODN and incubating at room temperature for 10-15 min in Opti-MEM (Corning) culture medium supplemented with 1mM MgCl₂. Confirmation of the linkage can be obtained by analyzing

using SDS-PAGE. For the fluorescent oligonucleotide reactions, 1.5 pmol of Alexa 488-conjugated dsODN or ssODN (IDT) can be incubated with 1.5 pmol Cas-PCV in the above conditions and separated by SDS-PAGE. Gels can be imaged using a 473 nm laser excitation on a Typhoon FLA9500 (GE).

[0143] An exemplary cleavage assay is as follows. A pcDNA3-eGFP vector or pcDNA5-GAPDH vector is linearized with BsaI or BspQI (NEB), respectively, and column purified. A concentration of 30 nM sgRNA, 30 nM Cas12 or other Cas protein, and 1x T4 ligase buffer are incubated for 10 min prior to adding linearized DNA to a final concentration of 3 nM. The reaction is incubated at 37 °C for 1 to 24 h, then separated by agarose gel electrophoresis and imaged using SYBR safe gel stain (Thermo Fisher). The percent cleaved is calculated by comparing densities of the uncleaved band and the top cleaved band using Image Lab software (Bio-Rad).

[0144] In some embodiments, the donor/insert polynucleotide is hybridized or otherwise complexed with one or more components of a CRISPR-Cas system e.g., hybridized to guide RNAs) immediately prior to delivery of the complex to e.g., a cell, or other vessel in which a target polynucleotide is present or potentially present. In some embodiments, the donor/insert polynucleotides are delivered separately (physically, spatially, and/or temporally) from the other components of a CRISPR-Cas system disclosed herein (including but not limited to a Cas protein, guide RNAs, or others). Such separation can allow for, among other things, control over the activity of the system. In some embodiments, the donor/insert polynucleotide is delivered 1-48 hours after delivery of a CRISPR-Cas system or encoding polynucleotide or vector.

[0145] In some embodiments, the donor/insert polynucleotide is configured to promote one DSB repair pathway over another. In some embodiments, the donor/insert polynucleotide is configured to promote HDR. In some embodiments, the donor/insert polynucleotide is attached to one or more HDR activators and/or NEHJ inhibitors. Attachment can be via a linker. Exemplary HDR activators and/or NEHJ inhibitors are described in greater detail elsewhere herein.

POLYNUCLEOTIDES ENCODING ENGINEERED TYPE V CAS POLYPEPTIDES

[0146] In another aspects, disclosed herein are directed to polynucleotides encoding the engineered reduced immunogenicity Type V Cas polypeptides, and CRISPR-Complexes described above. The polynucleotides may comprise coding sequences of one or more of

components of the CRISPR-Cas systems (e.g., one or more of the engineered Type V Cas polynucleotides, the at least one guide RNA, the donor/insert polynucleotides, and/or other components in the systems) as disclosed herein. Described in several example embodiments herein are nucleic acid molecules that encode one or more of the engineered Type V Cas polypeptides or other system polypeptides (e.g., heterologous functional domains) and/or guide molecules (e.g., guide RNAs) and/or donor polynucleotides and/or other CRISPR components. In certain example embodiments, a nucleic acid molecule encodes the engineered Type V Cas polypeptide (e.g., a Cas12 polypeptide, or domain or fragment thereof having Cas12 activity). In certain example embodiments, a single nucleic acid molecule encodes each of the engineered Type V Cas polypeptide and at least one guide RNA. In certain example embodiments, a single nucleic acid molecule encodes each of the engineered Type V Cas polypeptide, the at least one guide RNA, and the one or more donor polynucleotides.

[0147] Nucleic acid molecules may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of nucleic acid molecules: 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, which are incorporated by reference herein in their entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. A nucleic acid molecule 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 nucleic acid molecule may be further modified after polymerization, such as by conjugation with a labeling component. As used herein, unless otherwise indicated, 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.

[0148] In one embodiment, the nucleic acid molecule sequence is recombinant DNA. In further embodiments, the nucleic acid molecule sequence further comprises additional sequences as described elsewhere herein. In one embodiment, the nucleic acid sequence is synthesized in vitro.

[0149] The present disclosure provides nucleic acid molecules that encode one or more components of any of the engineered Type V Cas polypeptides or systems as referred to in any embodiment herein. In one embodiment, the nucleic acid molecules may comprise further regulatory sequences. By means of guidance and not limitation, the nucleic acid molecule 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 one embodiment, the nucleic acid molecule sequence may be a bicistronic expression construct. In further embodiments, the isolated nucleic acid molecule sequence may be incorporated in a cellular genome. In yet further embodiments, the isolated nucleic acid molecule sequence may be part of a cellular genome. In further embodiments, the isolated nucleic acid molecule sequence may be comprised in an artificial chromosome. In one embodiment, the 5' and/or 3' end of the isolated nucleic acid molecule sequence may be modified to improve the stability of the sequence or actively avoid degradation. In one embodiment, the isolated nucleic acid molecule sequence may be comprised in a bacteriophage. In other embodiments, the isolated nucleic acid molecule sequence may be contained in agrobacterium species. In one embodiment, the isolated nucleic acid molecule sequence is lyophilized.

Codon Optimization

[0150] Aspects of the invention relate to nucleic acid molecule that encode one or more components of one or more systems as described in any of the embodiments disclosed herein, wherein at least one or more regions of the nucleic acid molecule may be codon optimized for expression in eukaryotic cells. In one embodiment, the nucleic acid molecules that encode one or more components of one or more systems as described in any of the embodiments disclosed herein are optimized for expression in a mammalian cell or a plant cell.

[0151] 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 Type V Cas 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), which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein. 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 an engineered Type V Cas polypeptide corresponds to the most frequently used codon for a particular amino acid.

DELIVERY SYSTEMS

[0152] The present disclosure also provides delivery systems for introducing the engineered Type V Cas polypeptides, systems, and/or compositions disclosed 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.

[0153] In some embodiments, 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.

Cargos

[0154] The delivery systems may comprise one or more cargos. The cargos may comprise one or more components of the CRISPR-Cas systems and compositions herein. A cargo may comprise one or more CRISPR-Cas system components (e.g., proteins, guide molecules, donor polynucleotides, RNP complexes of same, or nucleic acids encoding same), vectors/systems thereof, mRNA thereof, RNP complexes thereof, or plasmids thereof. A cargo may comprise one or more of the following: i) a vector or vector system (viral or non-viral) comprising one or more nucleic acid molecules encoding one or more Cas proteins; ii) a vector or vector system (viral or non-viral) comprising one or more nucleic acid molecules encoding one or more guide molecules described herein, iii) mRNA of one or more Cas proteins; iv) one or more guide molecules; v) one or more Cas proteins; vi) one or more polynucleotides encoding one or more Cas proteins; vii) one or more polynucleotides encoding one or more guide molecules, or viii) any combination thereof. In some examples, a cargo may comprise a plasmid encoding one or more Cas proteins and one or more (e.g., a plurality of) guide molecules. In some embodiments,

a cargo may comprise mRNA encoding one or more Cas proteins and one or more guide molecules.

[0155] In some embodiments, a cargo may comprise one or more Cas proteins described herein and one or more guide 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.

[0156] In some embodiments, the cargo(s) can be any of the polynucleotide(s), e.g., CRISPR-Cas System polynucleotides, described herein.

Physical Delivery

[0157] In some embodiments, 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, Cas protein may be prepared in vitro, isolated, (refolded, purified if needed), and introduced to cells.

Microinjection

[0158] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In some embodiments, 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.

[0159] Plasmids comprising coding sequences for Cas proteins and/or guide RNAs, mRNAs, and/or guide 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 sgRNA directly to the nucleus and Cas-encoding mRNA to the cytoplasm, e.g., facilitating translation and shuttling of Cas to the nucleus.

[0160] 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 CRISPRa and CRISPRi.

Electroporation

[0161] In some embodiments, 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.

[0162] 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

[0163] 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

[0164] The cargos, e.g., nucleic acids and/or polypeptides, 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.

Transduction

[0165] The cargos, e.g. nucleic acids and/or polypeptides, can be introduced to cells by transduction by a viral or pseudoviral particle. Methods of packaging the cargos in viral particles can be accomplished using any suitable viral vector or vector systems. Such viral vector and vector systems are described in greater detail elsewhere herein. As used in this context herein “transduction” refers to the process by which foreign nucleic acids and/or proteins are introduced to a cell (prokaryote or eukaryote) by a viral or pseudo viral particle. After packaging in a viral particle or pseudo viral particle, the viral particles can be exposed to cells (e.g. in vitro, ex vivo, or in vivo) where the viral or pseudoviral particle infects the cell and delivers the cargo to the cell via transduction. Viral and pseudoviral particles can be optionally concentrated prior to exposure to target cells. In some embodiments, the virus titer of a composition containing viral and/or pseudoviral particles can be obtained and a specific titer be used to transduce cells.

Biolistics

[0166] The cargos, e.g. nucleic acids and/or polypeptides, can be introduced to cells using a biolistic method or technique. The term of art “biolistic”, as used herein refers to the delivery of nucleic acids to cells by high-speed particle bombardment. In some embodiments, the cargo(s) can be attached, associated with, or otherwise coupled to particles, which than can be delivered to the cell via a gene-gun (see e.g., Liang et al. 2018. Nat. Protocol. 13:413-430; Svitashv et al. 2016. Nat. Comm. 7:13274; Ortega-Escalante et al., 2019. Plant. J. 97:661-672). In some embodiments, the particles can be gold, tungsten, palladium, rhodium, platinum, or iridium particles.

Implantable Devices

[0167] In some embodiments, the delivery system can include an implantable device that incorporates or is coated with a CRISPR-Cas system or component thereof described herein. Various implantable devices are described in the art, and include any device, graft, or other composition that can be implanted into a subject.

Delivery Vehicles

[0168] 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 (e.g. virus particles), non-viral vehicles, and other delivery reagents described herein.

[0169] The delivery vehicles described herein can have a greatest dimension or greatest average dimension (e.g., diameter or greatest average diameter) of less than 100 microns (μm). In some embodiments, the delivery vehicles have a greatest dimension or greatest average dimension of less than 10 μm . In some embodiments, the delivery vehicles may have a greatest dimension or greatest average dimension of less than 2000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension or greatest average dimension of less than 1000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension or greatest average dimension (e.g., diameter or average 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 some embodiments, the delivery vehicles may have a greatest dimension or greatest average dimension ranging between 25 nm and 200 nm.

Vectors and Vector systems

[0170] Also provided herein are vectors that can contain one or more of the CRISPR-Cas system nucleic acid molecules described herein. Described in several example embodiments herein are vectors comprising nucleic acid molecules that encode one or more of the engineered Type V Cas polypeptides or other system polypeptides (e.g., heterologous functional domains) and/or guide molecules (e.g., guide RNAs) and/or donor polynucleotides and/or other CRISPR components. In certain example embodiments, a vector comprises a nucleic acid molecule encoding the engineered Type V Cas polypeptide (e.g., a Cas12 polypeptide, or domain or fragment thereof having Cas12 activity). In certain example embodiments, a single vector comprises one or more nucleic acid molecules (e.g., a single nucleic acid molecule) encoding each of the engineered Type V Cas polypeptide and at least one guide RNA. In certain example embodiments, a single vector comprises one or more nucleic acid molecules (e.g., a single nucleic acid molecule) encoding each of the engineered Type V Cas polypeptide, the at least

one guide RNA, and the one or more donor polynucleotides. In certain example embodiments, the vectors are viral vectors.

[0171] The vectors can be useful in producing bacterial, fungal, yeast, plant cells, animal cells, and transgenic animals that can express one or more components of the CRISPR-Cas system described herein. Within the scope of this disclosure are vectors containing one or more of the polynucleotide sequences described herein. One or more of the polynucleotides that are part of the CRISPR-Cas system described herein can be included in a vector or vector system. The vectors and/or vector systems can be used, for example, to express one or more of the polynucleotides in a cell, such as a producer cell, to produce CRISPR-Cas system containing virus particles described elsewhere herein. Other uses for the vectors and vector systems described herein are also within the scope of this disclosure. In general, and throughout this specification, the term “vector” refers to a tool that allows or facilitates the transfer of an entity from one environment to another. In some contexts which will be appreciated by those of ordinary skill in the art, “vector” can be a term of art to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A vector can be a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements.

[0172] Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a “plasmid,” which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.

Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as “expression vectors.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0173] Recombinant expression vectors can be composed of a nucleic acid (e.g. a polynucleotide) of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which can be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operably linked” and “operatively-linked” are used interchangeably herein and further defined elsewhere herein. In the context of a vector, 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). Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells. These and other embodiments of the vectors and vector systems are described elsewhere herein.

[0174] In some embodiments, the vector can be a bicistronic vector. In some embodiments, a bicistronic vector can be used for one or more elements of the CRISPR-Cas system described herein. In some embodiments, expression of elements of the CRISPR-Cas system described herein can be driven by the CBh promoter or other ubiquitous promoter. Where the element of the CRISPR-Cas system is an RNA, its expression can be driven by a Pol III promoter, such as a U6 promoter. In some embodiments, the two are combined.

[0175] In some embodiments, a vector capable of delivering an effector protein and optionally at least one CRISPR guide RNA to a cell can be composed of or contain a minimal promoter operably linked to a polynucleotide sequence encoding the effector protein and a second minimal promoter operably linked to a polynucleotide sequence encoding at least one guide RNA, wherein the length of the vector sequence comprising the minimal promoters and polynucleotide sequences is less than 4.4Kb. In an embodiment, the vector can be a viral vector. In certain embodiments, the viral vector is an adeno-associated virus (AAV) or an adenovirus vector. In another embodiment, the effector protein is a Cas protein. In a further embodiment, the CRISPR enzyme is Cas12.

[0176] In some embodiments, the vector capable of delivering a lentiviral vector for an effector protein and at least one CRISPR guide RNA to a cell can be composed of or contain a promoter operably linked to a polynucleotide sequence encoding Cas and a second promoter operably linked to a polynucleotide sequence encoding at least one guide RNA, wherein the polynucleotide sequences are in reverse orientation.

[0177] In one embodiment, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the CRISPR complex to the one or more target sequence(s) in a eukaryotic cell, wherein the CRISPR complex comprises a Cas enzyme complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas enzyme, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR complex comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said Cas CRISPR complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[0178] These and others are further detailed and described elsewhere herein.

Cell-based Vector Amplification and Expression

[0179] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into

a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). The vectors can be viral-based or non-viral based. In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism.

[0180] Vectors can be designed for expression of one or more elements of the CRISPR-Cas system described herein (e.g. nucleic acid transcripts, proteins, enzymes, and combinations thereof) in a suitable host cell. In some embodiments, the suitable host cell is a prokaryotic cell. Suitable host cells include, but are not limited to, bacterial cells, yeast cells, insect cells, and mammalian cells. In some embodiments, the suitable host cell is a eukaryotic cell.

[0181] In some embodiments, the suitable host cell is a suitable bacterial cell. Suitable bacterial cells include, but are not limited to, bacterial cells from the bacteria of the species *Escherichia coli*. Many suitable strains of *E. coli* are known in the art for expression of vectors. These include, but are not limited to Pir1, Stbl2, Stbl3, Stbl4, TOP10, XL1 Blue, and XL10 Gold. In some embodiments, the host cell is a suitable insect cell. Suitable insect cells include those from *Spodoptera frugiperda*. Suitable strains of *S. frugiperda* cells include, but are not limited to, Sf9 and Sf21. In some embodiments, the host cell is a suitable yeast cell. In some embodiments, the yeast cell can be from *Saccharomyces cerevisiae*. In some embodiments, the host cell is a suitable mammalian cell. Many types of mammalian cells have been developed to express vectors. Suitable mammalian cells include, but are not limited to, HEK293, Chinese Hamster Ovary Cells (CHOs), mouse myeloma cells, HeLa, U2OS, A549, HT1080, CAD, P19, NIH 3T3, L929, N2a, MCF-7, Y79, SO-Rb50, HepG G2, DIKX-X11, J558L, Baby hamster kidney cells (BHK), and chicken embryo fibroblasts (CEFs). Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990).

[0182] In some embodiments, the vector can be a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the

expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Yeast vectors can contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2 μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[0183] In some embodiments, the vector is a baculovirus vector or expression vector and can be suitable for expression of polynucleotides and/or proteins in insect cells. In some embodiments, the suitable host cell is an insect cell. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39). rAAV (recombinant Adeno-associated viral) vectors are preferably produced in insect cells, e.g., *Spodoptera frugiperda* Sf9 insect cells, grown in serum-free suspension culture. Serum-free insect cells can be purchased from commercial vendors, e.g., Sigma Aldrich (EX-CELL 405).

[0184] In some embodiments, the vector is a mammalian expression vector. In some embodiments, the mammalian expression vector is capable of expressing one or more polynucleotides and/or polypeptides in a mammalian cell. Examples of mammalian expression vectors include, but are not limited to, pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). The mammalian expression vector can include one or more suitable regulatory elements capable of controlling expression of the one or more polynucleotides and/or proteins in the mammalian cell. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. More detail on suitable regulatory elements are described elsewhere herein.

[0185] For other suitable expression vectors and vector systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING:

A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0186] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments can utilize viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety. In some embodiments, a regulatory element can be operably linked to one or more elements of a CRISPR-Cas system so as to drive expression of the one or more elements of the CRISPR-Cas system described herein.

[0187] In some embodiments, the vector can be a fusion vector or fusion expression vector. In some embodiments, fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus, carboxy terminus, or both of a recombinant protein. Such fusion vectors can serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. In some embodiments, expression of polynucleotides (such as non-coding polynucleotides) and

proteins in prokaryotes can be carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polynucleotides and/or proteins. In some embodiments, the fusion expression vector can include a proteolytic cleavage site, which can be introduced at the junction of the fusion vector backbone or other fusion moiety and the recombinant polynucleotide or protein to enable separation of the recombinant polynucleotide or protein from the fusion vector backbone or other fusion moiety subsequent to purification of the fusion polynucleotide or protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY* 185, Academic Press, San Diego, Calif. (1990) 60-89).

[0188] In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR-Cas system described herein are introduced into a host cell such that expression of the elements of the engineered delivery system described herein direct formation a CRISPR-Cas complex at one or more target sites. For example, a CRISPR-Cas effector protein described herein and a nucleic acid component (e.g., a guide polynucleotide) can each be operably linked to separate regulatory elements on separate vectors. RNA(s) of different elements of CRISPR-Cas system described herein can be delivered to an animal, plant, microorganism or cell thereof to produce an animal (e.g., a mammal, reptile, avian, etc.), plant, microorganism or cell thereof that constitutively, inducibly, or conditionally expresses different elements of the CRISPR-Cas system described herein that incorporates one or more elements of the CRISPR-Cas system described herein or contains one or more cells that incorporates and/or expresses one or more elements of the CRISPR-Cas system described herein.

[0189] In some embodiments, two or more of the elements expressed from the same or different regulatory element(s), can be combined in a single vector, with one or more additional vectors providing any components of the system not included in the first vector. CRISPR-Cas system polynucleotides that are combined in a single vector may be arranged in any suitable

orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding one or more CRISPR-Cas system proteins, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR-Cas system polynucleotides can be operably linked to and expressed from the same promoter.

Cell-Free Vector and Polynucleotide Expression

[0190] In some embodiments, the polynucleotide encoding one or more features of the CRISPR-Cas system can be expressed from a vector or suitable polynucleotide in a cell-free in vitro system. In other words, the polynucleotide can be transcribed and optionally translated in vitro. In vitro transcription/translation systems and appropriate vectors are generally known in the art and commercially available. Generally, in vitro transcription and in vitro translation systems replicate the processes of RNA and protein synthesis, respectively, outside of the cellular environment. Vectors and suitable polynucleotides for in vitro transcription can include T7, SP6, T3, promoter regulatory sequences that can be recognized and acted upon by an appropriate polymerase to transcribe the polynucleotide or vector.

[0191] In vitro translation can be stand-alone (e.g. translation of a purified polyribonucleotide) or linked/coupled to transcription. In some embodiments, the cell-free (or in vitro) translation system can include extracts from rabbit reticulocytes, wheat germ, and/or E. coli. The extracts can include various macromolecular components that are needed for translation of exogenous RNA (e.g. 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA, synthetases, initiation, elongation factors, termination factors, etc.). Other components can be included or added during the translation reaction, including but not limited to, amino acids, energy sources (ATP, GTP), energy regenerating systems (creatine phosphate and creatine phosphokinase (eukaryotic systems)) (phosphoenol pyruvate and pyruvate kinase for bacterial systems), and other co-factors (Mg²⁺, K⁺, etc.). As previously mentioned, in vitro translation can be based on RNA or DNA starting material. Some translation systems can utilize an RNA template as starting material (e.g. reticulocyte lysates and wheat germ extracts). Some translation systems can utilize a DNA template as a starting material (e.g. E coli-based systems). In these systems transcription and translation are coupled and DNA is first

transcribed into RNA, which is subsequently translated. Suitable standard and coupled cell-free translation systems are generally known in the art and are commercially available.

Vector Features

[0192] The vectors can include additional features that can confer one or more functionalities to the vector, the polynucleotide to be delivered, a virus particle produced therefrom, or polypeptide expressed thereof. Such features include, but are not limited to, regulatory elements, selectable markers, molecular identifiers (e.g. molecular barcodes), stabilizing elements, and the like. It will be appreciated by those skilled in the art that the design of the expression vector and additional features included can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc.

Regulatory Elements

[0193] In certain embodiments, the polynucleotides and/or vectors thereof described herein (such as the CRISPR-Cas system polynucleotides of the present invention) can include one or more regulatory elements that can be operatively linked to the polynucleotide. The term “regulatory element” is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences) and cellular localization signals (e.g. nuclear localization signals). 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 can direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and 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) (see, e.g., Boshart et al, Cell, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. Also encompassed by the term “regulatory element” are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981).

[0194] In some embodiments, the regulatory sequence can be a regulatory sequence described in U.S. Pat. No. 7,776,321, U.S. Pat. Pub. No. 2011/0027239, and International Patent Publication No. WO 2011/028929, the contents of which are incorporated by reference herein in their entirety. In some embodiments, the vector can contain a minimal promoter. In some embodiments, the minimal promoter is the Mecp2 promoter, tRNA promoter, or U6. In a further embodiment, the minimal promoter is tissue specific. In some embodiments, the length of the vector polynucleotide the minimal promoters and polynucleotide sequences is less than 4.4Kb.

[0195] To express a polynucleotide, the vector can include one or more transcriptional and/or translational initiation regulatory sequences, e.g. promoters, that direct the transcription of the gene and/or translation of the encoded protein in a cell. In some embodiments a constitutive promoter may be employed. Suitable constitutive promoters for mammalian cells are generally known in the art and include, but are not limited to SV40, CAG, CMV, EF-1 α , β -actin, RSV, and PGK. Suitable constitutive promoters for bacterial cells, yeast cells, and fungal cells are generally known in the art, such as a T-7 promoter for bacterial expression and an alcohol dehydrogenase promoter for expression in yeast.

[0196] In some embodiments, the regulatory element can be a regulated promoter. "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Regulated promoters include conditional promoters and inducible promoters. In some embodiments, conditional promoters can be employed to direct expression of a polynucleotide in a specific cell type, under certain environmental conditions, and/or during a specific state of development. Suitable tissue specific promoters can include, but are not limited to, liver specific promoters (e.g. APOA2, SERPIN A1 (hAAT), CYP3A4, and MIR122), pancreatic cell promoters (e.g. INS, IRS2, Pdx1, Alx3, Ppy), cardiac specific

promoters (e.g. Myh6 (alpha MHC), MYL2 (MLC-2v), TNI3 (cTnl), NPPA (ANF), Slc8a1 (Ncx1)), central nervous system cell promoters (SYN1, GFAP, INA, NES, MOBP, MBP, TH, FOXA2 (HNF3 beta)), skin cell specific promoters (e.g. FLG, K14, TGM3), immune cell specific promoters, (e.g. ITGAM, CD43 promoter, CD14 promoter, CD45 promoter, CD68 promoter), urogenital cell specific promoters (e.g. Pbsn, Upk2, Sbp, Fer114), endothelial cell specific promoters (e.g. ENG), pluripotent and embryonic germ layer cell specific promoters (e.g. Oct4, NANOG, Synthetic Oct4, T brachyury, NES, SOX17, FOXA2, MIR122), and muscle cell specific promoter (e.g. Desmin). Other tissue and/or cell specific promoters are generally known in the art and are within the scope of this disclosure.

[0197] Inducible/conditional promoters can be positively inducible/conditional promoters (e.g. a promoter that activates transcription of the polynucleotide upon appropriate interaction with an activated activator, or an inducer (compound, environmental condition, or other stimulus) or a negative/conditional inducible promoter (e.g. a promoter that is repressed (e.g. bound by a repressor) until the repressor condition of the promoter is removed (e.g. inducer binds a repressor bound to the promoter stimulating release of the promoter by the repressor or removal of a chemical repressor from the promoter environment). The inducer can be a compound, environmental condition, or other stimulus. Thus, inducible/conditional promoters can be responsive to any suitable stimuli such as chemical, biological, or other molecular agents, temperature, light, and/or pH. Suitable inducible/conditional promoters include, but are not limited to, Tet-On, Tet-Off, Lac promoter, pBad, AlcA, LexA, Hsp70 promoter, Hsp90 promoter, pDawn, XVE/OlexA, GVG, and pOp/LhGR.

[0198] Where expression in a plant cell is desired, the components of the CRISPR-Cas system described herein are typically placed under control of a plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged.

[0199] A constitutive plant promoter is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One non-limiting example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the CRISPR-Cas system components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter issue-

preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the CRISPR-Cas system are found in Kawamata et al., (1997) *Plant Cell Physiol* 38:792-803; Yamamoto et al., (1997) *Plant J* 12:255-65; Hire et al, (1992) *Plant Mol Biol* 20:207-18, Kuster et al, (1995) *Plant Mol Biol* 29:759-72, and Capana et al., (1994) *Plant Mol Biol* 25:681 -91.

[0200] Examples of promoters that are inducible and that can allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome), such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include one or more elements of the CRISPR-Cas system described herein, a light-responsive cytochrome heterodimer (e.g. from *Arabidopsis thaliana*), and a transcriptional activation/repression domain. In some embodiments, the vector can include one or more of the inducible DNA binding proteins provided in International Patent Publication No. WO 2014/018423 and US Patent Publication Nos., 2015/0291966, 2017/0166903, 2019/0203212, which describe e.g. embodiments of inducible DNA binding proteins and methods of use and can be adapted for use with the present invention.

[0201] In some embodiments, transient or inducible expression can be achieved by including, for example, chemical-regulated promoters, i.e. whereby the application of an exogenous chemical induces gene expression. Modulation of gene expression can also be obtained by including a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize *ln2-2* promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) *Plant Cell Physiol* 38:568-77), the maize *GST* promoter (*GST-ll-27*, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco *PR-1* a promoter (Ono et al., (2004) *Biosci Biotechnol Biochem* 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible

and tetracycline-repressible promoters (Gatz et al., (1991) Mol Gen Genet 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

[0202] In some embodiments, the polynucleotide, vector or system thereof can include one or more elements capable of translocating and/or expressing a CRISPR-Cas polynucleotide to/in a specific cell component or organelle. Such organelles can include, but are not limited to, nucleus, ribosome, endoplasmic reticulum, Golgi apparatus, chloroplast, mitochondria, vacuole, lysosome, cytoskeleton, plasma membrane, cell wall, peroxisome, centrioles, etc. Such regulatory elements can include, but are not limited to, nuclear localization signals (examples of which are described in greater detail elsewhere herein), any such as those that are annotated in the LocSigDB database (see e.g. <http://genome.unmc.edu/LocSigDB/> and Negi et al., 2015. Database. 2015: bav003; doi: 10.1093/database/bav003), nuclear export signals (e.g. LXXXLXXLXL (SEQ ID NO: 42) and others described elsewhere herein), endoplasmic reticulum localization/retention signals (e.g. KDEL (SEQ ID NO: 43), KDXX, KKXX, KXX, and others described elsewhere herein; and see e.g. Liu et al. 2007 Mol. Biol. Cell. 18(3):1073-1082 and Gorleku et al., 2011. J. Biol. Chem. 286:39573-39584), mitochondria (see e.g. Cell Reports. 22:2818-2826, particularly at Fig. 2; Doyle et al. 2013. PLoS ONE 8, e67938; Funes et al. 2002. J. Biol. Chem. 277:6051-6058; Matouschek et al. 1997. PNAS USA 85:2091-2095; Oca-Cossio et al., 2003. 165:707-720; Waltner et al., 1996. J. Biol. Chem. 271:21226-21230; Wilcox et al., 2005. PNAS USA 102:15435-15440; Galanis et al., 1991. FEBS Lett 282:425-430, peroxisome (e.g. (S/A/C)-(K/R/H)-(L/A), SLK, (R/K)-(L/V/I)-XXXXXX-(H/Q)-(L/A/F). Suitable protein targeting motifs can also be designed or identified using any suitable database or prediction tool, including but not limited to Minimotif Miner (<http://minimotifminer.org>, <http://KDLmitominer.mrc-mbu.cam.ac.uk/release-4.0/embodiment.do?name=Protein%20MTS>), LocDB (see above), PTSs predictor (), TargetP-2.0 (<http://www.cbs.dtu.dk/services/TargetP/>), ChloroP (<http://www.cbs.dtu.dk/services/ChloroP/>), NetNES (<http://www.cbs.dtu.dk/services/NetNES/>), Predotar (<https://urgi.versailles.inra.fr/predotar/>), and SignalP (<http://www.cbs.dtu.dk/services/SignalP/>).

Selectable Markers and Tags

[0203] One or more of the CRISPR-Cas system polynucleotides can be operably linked, fused to, or otherwise modified to include a polynucleotide that encodes or is a selectable marker or tag, which can be a polynucleotide or polypeptide. In some embodiments, the

polypeptide encoding a polypeptide selectable marker can be incorporated in the CRISPR-Cas system polynucleotide such that the selectable marker polypeptide, when translated, is inserted between two amino acids between the N- and C- terminus of the CRISPR-Cas system polypeptide or at the N- and/or C-terminus of the CRISPR-Cas system polypeptide. In some embodiments, the selectable marker or tag is a polynucleotide barcode or unique molecular identifier (UMI).

[0204] It will be appreciated that the polynucleotide encoding such selectable markers or tags can be incorporated into a polynucleotide encoding one or more components of the CRISPR-Cas system described herein in an appropriate manner to allow expression of the selectable marker or tag. Such techniques and methods are described elsewhere herein and will be instantly appreciated by one of ordinary skill in the art in view of this disclosure. Many such selectable markers and tags are generally known in the art and are intended to be within the scope of this disclosure.

[0205] Suitable selectable markers and tags include, but are not limited to, affinity tags, such as chitin binding protein (CBP), maltose binding protein (MBP), glutathione-S-transferase (GST), poly(His) tag; solubilization tags such as thioredoxin (TRX) and poly(NANP), MBP, and GST; chromatography tags such as those consisting of polyanionic amino acids, such as FLAG-tag; epitope tags such as V5-tag, Myc-tag, HA-tag and NE-tag; protein tags that can allow specific enzymatic modification (such as biotinylation by biotin ligase) or chemical modification (such as reaction with FLAsH-EDT2 for fluorescence imaging), DNA and/or RNA segments that contain restriction enzyme or other enzyme cleavage sites; DNA segments that encode products that provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO), hygromycin phosphotransferase (HPT)) and the like; DNA and/or RNA segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA and/or RNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β -galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), luciferase, and cell surface proteins); polynucleotides that can generate one or more new primer sites for PCR (e.g., the juxtaposition of two DNA sequences not previously juxtaposed), DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; epitope tags (e.g. GFP, FLAG- and His-tags), and, DNA sequences

that make a molecular barcode or unique molecular identifier (UMI), DNA sequences required for a specific modification (e.g., methylation) that allows its identification. Other suitable markers will be appreciated by those of skill in the art.

[0206] Selectable markers and tags can be operably linked to one or more components of the CRISPR-Cas system described herein via suitable linker, such as a glycine or glycine serine linkers as short as GS or GG up to (GGGGG)₃ (SEQ ID NO: 44) or (GGGGS)₃ (SEQ ID NO: 45). Other suitable linkers are described elsewhere herein.

[0207] The vector or vector system can include one or more polynucleotides encoding one or more targeting moieties. In some embodiments, the targeting moiety encoding polynucleotides can be included in the vector or vector system, such as a viral vector system, such that they are expressed within and/or on the virus particle(s) produced such that the virus particles can be targeted to specific cells, tissues, organs, etc. In some embodiments, the targeting moiety encoding polynucleotides can be included in the vector or vector system such that the CRISPR-Cas system polynucleotide(s) and/or products expressed therefrom include the targeting moiety and can be targeted to specific cells, tissues, organs, etc. In some embodiments, such as non-viral carriers, the targeting moiety can be attached to the carrier (e.g. polymer, lipid, inorganic molecule etc.) and can be capable of targeting the carrier and any attached or associated CRISPR-Cas system polynucleotide(s) to specific cells, tissues, organs, etc.

Codon Optimization of Vector Polynucleotides

[0208] As described elsewhere herein, the polynucleotide encoding one or more embodiments of the CRISPR-Cas system described herein can be codon optimized. In some embodiments, one or more polynucleotides contained in a vector (“vector polynucleotides”) described herein that are in addition to an optionally codon optimized polynucleotide encoding embodiments of the CRISPR-Cas system described herein can be codon optimized. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn

believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at www.kazusa.or.jp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid. As to codon usage in yeast, reference is made to the online Yeast Genome database available at yeastgenome.org/community/codon_usage.shtml, or *Codon selection in yeast*, Bennetzen and Hall, J Biol Chem. 1982 Mar 25;257(6):3026-31. As to codon usage in plants including algae, reference is made to *Codon usage in higher plants, green algae, and cyanobacteria*, Campbell and Gowri, Plant Physiol. 1990 Jan; 92(1): 1–11.; as well as *Codon usage in plant genes*, Murray et al, Nucleic Acids Res. 1989 Jan 25;17(2):477-98; or *Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages*, Morton BR, J Mol Evol. 1998 Apr;46(4):449-59.

[0209] The vector polynucleotide can be codon optimized for expression in a specific cell-type, tissue type, organ type, and/or subject type. In some embodiments, a codon optimized sequence is a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in a human or human cell), or for another eukaryote, such as another animal (e.g. a mammal or avian) as is described elsewhere herein. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific cell type. Such cell types can include, but are not limited to, epithelial cells (including skin cells, cells lining the gastrointestinal tract, cells lining other hollow organs), nerve cells (nerves, brain cells, spinal column cells, nerve support cells (e.g. astrocytes, glial cells, Schwann cells etc.) , muscle cells (e.g. cardiac muscle, smooth muscle cells, and skeletal muscle cells), connective tissue cells (fat and other soft tissue padding cells, bone cells, tendon cells, cartilage cells), blood cells,

stem cells and other progenitor cells, immune system cells, germ cells, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific tissue type. Such tissue types can include, but are not limited to, muscle tissue, connective tissue, connective tissue, nervous tissue, and epithelial tissue. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In some embodiments, the polynucleotide is codon optimized for a specific organ. Such organs include, but are not limited to, muscles, skin, intestines, liver, spleen, brain, lungs, stomach, heart, kidneys, gallbladder, pancreas, bladder, thyroid, bone, blood vessels, blood, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein.

[0210] In some embodiments, a vector polynucleotide is codon optimized for expression in particular cells, such as prokaryotic or 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 discussed herein, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate.

Vector Construction

[0211] The vectors described herein can be constructed using any suitable process or technique. In some embodiments, one or more suitable recombination and/or cloning methods or techniques can be used to the vector(s) described herein. Suitable recombination and/or cloning techniques and/or methods can include, but not limited to, those described in U.S. Patent Publication No. US 2004/0171156 A1. Other suitable methods and techniques are described elsewhere herein.

[0212] Construction of recombinant AAV vectors is described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989). Any of the techniques and/or methods can be used and/or adapted for constructing an AAV or other vector described herein. nAAV vectors are discussed elsewhere herein.

[0213] In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8,

9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide polynucleotides are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide s polynucleotides. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-polynucleotide-containing vectors may be provided, and optionally delivered to a cell.

[0214] Delivery vehicles, vectors, particles, nanoparticles, formulations, and components thereof for expression of one or more elements of a CRISPR-Cas system described herein are as used in the foregoing documents, such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667) and are discussed in greater detail herein.

Viral Vectors

[0215] In some embodiments, the vector is a viral vector. The term of art “viral vector” and as used herein in this context refers to polynucleotide based vectors that contain one or more elements from or based upon one or more elements of a virus that can be capable of expressing and packaging a polynucleotide, such as a CRISPR-Cas system polynucleotide of the present invention, into a virus particle and producing said virus particle when used alone or with one or more other viral vectors (such as in a viral vector system). Viral vectors and systems thereof can be used for producing viral particles for delivery of and/or expression of one or more components of the CRISPR-Cas system described herein. The viral vector can be part of a viral vector system involving multiple vectors. In some embodiments, systems incorporating multiple viral vectors can increase the safety of these systems. Suitable viral vectors can include retroviral-based vectors, lentiviral-based vectors, adenoviral-based vectors, adeno associated vectors, helper-dependent adenoviral (HdAd) vectors, hybrid adenoviral vectors, herpes simplex virus-based vectors, poxvirus-based vectors, and Epstein-Barr virus-based vectors. Other embodiments of viral vectors and viral particles produce therefrom are described elsewhere herein. In some embodiments, the viral vectors are configured to produce replication incompetent viral particles for improved safety of these systems.

[0216] In certain embodiments, the virus structural component, which can be encoded by one or more polynucleotides in a viral vector or vector system, comprises one or more capsid proteins including an entire capsid. In certain embodiments, such as wherein a viral capsid

comprises multiple copies of different proteins, the delivery system can provide one or more of the same protein or a mixture of such proteins. For example, AAV comprises 3 capsid proteins, VP1, VP2, and VP3, thus delivery systems of the invention can comprise one or more of VP1, and/or one or more of VP2, and/or one or more of VP3. Accordingly, the present invention is applicable to a virus within the family Adenoviridae, such as Atadenovirus, e.g., Ovine atadenovirus D, Aviadenovirus, e.g., Fowl aviadenovirus A, Ichtadenovirus, e.g., Sturgeon ichtadenovirus A, Mastadenovirus (which includes adenoviruses such as all human adenoviruses), e.g., Human mastadenovirus C, and Siadenovirus, e.g., Frog siadenovirus A. Thus, a virus of within the family Adenoviridae is contemplated as within the invention with discussion herein as to adenovirus applicable to other family members. Target-specific AAV capsid variants can be used or selected. Non-limiting examples include capsid variants selected to bind to chronic myelogenous leukemia cells, human CD34 PBPC cells, breast cancer cells, cells of lung, heart, dermal fibroblasts, melanoma cells, stem cell, glioblastoma cells, coronary artery endothelial cells and keratinocytes. See, e.g., Buning et al, 2015, *Current Opinion in Pharmacology* 24, 94-104. From teachings herein and knowledge in the art as to modifications of adenovirus (see, e.g., US Patents 9,410,129, 7,344,872, 7,256,036, 6,911,199, 6,740,525; Matthews, "Capsid-Incorporation of Antigens into Adenovirus Capsid Proteins for a Vaccine Approach," *Mol Pharm*, 8(1): 3-11 (2011)), as well as regarding modifications of AAV, the skilled person can readily obtain a modified adenovirus that has a large payload protein or a CRISPR-protein, despite that heretofore it was not expected that such a large protein could be provided on an adenovirus. And as to the viruses related to adenovirus mentioned herein, as well as to the viruses related to AAV mentioned elsewhere herein, the teachings herein as to modifying adenovirus and AAV, respectively, can be applied to those viruses without undue experimentation from this disclosure and the knowledge in the art.

[0217] In some embodiments, the viral vector is configured such that when the cargo is packaged the cargo(s) (e.g. one or more components of the CRISPR-Cas system, including but not limited to a Cas effector, is external to the capsid or virus particle. In the sense that it is not inside the capsid (enveloped or encompassed with the capsid) but is externally exposed so that it can contact the target genomic DNA. In some embodiments, the viral vector is configured such that all the cargo(s) are contained within the capsid after packaging.

Split Viral Vector Systems

[0218] When the CRISPR-Cas system viral vector or vector system (be it a retroviral (e.g. AAV) or lentiviral vector) is designed so as to position the cargo(s) (e.g., one or more CRISPR-Cas system components) at the internal surface of the capsid once formed, the cargo(s) will fill most or all of internal volume of the capsid. In other embodiments, the CRISPR protein may be modified or divided so as to occupy a less of the capsid internal volume. Accordingly, in certain embodiments, the CRISPR-Cas system or component thereof (e.g. a Cas effector protein) can be divided in two portions, one portion comprises in one viral particle or capsid and the second portion comprised in a second viral particle or capsid. In certain embodiments, by splitting the CRISPR-Cas system or component thereof in two portions, space is made available to link one or more heterologous domains to one or both CRISPR-Cas system component (e.g., Cas protein) portions. Such systems can be referred to as “split vector systems” or in the context of the present disclosure a “split CRISPR-Cas system” a “split CRISPR protein”, a “split Cas protein” and the like. This split protein approach is also described elsewhere herein. When the concept is applied to a vector system, it thus describes putting pieces of the split proteins on different vectors thus reducing the payload of any one vector. This approach can facilitate delivery of systems where the total system size is close to or exceeds the packaging capacity of the vector. This is independent of any regulation of the CRISPR-Cas system that can be achieved with a split system or split protein design.

[0219] Split CRISPR proteins that can be incorporated into the AAV or other vectors described herein are set forth elsewhere herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split CRISPR proteins are attached to a member of a specific binding pair, and when bound with each other, the members of the specific binding pair maintain the parts of the CRISPR protein in proximity. In certain embodiments, each part of a split CRISPR protein is associated with an inducible binding pair. An inducible binding pair is one which is capable of being switched “on” or “off” by a protein or small molecule that binds to both members of the inducible binding pair. In general, according to the invention, CRISPR proteins may preferably split between domains, leaving domains intact. Preferred, non-limiting examples of such CRISPR proteins include, without limitation, Cas protein, and orthologues. Preferred, non-limiting examples of split points include, with reference to SpCas9: a split position between 202A/203S; a split position between 255F/256D; a split position between 310E/311I; a split position between 534R/535K;

a split position between 572E/573C; a split position between 713S/714G; a split position between 1003L/104E; a split position between 1054G/1055E; a split position between 1114N/1115S; a split position between 1152K/1153S; a split position between 1245K/1246G; or a split between 1098 and 1099. Corresponding positions in Cas12 or other Cas proteins can be appreciated in view of these positions made with reference to SpCas9.

[0220] In some embodiments, any AAV serotype is preferred. In some embodiments, the VP2 domain associated with the CRISPR enzyme is an AAV serotype 2 VP2 domain. In some embodiments, the VP2 domain associated with the CRISPR enzyme is an AAV serotype 8 VP2 domain. The serotype can be a mixed serotype as is known in the art.

Retroviral and Lentiviral Vectors

[0221] Retroviral vectors can be composed of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Suitable retroviral vectors for the CRISPR-Cas systems can include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian immunodeficiency virus (SIV), human immunodeficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700). Selection of a retroviral gene transfer system may therefore depend on the target tissue.

[0222] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and are described in greater detail elsewhere herein. A retrovirus can also be engineered to allow for conditional expression of the inserted transgene, such that only certain cell types are infected by the lentivirus.

[0223] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. Advantages of using a lentiviral approach can include the ability to transduce or infect non-dividing cells and their ability to typically produce high viral titers, which can increase efficiency or efficacy of production and delivery. Suitable lentiviral vectors include, but are not limited to, human immunodeficiency virus (HIV)-based

lentiviral vectors, feline immunodeficiency virus (FIV)-based lentiviral vectors, simian immunodeficiency virus (SIV)-based lentiviral vectors, Moloney Murine Leukaemia Virus (Mo-MLV), Visna-maedi virus (VMV)-based lentiviral vector, carpine arthritis-encephalitis virus (CAEV)-based lentiviral vector, bovine immune deficiency virus (BIV)-based lentiviral vector, and Equine infectious anemia (EIAV)-based lentiviral vector. In some embodiments, an HIV-based lentiviral vector system can be used. In some embodiments, a FIV-based lentiviral vector system can be used.

[0224] In some embodiments, the lentiviral vector is an EIAV-based lentiviral vector or vector system. EIAV vectors have been used to mediate expression, packaging, and/or delivery in other contexts, such as for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285). In another embodiment, RetinoStat®, (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)), which describes RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the wet form of age-related macular degeneration. Any of these vectors described in these publications can be modified for the elements of the CRISPR-Cas system described herein.

[0225] In some embodiments, the lentiviral vector or vector system thereof can be a first-generation lentiviral vector or vector system thereof. First-generation lentiviral vectors can contain a large portion of the lentivirus genome, including the gag and pol genes, other additional viral proteins (e.g. VSV-G) and other accessory genes (e.g. vif, vpr, vpu, nef, and combinations thereof), regulatory genes (e.g. tat and/or rev) as well as the gene of interest between the LTRs. First generation lentiviral vectors can result in the production of virus particles that can be capable of replication in vivo, which may not be appropriate for some instances or applications.

[0226] In some embodiments, the lentiviral vector or vector system thereof can be a second-generation lentiviral vector or vector system thereof. Second-generation lentiviral vectors do not contain one or more accessory virulence factors and do not contain all components necessary for virus particle production on the same lentiviral vector. This can result in the production of a replication-incompetent virus particle and thus increase the safety of these systems over first-generation lentiviral vectors. In some embodiments, the second-generation vector lacks one or more accessory virulence factors (e.g. vif, vpr, vpu, nef, and combinations thereof). Unlike the first-generation lentiviral vectors, no single second-

generation lentiviral vector includes all features necessary to express and package a polynucleotide into a virus particle. In some embodiments, the envelope and packaging components are split between two different vectors with the gag, pol, rev, and tat genes being contained on one vector and the envelope protein (e.g. VSV-G) are contained on a second vector. The gene of interest, its promoter, and LTRs can be included on a third vector that can be used in conjunction with the other two vectors (packaging and envelope vectors) to generate a replication-incompetent virus particle.

[0227] In some embodiments, the lentiviral vector or vector system thereof can be a third-generation lentiviral vector or vector system thereof. Third-generation lentiviral vectors and vector systems thereof have increased safety over first- and second-generation lentiviral vectors and systems thereof because, for example, the various components of the viral genome are split between two or more different vectors but used together in vitro to make virus particles, they can lack the tat gene (when a constitutively active promoter is included up-stream of the LTRs), and they can include one or more deletions in the 3'LTR to create self-inactivating (SIN) vectors having disrupted promoter/enhancer activity of the LTR. In some embodiments, a third-generation lentiviral vector system can include (i) a vector plasmid that contains the polynucleotide of interest and upstream promoter that are flanked by the 5' and 3' LTRs, which can optionally include one or more deletions present in one or both of the LTRs to render the vector self-inactivating; (ii) a "packaging vector(s)" that can contain one or more genes involved in packaging a polynucleotide into a virus particle that is produced by the system (e.g. gag, pol, and rev) and upstream regulatory sequences (e.g. promoter(s)) to drive expression of the features present on the packaging vector, and (iii) an "envelope vector" that contains one or more envelope protein genes and upstream promoters. In certain embodiments, the third-generation lentiviral vector system can include at least two packaging vectors, with the gag-pol being present on a different vector than the rev gene.

[0228] In some embodiments, 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) can be used/and or adapted to the CRISPR-Cas system of the present invention.

[0229] In some embodiments, the pseudotype and infectivity or tropism of a lentivirus particle can be tuned by altering the type of envelope protein(s) included in the lentiviral vector or system thereof. As used herein, an "envelope protein" or "outer protein" means a protein

exposed at the surface of a viral particle that is not a capsid protein. For example, envelope or outer proteins typically comprise proteins embedded in the envelope of the virus. In some embodiments, a lentiviral vector or vector system thereof can include a VSV-G envelope protein. VSV-G mediates viral attachment to an LDL receptor (LDLR) or an LDLR family member present on a host cell, which triggers endocytosis of the viral particle by the host cell. Because LDLR is expressed by a wide variety of cells, viral particles expressing the VSV-G envelope protein can infect or transduce a wide variety of cell types. Other suitable envelope proteins can be incorporated based on the host cell that a user desires to be infected by a virus particle produced from a lentiviral vector or system thereof described herein and can include, but are not limited to, feline endogenous virus envelope protein (RD114) (see e.g. Hanawa et al. *Molec. Ther.* 2002 5(3) 242-251), modified Sindbis virus envelope proteins (see e.g. Morizono et al. 2010. *J. Virol.* 84(14) 6923-6934; Morizono et al. 2001. *J. Virol.* 75:8016-8020; Morizono et al. 2009. *J. Gene Med.* 11:549-558; Morizono et al. 2006 *Virology* 355:71-81; Morizono et al. *J. Gene Med.* 11:655-663, Morizono et al. 2005 *Nat. Med.* 11:346-352), baboon retroviral envelope protein (see e.g. Girard-Gagnepain et al. 2014. *Blood.* 124: 1221-1231); Tupaia paramyxovirus glycoproteins (see e.g. Enkirch T. et al., 2013. *Gene Ther.* 20:16-23); measles virus glycoproteins (see e.g. Funke et al. 2008. *Molec. Ther.* 16(8): 1427-1436), rabies virus envelope proteins, MLV envelope proteins, Ebola envelope proteins, baculovirus envelope proteins, filovirus envelope proteins, hepatitis E1 and E2 envelope proteins, gp41 and gp120 of HIV, hemagglutinin, neuraminidase, M2 proteins of influenza virus, and combinations thereof.

[0230] In some embodiments, the tropism of the resulting lentiviral particle can be tuned by incorporating cell targeting peptides into a lentiviral vector such that the cell targeting peptides are expressed on the surface of the resulting lentiviral particle. In some embodiments, a lentiviral vector can contain an envelope protein that is fused to a cell targeting protein (see e.g. Buchholz et al. 2015. *Trends Biotechnol.* 33:777-790; Bender et al. 2016. *PLoS Pathog.* 12(e1005461); and Friedrich et al. 2013. *Mol. Ther.* 21: 849-859).

[0231] In some embodiments, a split-intein-mediated approach to target lentiviral particles to a specific cell type can be used (see e.g. Chamoun-Emaneulli et al. 2015. *Biotechnol. Bioeng.* 112:2611-2617, Ramirez et al. 2013. *Protein. Eng. Des. Sel.* 26:215-233. In these embodiments, a lentiviral vector can contain one half of a splicing-deficient variant of the naturally split intein from *Nostoc punctiforme* fused to a cell targeting peptide and the same or

different lentiviral vector can contain the other half of the split intein fused to an envelope protein, such as a binding-deficient, fusion-competent virus envelope protein. This can result in production of a virus particle from the lentiviral vector or vector system that includes a split intein that can function as a molecular Velcro linker to link the cell-binding protein to the pseudotyped lentivirus particle. This approach can be advantageous for use where surface-incompatibilities can restrict the use of, e.g., cell targeting peptides.

[0232] In some embodiments, a covalent-bond-forming protein-peptide pair can be incorporated into one or more of the lentiviral vectors described herein to conjugate a cell targeting peptide to the virus particle (see e.g. Kasaraneni et al. 2018. Sci. Reports (8) No. 10990). In some embodiments, a lentiviral vector can include an N-terminal PDZ domain of InaD protein (PDZ1) and its pentapeptide ligand (TEFCA (SEQ ID NO: 46)) from NorpA, which can conjugate the cell targeting peptide to the virus particle via a covalent bond (e.g. a disulfide bond). In some embodiments, the PDZ1 protein can be fused to an envelope protein, which can optionally be binding deficient and/or fusion competent virus envelope protein and included in a lentiviral vector. In some embodiments, the TEFCA (SEQ ID NO: 46) can be fused to a cell targeting peptide and the TEFCA-CPT (SEQ ID NO: 47) fusion construct can be incorporated into the same or a different lentiviral vector as the PDZ1-envelope protein construct. During virus production, specific interaction between the PDZ1 and TEFCA (SEQ ID NO: 46) facilitates producing virus particles covalently functionalized with the cell targeting peptide and thus capable of targeting a specific cell-type based upon a specific interaction between the cell targeting peptide and cells expressing its binding partner. This approach can be advantageous for use where surface-incompatibilities can restrict the use of, e.g., cell targeting peptides.

[0233] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 and US Patent No. US7259015. Any of these systems or a variant thereof can be used to deliver an CRISPR-Cas system polynucleotide described herein to a cell.

[0234] In some embodiments, a lentiviral vector system can include one or more transfer plasmids. Transfer plasmids can be generated from various other vector backbones and can include one or more features that can work with other retroviral and/or lentiviral vectors in the system that can, for example, improve safety of the vector and/or vector system, increase virial titers, and/or increase or otherwise enhance expression of the desired insert to be expressed and/or packaged into the viral particle. Suitable features that can be included in a transfer plasmid can include, but are not limited to, 5'LTR, 3'LTR, SIN/LTR, origin of replication (Ori), selectable marker genes (e.g. antibiotic resistance genes), Psi (Ψ), RRE (rev response element), cPPT (central polypurine tract), promoters, WPRE (woodchuck hepatitis post-transcriptional regulatory element), SV40 polyadenylation signal, pUC origin, SV40 origin, F1 origin, and combinations thereof.

[0235] In another embodiment, Cocal vesiculovirus envelope pseudotyped retroviral or lentiviral vector particles are contemplated (see, e.g., US Patent Publication No. 20120164118 assigned to the Fred Hutchinson Cancer Research Center). Cocal virus is in the Vesiculovirus genus and is a causative agent of vesicular stomatitis in mammals. Cocal virus was originally isolated from mites in Trinidad (Jonkers et al., *Am. J. Vet. Res.* 25:236-242 (1964)), and infections have been identified in Trinidad, Brazil, and Argentina from insects, cattle, and horses. Many of the vesiculoviruses that infect mammals have been isolated from naturally infected arthropods, suggesting that they are vector-borne. Antibodies to vesiculoviruses are common among people living in rural areas where the viruses are endemic and laboratory-acquired; infections in humans usually result in influenza-like symptoms. The Cocal virus envelope glycoprotein shares 71.5% identity at the amino acid level with VSV-G Indiana, and phylogenetic comparison of the envelope gene of vesiculoviruses shows that Cocal virus is serologically distinct from, but most closely related to, VSV-G Indiana strains among the vesiculoviruses. Jonkers et al., *Am. J. Vet. Res.* 25:236-242 (1964) and Travassos da Rosa et al., *Am. J. Tropical Med. & Hygiene* 33:999-1006 (1984). The Cocal vesiculovirus envelope pseudotyped retroviral vector particles may include for example, lentiviral, alpharetroviral, betaretroviral, gammaretroviral, deltaretroviral, and epsilon-retroviral vector particles that may comprise retroviral Gag, Pol, and/or one or more accessory protein(s) and a Cocal vesiculovirus envelope protein. In certain embodiments of these embodiments, the Gag, Pol, and accessory proteins are lentiviral and/or gammaretroviral. In some embodiments, a retroviral vector can

contain encoding polypeptides for one or more Cocal vesiculovirus envelope proteins such that the resulting viral or pseudoviral particles are Cocal vesiculovirus envelope pseudotyped.

Adenoviral vectors, Helper-dependent Adenoviral vectors, and Hybrid Adenoviral Vectors

[0236] In some embodiments, the vector can be an adenoviral vector. In some embodiments, the adenoviral vector can include elements such that the virus particle produced using the vector or system thereof can be serotype 2 or serotype 5. In some embodiments, the polynucleotide to be delivered via the adenoviral particle can be up to about 8 kb. Thus, in some embodiments, an adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 8 kb. Adenoviral vectors have been used successfully in several contexts (see e.g. Teramoto et al. 2000. *Lancet*. 355:1911-1912; Lai et al. 2002. *DNA Cell. Biol.* 21:895-913; Flotte et al., 1996. *Hum. Gene. Ther.* 7:1145-1159; and Kay et al. 2000. *Nat. Genet.* 24:257-261).

[0237] In some embodiments the vector can be a helper-dependent adenoviral vector or system thereof. These are also referred to in the art as “gutless” or “guttled” vectors and are a modified generation of adenoviral vectors (see e.g. Thrasher et al. 2006. *Nature*. 443:E5-7). In certain embodiments of the helper-dependent adenoviral vector system one vector (the helper) can contain all the viral genes required for replication but contains a conditional gene defect in the packaging domain. The second vector of the system can contain only the ends of the viral genome, one or more CRISPR-Cas polynucleotides, and the native packaging recognition signal, which can allow selective packaged release from the cells (see e.g. Cideciyan et al. 2009. *N Engl J Med.* 361:725-727). Helper-dependent adenoviral vector systems have been successful for gene delivery in several contexts (see e.g. Simonelli et al. 2010. *J Am Soc Gene Ther.* 18:643-650; Cideciyan et al. 2009. *N Engl J Med.* 361:725-727; Crane et al. 2012. *Gene Ther.* 19(4):443-452; Alba et al. 2005. *Gene Ther.* 12:18-S27; Croyle et al. 2005. *Gene Ther.* 12:579-587; Amalfitano et al. 1998. *J. Virol.* 72:926-933; and Morral et al. 1999. *PNAS.* 96:12816-12821). The techniques and vectors described in these publications can be adapted for inclusion and delivery of the CRISPR-Cas system polynucleotides described herein. In some embodiments, the polynucleotide to be delivered via the viral particle produced from a helper-dependent adenoviral vector or system thereof can be up to about 37 kb. Thus, in some embodiments, a adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 37 kb (see e.g. Rosewell et al. 2011. *J. Genet. Syndr. Gene Ther. Suppl.* 5:001).

[0238] In some embodiments, the vector is a hybrid-adenoviral vector or system thereof. Hybrid adenoviral vectors are composed of the high transduction efficiency of a gene-deleted adenoviral vector and the long-term genome-integrating potential of adeno-associated, retroviruses, lentivirus, and transposon based-gene transfer. In some embodiments, such hybrid vector systems can result in stable transduction and limited integration site. See e.g. Balague et al. 2000. *Blood*. 95:820-828; Morral et al. 1998. *Hum. Gene Ther.* 9:2709-2716; Kubo and Mitani. 2003. *J. Virol.* 77(5): 2964-2971; Zhang et al. 2013. *PloS One*. 8(10) e76771; and Cooney et al. 2015. *Mol. Ther.* 23(4):667-674), whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention. In some embodiments, a hybrid-adenoviral vector can include one or more features of a retrovirus and/or an adeno-associated virus. In some embodiments the hybrid-adenoviral vector can include one or more features of a spuma retrovirus or foamy virus (FV). See e.g. Ehrhardt et al. 2007. *Mol. Ther.* 15:146-156 and Liu et al. 2007. *Mol. Ther.* 15:1834-1841, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention. Advantages of using one or more features from the FVs in the hybrid-adenoviral vector or system thereof can include the ability of the viral particles produced therefrom to infect a broad range of cells, a large packaging capacity as compared to other retroviruses, and the ability to persist in quiescent (non-dividing) cells. See also e.g. Ehrhardt et al. 2007. *Mol. Ther.* 15:146-156 and Shuji et al. 2011. *Mol. Ther.* 19:76-82, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention.

Adeno Associated Viral (AAV) Vectors

[0239] In an embodiment, the vector can be an adeno-associated virus (AAV) vector. See, e.g., West et al., *Virology* 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, *Human Gene Therapy* 5:793-801 (1994); and Muzyczka, *J. Clin. Invest.* 94:1351 (1994). Although similar to adenoviral vectors in some of their features, AAVs have some deficiency in their replication and/or pathogenicity and thus can be safer than adenoviral vectors. In some embodiments the AAV can integrate into a specific site on chromosome 19 of a human cell with no observable side effects. In some embodiments, the capacity of the AAV vector, system thereof, and/or AAV particles can be up to about 4.7 kb.

[0240] The AAV vector or system thereof can include one or more regulatory molecules. In some embodiments the regulatory molecules can be promoters, enhancers, repressors, and

the like, which are described in greater detail elsewhere herein. In some embodiments, the AAV vector or system thereof can include one or more polynucleotides that can encode one or more regulatory proteins. In some embodiments, the one or more regulatory proteins can be selected from Rep78, Rep68, Rep52, Rep40, variants thereof, and combinations thereof.

[0241] The AAV vector or system thereof can include one or more polynucleotides that can encode one or more capsid proteins. The capsid proteins can be selected from VP1, VP2, VP3, and combinations thereof. The capsid proteins can be capable of assembling into a protein shell of the AAV virus particle. In some embodiments, the AAV capsid can contain 60 capsid proteins. In some embodiments, the ratio of VP1:VP2:VP3 in a capsid can be about 1:1:10.

[0242] In some embodiments, the AAV vector or system thereof can include one or more adenovirus helper factors or polynucleotides that can encode one or more adenovirus helper factors. Such adenovirus helper factors can include, but are not limited, E1A, E1B, E2A, E4ORF6, and VA RNAs. In some embodiments, a producing host cell line expresses one or more of the adenovirus helper factors.

[0243] The AAV vector or system thereof can be configured to produce AAV particles having a specific serotype. In some embodiments, the serotype can be AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, AAV-9 or any combinations thereof. In some embodiments, the AAV can be AAV1, AAV-2, AAV-5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV-1, AAV-2, AAV-5 or any combination thereof for targeting brain and/or neuronal cells; and one can select AAV-4 for targeting cardiac tissue; and one can select AAV8 for delivery to the liver. Thus, in some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the brain and/or neuronal cells can be configured to generate AAV particles having serotypes 1, 2, 5 or a hybrid capsid AAV-1, AAV-2, AAV-5 or any combination thereof. In some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting cardiac tissue can be configured to generate an AAV particle having an AAV-4 serotype. In some embodiments, an AAV vector or system thereof capable of producing AAV particles capable of targeting the liver can be configured to generate an AAV having an AAV-8 serotype. In some embodiments, the AAV vector is a hybrid AAV vector or system thereof. Hybrid AAVs are AAVs that include genomes with elements from one serotype that are packaged into a capsid derived from at least one different serotype. For example, if it is the rAAV2/5 that is to

be produced, and if the production method is based on the helper-free, transient transfection method discussed above, the 1st plasmid and the 3rd plasmid (the adeno helper plasmid) will be the same as discussed for rAAV2 production. However, the second plasmid, the pRepCap will be different. In this plasmid, called pRep2/Cap5, the Rep gene is still derived from AAV2, while the Cap gene is derived from AAV5. The production scheme is the same as the above-mentioned approach for AAV2 production. The resulting rAAV is called rAAV2/5, in which the genome is based on recombinant AAV2, while the capsid is based on AAV5. It is assumed the cell or tissue-tropism displayed by this AAV2/5 hybrid virus should be the same as that of AAV5.

[0244] A tabulation of certain AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008), which is recapitulated in **Table 3** below.

[0245] **Table 3.**

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

[0246] In some embodiments, the AAV vector or system thereof is configured as a “gutless” vector, similar to that described in connection with a retroviral vector. In some embodiments, the “gutless” AAV vector or system thereof can have the cis-acting viral DNA elements involved in genome amplification and packaging in linkage with the heterologous sequences of interest (e.g. the CRISPR-Cas system polynucleotide(s)).

[0247] In some embodiments, the AAV vectors are produced in insect cells, e.g., *Spodoptera frugiperda* Sf9 insect cells, grown in serum-free suspension culture. Serum-free insect cells can be purchased from commercial vendors, e.g., Sigma Aldrich (EX-CELL 405).

[0248] In some embodiments, an AAV vector or vector system can contain or consists essentially of one or more polynucleotides encoding one or more components of a CRISPR system. In some embodiments, the AAV vector or vector system can contain a plurality of cassettes comprising or consisting a first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding a CRISPR-associated (Cas) protein (putative nuclease or helicase proteins), e.g., a Cas protein and a terminator, and a two, or more, advantageously up to the packaging size limit of the vector, e.g., in total (including the first cassette) five, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector), or two or more individual rAAVs, each containing one or more than one cassette of a CRISPR system, e.g., a first rAAV containing the first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding Cas, e.g., a Cas and a terminator, and a second rAAV containing a plurality, four, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector). As rAAV is a DNA virus, the nucleic acid molecules in the herein discussion concerning AAV or rAAV are advantageously DNA. In some embodiments, the promoter is a tissue specific promoter or another tissue specific regulatory element. Suitable tissue specific regulatory elements, including promoters, are described in greater detail elsewhere herein.

[0249] In another embodiment, the invention provides a non-naturally occurring or engineered CRISPR protein associated with Adeno Associated Virus (AAV), e.g., an AAV comprising a CRISPR protein as a fusion, with or without a linker, to or with an AAV capsid protein such as VP1, VP2, and/or VP3; and, for shorthand purposes, such a non-naturally occurring or engineered CRISPR protein is herein termed a “AAV-CRISPR protein” More in particular, modifying the knowledge in the art, e.g., Rybniker et al., “Incorporation of Antigens

into Viral Capsids Augments Immunogenicity of Adeno-Associated Virus Vector-Based Vaccines,” J Virol. Dec 2012; 86(24): 13800–13804, Lux K, et al. 2005. Green fluorescent protein-tagged adeno-associated virus particles allow the study of cytosolic and nuclear trafficking. J. Virol. 79:11776–11787, Munch RC, et al. 2012. “Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer.” Mol. Ther. [Epub ahead of print.] doi:10.1038/mt.2012.186 and Warrington KH, Jr, et al. 2004. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. J. Virol. 78:6595–6609, each incorporated herein by reference, one can obtain a modified AAV capsid of the invention. It will be understood by those skilled in the art that the modifications described herein if inserted into the AAV cap gene may result in modifications in the VP1, VP2 and/or VP3 capsid subunits. Alternatively, the capsid subunits can be expressed independently to achieve modification in only one or two of the capsid subunits (VP1, VP2, VP3, VP1+VP2, VP1+VP3, or VP2+VP3). One can modify the cap gene to have expressed at a desired location a non-capsid protein advantageously a large payload protein, such as a CRISPR-protein. Likewise, these can be fusions, with the protein, e.g., large payload protein such as a CRISPR-protein fused in a manner analogous to prior art fusions. See, e.g., US Patent Publication 20090215879; Nance et al., “Perspective on Adeno-Associated Virus Capsid Modification for Duchenne Muscular Dystrophy Gene Therapy,” Hum Gene Ther. 26(12):786–800 (2015) and documents cited therein, incorporated herein by reference. The skilled person, from this disclosure and the knowledge in the art can make and use modified AAV or AAV capsid as in the herein invention, and through this disclosure one knows now that large payload proteins can be fused to the AAV capsid. Applicants provide AAV capsid -CRISPR protein (e.g., Cas), dCas fusions and those AAV-capsid CRISPR protein (e.g., Casfusions can be a recombinant AAV that contains nucleic acid molecule(s) encoding or providing CRISPR-Cas or CRISPR system or complex RNA guide(s), whereby the CRISPR protein (e.g., Cas) fusion delivers a CRISPR-Cas or CRISPR system complex (e.g., the CRISPR protein or Cas is provided by the fusion, e.g., VP1, VP2, or VP3 fusion, and the guide RNA is provided by the coding of the recombinant virus, whereby in vivo, in a cell, the CRISPR-Cas or CRISPR system is assembled from the nucleic acid molecule(s) of the recombinant providing the guide RNA and the outer surface of the virus providing the CRISPR-Enzyme (e.g., Cas). Such as complex may herein be termed an “AAV-CRISPR system” or an “AAV-CRISPR-Cas” or “AAV-CRISPR complex” or AAV-CRISPR-Cas

complex.” Accordingly, the instant invention is also applicable to a virus in the genus Dependoparvovirus or in the family Parvoviridae, for instance, AAV, or a virus of Amdoparvovirus, e.g., Carnivore amdoparvovirus 1, a virus of Aveparvovirus, e.g., Galliform aveparvovirus 1, a virus of Bocaparvovirus, e.g., Ungulate bocaparvovirus 1, a virus of Copiparvovirus, e.g., Ungulate copiparvovirus 1, a virus of Dependoparvovirus, e.g., Adeno-associated dependoparvovirus A, a virus of Erythroparvovirus, e.g., Primate erythroparvovirus 1, a virus of Protoparvovirus, e.g., Rodent protoparvovirus 1, a virus of Tetraparvovirus, e.g., Primate tetraparvovirus 1. Thus, a virus of within the family Parvoviridae or the genus Dependoparvovirus or any of the other foregoing genera within Parvoviridae is contemplated as within the invention with discussion herein as to AAV applicable to such other viruses.

[0250] In some embodiments, the CRISPR enzyme is external to the capsid or virus particle. In the sense that it is not inside the capsid (enveloped or encompassed with the capsid) but is externally exposed so that it can contact the target genomic DNA. In some embodiments, the CRISPR enzyme is associated with the AAV VP2 domain by way of a fusion protein. In some embodiments, the association may be considered to be a modification of the VP2 domain. Where reference is made herein to a modified VP2 domain, then this will be understood to include any association discussed herein of the VP2 domain and the CRISPR enzyme. In some embodiments, the AAV VP2 domain may be associated (or tethered) to the CRISPR enzyme via a connector protein, for example using a system such as the streptavidin-biotin system. In an embodiment, the present invention provides a polynucleotide encoding the present CRISPR enzyme and associated AAV VP2 domain. In one embodiment, the invention provides a non-naturally occurring modified AAV having a VP2-CRISPR enzyme capsid protein, wherein the CRISPR enzyme is part of or tethered to the VP2 domain. In some preferred embodiments, the CRISPR enzyme is fused to the VP2 domain so that, in another embodiment, the invention provides a non-naturally occurring modified AAV having a VP2-CRISPR enzyme fusion capsid protein. Thus, reference herein to a VP2-CRISPR enzyme capsid protein may also include a VP2-CRISPR enzyme fusion capsid protein. In some embodiments, the VP2-CRISPR enzyme capsid protein further comprises a linker, whereby the VP2-CRISPR enzyme is distanced from the remainder of the AAV. In some embodiments, the VP2-CRISPR enzyme capsid protein further comprises at least one protein complex, e.g., CRISPR complex, such as a CRISPR-Cas complex guide RNA that targets a particular DNA, TALE, etc. A CRISPR complex, such as CRISPR-Cas system comprising the VP2-CRISPR enzyme capsid protein

and at least one CRISPR complex, such as a CRISPR-Cas complex guide RNA that targets a particular DNA, is also provided in one embodiment.

[0251] In one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a CRISPR enzyme which is part of or tethered to an AAV capsid domain, i.e., VP1, VP2, or VP3 domain of Adeno-Associated Virus (AAV) capsid. In some embodiments, part of or tethered to an AAV capsid domain includes associated with associated with a AAV capsid domain. In some embodiments, the CRISPR enzyme may be fused to the AAV capsid domain. In some embodiments, the fusion may be to the N-terminal end of the AAV capsid domain. As such, in some embodiments, the C- terminal end of the CRISPR enzyme is fused to the N- terminal end of the AAV capsid domain. In some embodiments, an NLS and/or a linker (such as a GlySer linker) may be positioned between the C- terminal end of the CRISPR enzyme and the N- terminal end of the AAV capsid domain. In some embodiments, the fusion may be to the C-terminal end of the AAV capsid domain. In some embodiments, this is not preferred due to the fact that the VP1, VP2 and VP3 domains of AAV are alternative splices of the same RNA and so a C- terminal fusion may affect all three domains. In some embodiments, the AAV capsid domain is truncated. In some embodiments, some or all of the AAV capsid domain is removed. In some embodiments, some of the AAV capsid domain is removed and replaced with a linker (such as a GlySer linker), typically leaving the N- terminal and C- terminal ends of the AAV capsid domain intact, such as the first 2, 5 or 10 amino acids. In this way, the internal (non-terminal) portion of the VP3 domain may be replaced with a linker. It is particularly preferred that the linker is fused to the CRISPR protein. A branched linker may be used, with the CRISPR protein fused to the end of one of the branches. This allows for some degree of spatial separation between the capsid and the CRISPR protein. In this way, the CRISPR protein is part of (or fused to) the AAV capsid domain.

[0252] In other embodiments, the CRISPR enzyme may be fused in frame within, i.e. internal to, the AAV capsid domain. Thus, in some embodiments, the AAV capsid domain again preferably retains its N- terminal and C- terminal ends. In this case, a linker is preferred, in some embodiments, either at one or both ends of the CRISPR enzyme. In this way, the CRISPR enzyme is again part of (or fused to) the AAV capsid domain. In certain embodiments, the positioning of the CRISPR enzyme is such that the CRISPR enzyme is at the external surface of the viral capsid once formed. In one embodiment, the invention provides a non-

naturally occurring or engineered composition comprising a CRISPR enzyme associated with a AAV capsid domain of Adeno-Associated Virus (AAV) capsid. Here, associated may mean in some embodiments fused, or in some embodiments bound to, or in some embodiments tethered to. The CRISPR protein may, in some embodiments, be tethered to the VP1, VP2, or VP3 domain. This may be via a connector protein or tethering system such as the biotin-streptavidin system. In one example, a biotinylation sequence (15 amino acids) could therefore be fused to the CRISPR protein. When a fusion of the AAV capsid domain, especially the N-terminus of the AAV capsid domain, with streptavidin is also provided, the two will therefore associate with very high affinity. Thus, in some embodiments, provided is a composition or system comprising a CRISPR protein-biotin fusion and a streptavidin- AAV capsid domain arrangement, such as a fusion. The CRISPR protein-biotin and streptavidin- AAV capsid domain forms a single complex when the two parts are brought together. NLSs may also be incorporated between the CRISPR protein and the biotin; and/or between the streptavidin and the AAV capsid domain.

[0253] As such, provided is a fusion of a CRISPR enzyme with a connector protein specific for a high affinity ligand for that connector, whereas the AAV VP2 domain is bound to said high affinity ligand. For example, streptavidin may be the connector fused to the CRISPR enzyme, while biotin may be bound to the AAV VP2 domain. Upon co-localization, the streptavidin will bind to the biotin, thus connecting the CRISPR enzyme to the AAV VP2 domain. The reverse arrangement is also possible. In some embodiments, a biotinylation sequence (15 amino acids) could therefore be fused to the AAV VP2 domain, especially the N-terminus of the AAV VP2 domain. A fusion of the CRISPR enzyme with streptavidin is also preferred, in some embodiments. In some embodiments, the biotinylated AAV capsids with streptavidin-CRISPR enzyme are assembled *in vitro*. This way the AAV capsids should assemble in a straightforward manner and the CRISPR enzyme-streptavidin fusion can be added after assembly of the capsid. In other embodiments a biotinylation sequence (15 amino acids) could therefore be fused to the CRISPR enzyme, together with a fusion of the AAV VP2 domain, especially the N- terminus of the AAV VP2 domain, with streptavidin. For simplicity, a fusion of the CRISPR enzyme and the AAV VP2 domain is preferred in some embodiments. In some embodiments, the fusion may be to the N- terminal end of the CRISPR enzyme. In other words, in some embodiments, the AAV and CRISPR enzyme are associated via fusion. In some embodiments, the AAV and CRISPR enzyme are associated via fusion including a

linker. Suitable linkers are discussed herein but include GlySer linkers. Fusion to the N- term of AAV VP2 domain is preferred, in some embodiments. In some embodiments, the CRISPR enzyme comprises at least one Nuclear Localization Signal (NLS). In a further embodiment, the present invention provides compositions comprising the CRISPR enzyme and associated AAV VP2 domain or the polynucleotides or vectors described herein. Such compositions and formulations are discussed elsewhere herein.

[0254] An alternative tether may be to fuse or otherwise associate the AAV capsid domain to an adaptor protein which binds to or recognizes to a corresponding RNA sequence or motif. In some embodiments, the adaptor is or comprises a binding protein which recognizes and binds (or is bound by) an RNA sequence specific for said binding protein. In some embodiments, a preferred example is the MS2 (*see* Konermann et al. Dec 2014, cited *infra*, incorporated herein by reference) binding protein which recognizes and binds (or is bound by) an RNA sequence specific for the MS2 protein.

[0255] With the AAV capsid domain associated with the adaptor protein, the CRISPR protein may, in some embodiments, be tethered to the adaptor protein of the AAV capsid domain. The CRISPR protein may, in some embodiments, be tethered to the adaptor protein of the AAV capsid domain via the CRISPR enzyme being in a complex with a modified guide, *see* Konermann *et al.* The modified guide is, in some embodiments, a sgRNA. In some embodiments, the modified guide comprises a distinct RNA sequence; *see*, e.g., International Patent Application No. PCT/US14/70175, incorporated herein by reference.

[0256] In some embodiments, distinct RNA sequence is an aptamer. Thus, corresponding aptamer- adaptor protein systems are preferred. One or more functional domains may also be associated with the adaptor protein. An example of a preferred arrangement would be: [AAV AAV capsid domain - adaptor protein] - [modified guide - CRISPR protein]

[0257] In certain embodiments, the positioning of the CRISPR protein is such that the CRISPR protein is at the internal surface of the viral capsid once formed. In one embodiment, the invention provides a non-naturally occurring or engineered composition comprising a CRISPR protein associated with an internal surface of an AAV capsid domain. Here again, associated may mean in some embodiments fused, or in some embodiments bound to, or in some embodiments tethered to. The CRISPR protein may, in some embodiments, be tethered to the VP1, VP2, or VP3 domain such that it locates to the internal surface of the viral capsid

once formed. This may be via a connector protein or tethering system such as the biotin-streptavidin system as described above and/or elsewhere herein.

[0258] In one embodiment, the invention provides an engineered, non-naturally occurring CRISPR-Cas system comprising a AAV-Cas protein and a guide RNA that targets a DNA molecule encoding a gene product in a cell, whereby the guide RNA targets the DNA molecule encoding the gene product and the Cas protein cleaves the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the guide RNA do not naturally occur together. The invention comprehends the guide RNA comprising a guide sequence fused to a tracr sequence. In a preferred embodiment the Cas protein is a Cas protein. In some embodiments, the polynucleotide encoding the Cas protein is codon optimized for expression in a eukaryotic cell. In some embodiments, the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment, the expression of the gene product is decreased.

[0259] In another embodiment, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a first regulatory element operably linked to a CRISPR-Cas system guide RNA that targets a DNA molecule encoding a gene product and an AAV-Cas protein. The components may be located on same or different vectors of the system or may be the same vector whereby the AAV-Cas protein also delivers the RNA of the CRISPR system. The guide RNA targets the DNA molecule encoding the gene product in a cell and the AAV-Cas protein may cleaves the DNA molecule encoding the gene product (it may cleave one or both strands or have substantially no nuclease activity), whereby expression of the gene product is altered; and, wherein the AAV-Cas protein and the guide RNA do not naturally occur together. The invention comprehends the guide RNA comprising a guide sequence fused to a tracr sequence. In an embodiment of the invention the AAV-Cas protein is a Type V AAV-CRISPR-Cas protein and in a preferred embodiment the AAV-Cas protein is an AAV-Cas protein. The invention further comprehends the coding for the AAV-Cas protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[0260] In one embodiment, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably

linked to a tracr mate sequence and one or more insertion sites for inserting one or more guide sequences upstream of the tracr mate sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a AAV-CRISPR complex to a target sequence in a eukaryotic cell, wherein the CRISPR complex comprises a AAV-CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence; and (b) said AAV-CRISPR enzyme comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on or in the same or different vectors of the system. In some embodiments, component (a) further comprises the tracr sequence downstream of the tracr mate sequence under the control of the first regulatory element. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of an AAV-CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the system comprises the tracr sequence under the control of a third regulatory element, such as a polymerase III promoter. In some embodiments, the tracr sequence exhibits at least 50%, 60%, 70%, 80%, 90%, 95%, or 99% of sequence complementarity along the length of the tracr mate sequence when optimally aligned. Determining optimal alignment is within the purview of one of skill in the art. For example, there are publicly and commercially available alignment algorithms and programs such as, but not limited to, ClustalW, Smith-Waterman in matlab, Bowtie, Geneious, Biopython and SeqMan. In some embodiments, the AAV-CRISPR complex comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR complex in a detectable amount in the nucleus of a eukaryotic cell. Without wishing to be bound by theory, it is believed that a nuclear localization sequence is not necessary for AAV-CRISPR complex activity in eukaryotes, but that including such sequences enhances activity of the system, especially as to targeting nucleic acid molecules in the nucleus and/or having molecules exit the nucleus. In some embodiments, the AAV-CRISPR enzyme is an AAV-Cas enzyme. In some embodiments, the AAV-Cas enzyme is derived from a Cas12 protein (e.g., a *Lachnospiraceae bacterium* Cas12a (LbCas12a), *Acidaminococcus* sp. (AsCas12a), *Francisella novicida* Cas12a (FnCas12a), *Moraxella bovoculi* Cas12a (MbCas12), or *Eubacterium rectale* Cas12a (MAD7)) (e.g., a Cas12 protein of one of these organisms modified to have or be associated with at least one AAV) and may include further mutations

or alterations or be a chimeric Cas12. The enzyme may be an AAV-Cas12 homolog or ortholog. In some embodiments, the AAV-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the AAV-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the AAV-CRISPR enzyme lacks DNA strand cleavage activity. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 15, 16, 17, 18, 19, 20, 25 nucleotides, or between 10-30, or between 15-25, or between 15-20 nucleotides in length.

[0261] In general, in some embodiments, the AAV further comprises a repair template. It will be appreciated that comprises here may mean encompassed within the viral capsid or that the virus encodes the comprised protein. In some embodiments, one or more, preferably two or more guide RNAs, may be comprised/encompassed within the AAV vector. Two may be preferred, in some embodiments, as it allows for multiplexing or dual nickase approaches. Particularly for multiplexing, two or more guides may be used. In fact, in some embodiments, three or more, four or more, five or more, or even six or more guide RNAs may be comprised/encompassed within the AAV. More space has been freed up within the AAV by virtue of the fact that the AAV no longer needs to comprise/encompass the CRISPR enzyme. In each of these instances, a repair template may also be provided comprised/encompassed within the AAV. In some embodiments, the repair template corresponds to or includes the DNA target.

Herpes Simplex Viral Vectors

[0262] In some embodiments, the vector can be a Herpes Simplex Viral (HSV)-based vector or system thereof. HSV systems can include the disabled infections single copy (DISC) viruses, which are composed of a glycoprotein H defective mutant HSV genome. When the defective HSV is propagated in complementing cells, virus particles can be generated that are capable of infecting subsequent cells permanently replicating their own genome but are not capable of producing more infectious particles. See e.g. 2009. Trobridge. Exp. Opin. Biol. Ther. 9:1427-1436, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention. In some embodiments where an HSV vector or system thereof is utilized, the host cell can be a complementing cell. In some embodiments, HSV vector or system thereof can be capable of producing virus

particles capable of delivering a polynucleotide cargo of up to 150 kb. Thus, in some embodiment the CRISPR-Cas system polynucleotide(s) included in the HSV-based viral vector or system thereof can sum from about 0.001 to about 150 kb. HSV-based vectors and systems thereof have been successfully used in several contexts including various models of neurologic disorders. See e.g. Cockrell et al. 2007. *Mol. Biotechnol.* 36:184-204; Kafri T. 2004. *Mol. Biol.* 246:367-390; Balaggan and Ali. 2012. *Gene Ther.* 19:145-153; Wong et al. 2006. *Hum. Gen. Ther.* 2002. 17:1-9; Azzouz et al. *J. Neurosci.* 22L10302-10312; and Betchen and Kaplitt. 2003. *Curr. Opin. Neurol.* 16:487-493, whose techniques and vectors described therein can be modified and adapted for use in the CRISPR-Cas system of the present invention.

Poxvirus Vectors

[0263] In some embodiments, the vector can be a poxvirus vector or system thereof. In some embodiments, the poxvirus vector can result in cytoplasmic expression of one or more CRISPR-Cas system polynucleotides of the present invention. In some embodiments the capacity of a poxvirus vector or system thereof can be about 25 kb or more. In some embodiments, a poxvirus vector or system thereof can include one or more CRISPR-Cas system polynucleotides described herein.

Viral Vectors for delivery to plants

[0264] The systems and compositions may be delivered to plant cells using viral vehicles. In particular embodiments, 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.

Virus Particle Production from Viral Vectors

Retroviral Production

[0265] In some embodiments, one or more viral vectors and/or system thereof can be delivered to a suitable cell line for production of virus particles containing the polynucleotide or other payload to be delivered to a host cell. Suitable host cells for virus production from

viral vectors and systems thereof described herein are known in the art and are commercially available. For example, suitable host cells include HEK 293 cells and its variants (HEK 293T and HEK 293TN cells). In some embodiments, the suitable host cell for virus production from viral vectors and systems thereof described herein can stably express one or more genes involved in packaging (e.g. pol, gag, and/or VSV-G) and/or other supporting genes.

[0266] In some embodiments, after delivery of one or more viral vectors to the suitable host cells for or virus production from viral vectors and systems thereof, the cells are incubated for an appropriate length of time to allow for viral gene expression from the vectors, packaging of the polynucleotide to be delivered (e.g. an CRISPR-Cas system polynucleotide), and virus particle assembly, and secretion of mature virus particles into the culture media. Various other methods and techniques are generally known to those of ordinary skill in the art.

[0267] Mature virus particles can be collected from the culture media by a suitable method. In some embodiments, this can involve centrifugation to concentrate the virus. The titer of the composition containing the collected virus particles can be obtained using a suitable method. Such methods can include transducing a suitable cell line (e.g. NIH 3T3 cells) and determining transduction efficiency, infectivity in that cell line by a suitable method. Suitable methods include PCR-based methods, flow cytometry, and antibiotic selection-based methods. Various other methods and techniques are generally known to those of ordinary skill in the art. The concentration of virus particle can be adjusted as needed. In some embodiments, the resulting composition containing virus particles can contain 1×10^1 - 1×10^{20} particles/mL.

[0268] Lentiviruses may be prepared from any lentiviral vector or vector system described herein. In one example embodiment, after cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) can be seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, the media can be changed to OptiMEM (serum-free) media and transfection of the lentiviral vectors can done 4 hours later. Cells can be transfected with 10 µg of lentiviral transfer plasmid (pCasES10) and the appropriate packaging plasmids (e.g., 5 µg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat)). Transfection can be carried out in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100ul Plus reagent). After 6 hours, the media can be changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods can use serum during cell culture, but serum-free methods are preferred.

[0269] Following transfection and allowing the producing cells (also referred to as packaging cells) to package and produce virus particles with packaged cargo, the lentiviral particles can be purified. In an exemplary embodiment, virus-containing supernatants can be harvested after 48 hours. Collected virus-containing supernatants can first be cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They can then be spun in an ultracentrifuge for 2 hours at 24,000 rpm. The resulting virus-containing pellets can be resuspended in 50ul of DMEM overnight at 4 degrees C. They can be then aliquoted and used immediately or immediately frozen at -80 degrees C for storage.

AAV Particle Production

[0270] There are two main strategies for producing AAV particles from AAV vectors and systems thereof, such as those described herein, which depend on how the adenovirus helper factors are provided (helper v. helper free). In some embodiments, a method of producing AAV particles from AAV vectors and systems thereof can include adenovirus infection into cell lines that stably harbor AAV replication and capsid encoding polynucleotides along with AAV vector containing the polynucleotide to be packaged and delivered by the resulting AAV particle (e.g. the CRISPR-Cas system polynucleotide(s)). In some embodiments, a method of producing AAV particles from AAV vectors and systems thereof can be a “helper free” method, which includes co-transfection of an appropriate producing cell line with three vectors (e.g. plasmid vectors): (1) an AAV vector that contains a polynucleotide of interest (e.g. the CRISPR-Cas system polynucleotide(s)) between 2 ITRs; (2) a vector that carries the AAV Rep-Cap encoding polynucleotides; and (helper polynucleotides. One of skill in the art will appreciate various methods and variations thereof that are both helper and -helper free and as well as the different advantages of each system.

Non-Viral Vectors

[0271] In some embodiments, the vector is a non-viral vector or vector system. The term of art “Non-viral vector” and as used herein in this context refers to molecules and/or compositions that are vectors but that are not based on one or more component of a virus or virus genome (excluding any nucleotide to be delivered and/or expressed by the non-viral vector) that can be capable of incorporating CRISPR-Cas polynucleotide(s) and delivering said CRISPR-Cas polynucleotide(s) to a cell and/or expressing the polynucleotide in the cell. It will be appreciated that this does not exclude vectors containing a polynucleotide designed to target a virus-based polynucleotide that is to be delivered. For example, if a gRNA to be delivered is

directed against a virus component and it is inserted or otherwise coupled to an otherwise non-viral vector or carrier, this would not make said vector a “viral vector”. Non-viral vectors can include, without limitation, naked polynucleotides and polynucleotide (non-viral) based vector and vector systems.

Naked Polynucleotides

[0272] In some embodiments one or more CRISPR-Cas system polynucleotides described elsewhere herein can be included in a naked polynucleotide. The term of art “naked polynucleotide” as used herein refers to polynucleotides that are not associated with another molecule (e.g. proteins, lipids, and/or other molecules) that can often help protect it from environmental factors and/or degradation. As used herein, associated with includes, but is not limited to, linked to, adhered to, adsorbed to, enclosed in, enclosed in or within, mixed with, and the like. Naked polynucleotides that include one or more of the CRISPR-Cas system polynucleotides described herein can be delivered directly to a host cell and optionally expressed therein. The naked polynucleotides can have any suitable two- and three-dimensional configurations. By way of non-limiting examples, naked polynucleotides can be single-stranded molecules, double stranded molecules, circular molecules (e.g. plasmids and artificial chromosomes), molecules that contain portions that are single stranded and portions that are double stranded (e.g. ribozymes), and the like. In some embodiments, the naked polynucleotide contains only the CRISPR-Cas system polynucleotide(s) of the present invention. In some embodiments, the naked polynucleotide can contain other nucleic acids and/or polynucleotides in addition to the CRISPR-Cas system polynucleotide(s) of the present invention. The naked polynucleotides can include one or more elements of a transposon system. Transposons and system thereof are described in greater detail elsewhere herein.

Non-Viral Polynucleotide Vectors

[0273] In some embodiments, one or more of the CRISPR-Cas system polynucleotides can be included in a non-viral polynucleotide vector. Suitable non-viral polynucleotide vectors include, but are not limited to, transposon vectors and vector systems, plasmids, bacterial artificial chromosomes, yeast artificial chromosomes, AR(antibiotic resistance)-free plasmids and miniplasmids, circular covalently closed vectors (e.g. minicircles, minivectors, miniknots,), linear covalently closed vectors (“dumbbell shaped”), MIDGE (minimalistic immunologically defined gene expression) vectors, MiLV (micro-linear vector) vectors,

Ministrings, mini-intronic plasmids, PSK systems (post-segregationally killing systems), ORT (operator repressor titration) plasmids, and the like. See e.g. Hardee et al. 2017. *Genes*. 8(2):65.

[0274] In some embodiments, the non-viral polynucleotide vector can have a conditional origin of replication. In some embodiments, the non-viral polynucleotide vector can be an ORT plasmid. In some embodiments, the non-viral polynucleotide vector can have a minimalistic immunologically defined gene expression. In some embodiments, the non-viral polynucleotide vector can have one or more post-segregationally killing system genes. In some embodiments, the non-viral polynucleotide vector is AR-free. In some embodiments, the non-viral polynucleotide vector is a minivector. In some embodiments, the non-viral polynucleotide vector includes a nuclear localization signal. In some embodiments, the non-viral polynucleotide vector can include one or more CpG motifs. In some embodiments, the non-viral polynucleotide vectors can include one or more scaffold/matrix attachment regions (S/MARs). See e.g. Mirkovitch et al. 1984. *Cell*. 39:223-232, Wong et al. 2015. *Adv. Genet*. 89:113-152, whose techniques and vectors can be adapted for use in the present invention. S/MARs are AT-rich sequences that play a role in the spatial organization of chromosomes through DNA loop base attachment to the nuclear matrix. S/MARs are often found close to regulatory elements such as promoters, enhancers, and origins of DNA replication. Inclusion of one or S/MARs can facilitate a once-per-cell-cycle replication to maintain the non-viral polynucleotide vector as an episome in daughter cells. In certain embodiments, the S/MAR sequence is located downstream of an actively transcribed polynucleotide (e.g. one or more CRISPR-Cas system polynucleotides of the present invention) included in the non-viral polynucleotide vector. In some embodiments, the S/MAR can be a S/MAR from the beta-interferon gene cluster. See e.g. Verghese et al. 2014. *Nucleic Acid Res*. 42:e53; Xu et al. 2016. *Sci. China Life Sci*. 59:1024-1033; Jin et al. 2016. 8:702-711; Koirala et al. 2014. *Adv. Exp. Med. Biol*. 801:703-709; and Nehlsen et al. 2006. *Gene Ther. Mol. Biol*. 10:233-244, whose techniques and vectors can be adapted for use in the present invention.

[0275] In some embodiments, the non-viral vector is a transposon vector or system thereof. 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 and DNA transposons. 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. In some embodiments, the non-viral polynucleotide vector can be a retrotransposon vector. In some embodiments, the retrotransposon vector includes long terminal repeats. In some embodiments, the retrotransposon vector does not include long terminal repeats. In some embodiments, the non-viral polynucleotide vector can be a DNA transposon vector. DNA transposon vectors can include a polynucleotide sequence encoding a transposase. In some embodiments, the transposon vector is configured as a non-autonomous transposon vector, meaning that the transposition does not occur spontaneously on its own. In some of these embodiments, the transposon vector lacks one or more polynucleotide sequences encoding proteins required for transposition. In some embodiments, the non-autonomous transposon vectors lack one or more Ac elements.

[0276] In some embodiments a non-viral polynucleotide transposon vector system can include a first polynucleotide vector that contains the CRISPR-Cas system polynucleotide(s) of the present invention flanked on the 5' and 3' ends by transposon terminal inverted repeats (TIRs) and a second polynucleotide vector that includes a polynucleotide capable of encoding a transposase coupled to a promoter to drive expression of the transposase. When both are expressed in the same cell the transposase can be expressed from the second vector and can transpose the material between the TIRs on the first vector (e.g. the CRISPR-Cas system polynucleotide(s) of the present invention) and integrate it into one or more positions in the host cell's genome. In some embodiments the transposon vector or system thereof can be configured as a gene trap. In some embodiments, the TIRs can be configured to flank a strong splice acceptor site followed by a reporter and/or other gene (e.g. one or more of the CRISPR-Cas system polynucleotide(s) of the present invention) and a strong poly A tail. When transposition occurs while using this vector or system thereof, the transposon can insert into an intron of a gene and the inserted reporter or other gene can provoke a mis-splicing process and as a result it in activates the trapped gene.

[0277] Any suitable transposon system can be used. Suitable transposon and systems thereof can include 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.

Non-Vector Delivery Vehicles

[0278] 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 particles (e.g., lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, metal nanoparticles, streptolysin O, multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Particles

[0279] In certain embodiments, the delivery vehicles are or comprise particles (e.g., delivery particles). Described in several example embodiments herein are particles comprising one or more cargo(s) and/or vector(s) as described herein. In certain example embodiments, a delivery particle comprises the Type V Cas polynucleotide and/or the at least one guide molecule as described herein, optionally in the form of a RNP complex as described herein. In certain embodiments, the delivery particle further comprises a donor polynucleotide, optionally hybridized to the at least one guide molecule. In certain example embodiments, a delivery particle comprises one or more nucleic acid molecules described herein, and/or one or more delivery vehicles (e.g., mRNA, viral vectors, plasmids) described herein. In certain example embodiments, a delivery particle comprises a nucleic acid molecule encoding the engineered Type V Cas polypeptide (e.g., a Cas12 polypeptide, or domain or fragment thereof having Cas12 activity), or a delivery vehicle comprising said nucleic acid molecule. In certain example embodiments, said delivery vehicle is a viral vector. In certain example embodiments, a delivery particle comprises a single nucleic acid molecule encoding each of the engineered Type V Cas polypeptide and at least one guide RNA, or a delivery vehicle comprising said nucleic acid molecule. In certain example embodiments, said delivery vehicle is a viral vector. In certain example embodiments, the delivery particle comprises a single nucleic acid molecule encoding each of the engineered Type V Cas polypeptide, the at least one guide RNA, and the one or more donor polynucleotides, or a delivery vehicle comprising said nucleic acid molecule. In certain example embodiments, said delivery vehicle is a viral vector.

[0280] For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension or greatest average dimension (e.g., diameter or greatest average

diameter) no greater than 1000 nm. 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).

[0281] Nanoparticles may also be used to deliver the compositions and systems to cells, as described in WO 2008042156, US 20130185823, and WO2015089419. In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain embodiments, nanoparticles of the invention have a greatest dimension or greatest average dimension (e.g., diameter or average diameter) of 500 nm or less. In other embodiments, nanoparticles of the invention have a greatest dimension or greatest average dimension ranging between 25 nm and 200 nm. In other embodiments, nanoparticles of the invention have a greatest dimension or greatest average dimension of 100 nm or less. In other embodiments, nanoparticles of the invention have a greatest dimension or greatest average dimensions ranging between 35 nm and 60 nm. It will be appreciated that reference made herein to particles or nanoparticles can be interchangeable, where appropriate. Nanoparticles made of semiconducting material may also be labeled quantum dots if they are small enough (typically sub 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present invention. Semi-solid and soft nanoparticles have been manufactured and are within the scope of the present invention. Nanoparticles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

[0282] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), ultraviolet-visible spectroscopy, dual polarization interferometry and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo

(herein cargo refers to e.g., one or more components of CRISPR-Cas system e.g., CRISPR enzyme or mRNA or guide RNA, or any combination thereof, and may include additional carriers and/or excipients) to provide particles of an optimal size for delivery for any *in vitro*, *ex vivo* and/or *in vivo* application of the present invention. In certain preferred embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS). Mention is made of US Patent No. 8,709,843; US Patent No. 6,007,845; US Patent No. 5,855,913; US Patent No. 5,985,309; US Patent No. 5,543,158; and the publication by James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84, describing particles, methods of making and using them and measurements thereof.

Lipid Particles

[0283] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, International Patent Publication Nos. WO 91/17424 and WO 91/16024. The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

Lipid nanoparticles (LNPs)

[0284] 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.

[0285] In some examples, LNPs may be used for delivering DNA molecules (e.g., those comprising coding sequences of Cas and/or gRNA) and/or RNA molecules (e.g., mRNA of Cas, gRNAs). In certain cases, LNPs may be used for delivering RNP complexes of Cas/gRNA.

[0286] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleoyloxy-3-N,N- dimethylaminopropane (DLinDMA), 1,2-dilinoleoyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-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-DOMG, 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).

[0287] In some embodiments, an LNP delivery vehicle can be used to deliver a virus particle containing a CRISPR-Cas system and/or component(s) thereof. In some embodiments, the virus particle(s) can be adsorbed to the lipid particle, such as through electrostatic interactions, and/or can be attached to the liposomes via a linker.

[0288] In some embodiments, the LNP contains a nucleic acid, wherein the charge ratio of nucleic acid backbone phosphates to cationic lipid nitrogen atoms is about 1: 1.5 – 7 or about 1:4.

[0289] In some embodiments, the LNP also includes a shielding compound, which is removable from the lipid composition under *in vivo* conditions. In some embodiments, the shielding compound is a biologically inert compound. In some embodiments, the shielding compound does not carry any charge on its surface or on the molecule as such. In some embodiments, the shielding compounds are polyethyleneglycoles (PEGs), hydroxyethylglucose (HEG) based polymers, polyhydroxyethyl starch (polyHES) and polypropylene. In some embodiments, the PEG, HEG, polyHES, and a polypropylene weight between about 500 to 10,000 Da or between about 2000 to 5000 Da. In some embodiments, the shielding compound is PEG2000 or PEG5000.

[0290] In some embodiments, the LNP can include one or more helper lipids. In some embodiments, the helper lipid can be a phosphor lipid or a steroid. In some embodiments, the helper lipid is between about 20 mol % to 80 mol % of the total lipid content of the composition. In some embodiments, the helper lipid component is between about 35 mol % to 65 mol % of the total lipid content of the LNP. In some embodiments, the LNP includes lipids at 50 mol% and the helper lipid at 50 mol% of the total lipid content of the LNP.

[0291] Other non-limiting, exemplary LNP delivery vehicles are described in U.S. Patent Publication Nos. US 20160174546, US 20140301951, US 20150105538, US 20150250725,

Wang et al., *J. Control Release*, 2017 Jan 31. pii: S0168-3659(17)30038-X. doi: 10.1016/j.jconrel.2017.01.037. [Epub ahead of print]; Altinoğlu et al., *Biomater Sci.*, 4(12):1773-80, Nov. 15, 2016; Wang et al., *PNAS*, 113(11):2868-73 March 15, 2016; Wang et al., *PloS One*, 10(11): e0141860. doi: 10.1371/journal.pone.0141860. eCollection 2015, Nov. 3, 2015; Takeda et al., *Neural Regen Res.* 10(5):689-90, May 2015; Wang et al., *Adv. Healthc Mater.*, 3(9):1398-403, Sep. 2014; and Wang et al., *Angew Chem Int Ed Engl.*, 53(11):2893-8, Mar. 10, 2014; James E. Dahlman and Carmen Barnes et al. *Nature Nanotechnology* (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84; Coelho et al., *N Engl J Med* 2013; 369:819-29; Aleku et al., *Cancer Res.*, 68(23): 9788-98 (Dec. 1, 2008), Strumberg et al., *Int. J. Clin. Pharmacol. Ther.*, 50(1): 76-8 (Jan. 2012), Schultheis et al., *J. Clin. Oncol.*, 32(36): 4141-48 (Dec. 20, 2014), and Fehring et al., *Mol. Ther.*, 22(4): 811-20 (Apr. 22, 2014); Novobrantseva, *Molecular Therapy–Nucleic Acids* (2012) 1, e4; doi:10.1038/mtna.2011.3; WO2012135025; US 20140348900; US 20140328759; US 20140308304; WO 2005/105152; WO 2006/069782; WO 2007/121947; US 2015/082080; US 20120251618; 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316;

Liposomes

[0292] In some embodiments, a lipid particle may be a 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 some embodiments, 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).

[0293] 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.

[0294] 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.

[0295] In some embodiments, a liposome delivery vehicle can be used to deliver a virus particle containing a CRISPR-Cas system and/or component(s) thereof. In some embodiments, the virus particle(s) can be adsorbed to the liposome, such as through electrostatic interactions, and/or can be attached to the liposomes via a linker.

[0296] In some embodiments, the liposome can be a Trojan Horse liposome (also known in the art as Molecular Trojan Horses), see e.g. <http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long>, the teachings of which can be applied and/or adapted to generate and/or deliver the CRISPR-Cas systems described herein.

[0297] Other non-limiting, exemplary liposomes can be those as set forth in Wang et al., ACS Synthetic Biology, 1, 403-07 (2012); Wang et al., PNAS, 113(11) 2868-2873 (2016); Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679; WO 2008/042973; US Pat. No. 8,071,082; WO 2014/186366; 20160257951; US20160129120; US 20160244761; 20120251618; WO2013/093648; Lipofectin (a combination of DOTMA and DOPE), Lipofectase, LIPOFECTAMINE.RTM. (e.g., LIPOFECTAMINE.RTM. 2000, LIPOFECTAMINE.RTM. 3000, LIPOFECTAMINE.RTM. RNAiMAX, LIPOFECTAMINE.RTM. LTX), SAINT-RED (Synvolux Therapeutics, Groningen Netherlands), DOPE, Cytofectin (Gilead Sciences, Foster City, Calif.), and Eufectins (JBL, San Luis Obispo, Calif.).

Stable nucleic-acid-lipid particles (SNALPs)

[0298] In some embodiments, the lipid particles may be stable nucleic acid lipid particles (SNALPs). SNALPs may comprise an ionizable lipid (DLinDMA) (e.g., cationic at low pH), a neutral helper lipid, cholesterol, a diffusible polyethylene glycol (PEG)-lipid, or any combination thereof. In some examples, SNALPs may comprise synthetic cholesterol, dipalmitoylphosphatidylcholine, 3-N-[(w-methoxy polyethylene glycol)2000]carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane. In some examples, SNALPs may comprise synthetic cholesterol, 1,2-distearoyl-sn-glycero-3-phosphocholine, PEG- cDMA, and 1,2-dilinoleyloxy-3-(N,N-dimethyl)aminopropane (DLinDMAo).

[0299] Other non-limiting, exemplary SNALPs that can be used to deliver the CRISPR-Cas systems described herein can be any such SNALPs as described in Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005, Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006; Geisbert et al., Lancet 2010; 375: 1896-905; Judge, J. Clin. Invest. 119:661-673

(2009); and Semple et al., Nature Nanotechnology, Volume 28 Number 2 February 2010, pp. 172-177.

Other Lipids

[0300] The lipid particles may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]- dioxolane (DLin-KC2-DMA), DLin-KC2-DMA4, C12- 200 and colipids disteoylphosphatidyl choline, cholesterol, and PEG-DMG.

[0301] In some embodiments, the delivery vehicle can be or include a lipidoid, such as any of those set forth in, for example, US 20110293703.

[0302] In some embodiments, the delivery vehicle can be or include an amino lipid, such as any of those set forth in, for example, Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 – 8533.

[0303] In some embodiments, the delivery vehicle can be or include a lipid envelope, such as any of those set forth in, for example, Korman et al., 2011. Nat. Biotech. 29:154-157.

Sugar-Based Particles

[0304] In some embodiments, the delivery vehicle can be a sugar-based particle. In some embodiments, the sugar-based particles can be or include GalNAc, such as any of those described in WO2014118272; US 20020150626; Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455.

Polymer-based Particles

[0305] In some embodiments, the delivery vehicles may comprise polymer-based particles (e.g., nanoparticles). In some embodiments, the polymer-based particles may mimic a viral mechanism of membrane fusion. The polymer-based particles may be a synthetic copy of Influenza virus machinery and form transfection complexes with various types of nucleic acids ((siRNA, miRNA, plasmid DNA or shRNA, mRNA) that cells take up via the endocytosis pathway, a process that involves the formation of an acidic compartment. The low pH in late endosomes acts as a chemical switch that renders the particle surface hydrophobic and facilitates membrane crossing. Once in the cytosol, the particle releases its payload for cellular action. This Active Endosome Escape technology is safe and maximizes transfection efficiency as it is using a natural uptake pathway. In some embodiments, the polymer-based particles may comprise alkylated and carboxyalkylated branched polyethylenimine. In some examples, the

polymer-based particles are VIROMER, e.g., VIROMER RNAi, VIROMER RED, VIROMER mRNA, VIROMER CRISPR. Example methods of delivering the systems and compositions herein include those described in Bawage SS et al., Synthetic mRNA expressed Cas13a mitigates RNA virus infections, www.biorxiv.org/content/10.1101/370460v1.full doi: doi.org/10.1101/370460, Viromer® RED, a powerful tool for transfection of keratinocytes. doi: 10.13140/RG.2.2.16993.61281, Viromer® Transfection - Factbook 2018: technology, product overview, users' data., doi:10.13140/RG.2.2.23912.16642. Other exemplary and non-limiting polymeric particles are described in US 20170079916, US 20160367686, US 20110212179, US 20130302401, 6,007,845, 5,855,913, 5,985,309, 5,543,158, WO2012135025, US 20130252281, US 20130245107, US 20130244279; US 20050019923, 20080267903.

Metal Nanoparticles

[0306] In some embodiments, the delivery vehicles comprise gold nanoparticles (also referred to AuNPs or colloidal gold). Gold nanoparticles may form complex with cargos, e.g., Cas:gRNA RNP. Gold nanoparticles may be coated, e.g., coated in a silicate and an endosomal disruptive polymer, PAsp(DET). Examples of gold nanoparticles include AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, and those described in Mout R, et al. (2017). ACS Nano 11:2452–8; Lee K, et al. (2017). Nat Biomed Eng 1:889–901. Other metal nanoparticles can also be complexed with cargo(s). Such metal particles include tungsten, palladium, rhodium, platinum, and iridium particles. Other non-limiting, exemplary metal nanoparticles are described in US 20100129793.

Multifunctional Envelope-Type Nanodevice (MEND)

[0307] The delivery vehicles may comprise multifunctional envelope-type nanodevice (MENDs). MENDs may comprise condensed plasmid DNA, a PLL core, and a lipid film shell. A MEND may further comprise cell-penetrating peptide (e.g., stearyl octaarginine) described herein. The cell penetrating peptide may be in the lipid shell. The lipid envelope may be modified with one or more functional components, e.g., one or more of: polyethylene glycol (e.g., to increase vascular circulation time), ligands for targeting of specific tissues/cells, additional cell-penetrating peptides (e.g., for greater cellular delivery), lipids to enhance endosomal escape, and nuclear delivery tags. In some examples, the MEND may be a tetra-lamellar MEND (T-MEND), which may target the cellular nucleus and mitochondria. In certain examples, a MEND may be a PEG-peptide-DOPE-conjugated MEND (PPD-MEND), which

may target bladder cancer cells. Examples of MENDs include those described in Kogure K, et al. (2004). *J Control Release* 98:317–23; Nakamura T, et al. (2012). *Acc Chem Res* 45:1113–21.

Lipid-coated mesoporous silica particles

[0308] The delivery vehicles may comprise lipid-coated mesoporous silica particles. Lipid-coated mesoporous silica particles may comprise a mesoporous silica nanoparticle core and a lipid membrane shell. The silica core may have a large internal surface area, leading to high cargo loading capacities. In some embodiments, pore sizes, pore chemistry, and overall particle sizes may be modified for loading different types of cargos. The lipid coating of the particle may also be modified to maximize cargo loading, increase circulation times, and provide precise targeting and cargo release. Examples of lipid-coated mesoporous silica particles include those described in Du X, et al. (2014). *Biomaterials* 35:5580–90; Durfee PN, et al. (2016). *ACS Nano* 10:8325–45.

Inorganic nanoparticles

[0309] The delivery vehicles may comprise inorganic nanoparticles. Examples of inorganic nanoparticles include carbon nanotubes (CNTs) (e.g., as described in Bates K and Kostarelos K. (2013). *Adv Drug Deliv Rev* 65:2023–33.), bare mesoporous silica nanoparticles (MSNPs) (e.g., as described in Luo GF, et al. (2014). *Sci Rep* 4:6064), and dense silica nanoparticles (SiNPs) (as described in Luo D and Saltzman WM. (2000). *Nat Biotechnol* 18:893–5).

Exosomes

[0310] 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.

[0311] 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.

[0312] Other non-limiting, exemplary exosomes include any of those set forth in Alvarez-Erviti et al. 2011, Nat Biotechnol 29: 341; [1401] El-Andaloussi et al. (Nature Protocols 7:2112–2126(2012); and Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130).

Spherical Nucleic Acids (SNAs)

[0313] In some embodiments, the delivery vehicle can be a SNA. SNAs are three dimensional nanostructures that can be composed of densely functionalized and highly oriented nucleic acids that can be covalently attached to the surface of spherical nanoparticle cores. The core of the spherical nucleic acid can impart the conjugate with specific chemical and physical properties, and it can act as a scaffold for assembling and orienting the oligonucleotides into a dense spherical arrangement that gives rise to many of their functional properties, distinguishing them from all other forms of matter. In some embodiments, the core is a crosslinked polymer. Non-limiting, exemplary SNAs can be any of those set forth in Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., and Small, 10:186-192.

Self-Assembling Nanoparticles

[0314] In some embodiments, the delivery vehicle is a self-assembling nanoparticle. The self-assembling nanoparticles can contain one or more polymers. The self-assembling nanoparticles can be PEGylated. Self-assembling nanoparticles are known in the art. Non-limiting, exemplary self-assembling nanoparticles can any as set forth in Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19, Bartlett et al. (PNAS, September 25, 2007, vol. 104, no. 39); Davis et al., Nature, Vol 464, 15 April 2010.

Lipoplexes/polyplexes

[0315] In some embodiments, the delivery vehicles comprise lipoplexes and/or polyplexes. Lipoplexes may bind to negatively charged cell membrane and induce endocytosis into the

cells. Examples of lipoplexes may be complexes comprising lipid(s) and non-lipid components. Examples of lipoplexes and polyplexes include FuGENE-6 reagent, a non-liposomal solution containing lipids and other components, zwitterionic amino lipids (ZALs), Ca²⁺ (e.g., forming DNA/Ca²⁺ microcomplexes), polyethenimine (PEI) (e.g., branched PEI), and poly(L-lysine) (PLL).

Cell Penetrating Peptides

[0316] In some embodiments, the delivery vehicles comprise cell penetrating peptides (CPPs). CPPs are short peptides that facilitate cellular uptake of various molecular cargo (e.g., from nanosized particles to small chemical molecules and large fragments of DNA).

[0317] CPPs may be of different sizes, amino acid sequences, and charges. In some examples, CPPs can translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPPs may be introduced into cells via different mechanisms, e.g., direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

[0318] CPPs may have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. Another type of CPPs is the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1). Examples of CPPs include Penetratin, Tat (48-60), Transportan, and (R-Ahx-R4) (Ahx refers to aminohexanoyl), Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β 3 signal peptide sequence, polyarginine peptide Arg5 sequence, Guanine rich-molecular transporters, and sweet arrow peptide. Examples of CPPs and related applications also include those described in US Patent 8,372,951.

[0319] CPPs can be used for *in vitro*, and *ex vivo* work quite readily, and extensive optimization for each cargo and cell type is usually required. In some examples, CPPs may be covalently attached to the Cas protein directly, which is then complexed with the gRNA and delivered to cells. In some examples, separate delivery of CPP–Cas and CPP–gRNA to multiple cells may be performed. CPP may also be used to delivery RNPs.

[0320] CPPs may be used to deliver the compositions and systems to plants. In some examples, CPPs may be used to deliver the components to plant protoplasts, which are then regenerated to plant cells and further to plants.

DNA Nanoclews

[0321] In some embodiments, the delivery vehicles comprise DNA nanoclews. A DNA nanoclew refers to a sphere-like structure of DNA (e.g., with a shape of a ball of yarn). The nanoclew may be synthesized by rolling circle amplification with palindromic sequences that aid in the self-assembly of the structure. The sphere may then be loaded with a payload. An example of DNA nanoclew is described in Sun W et al, J Am Chem Soc. 2014 Oct 22;136(42):14722-5; and Sun W et al, Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. DNA nanoclew may have a palindromic sequences to be partially complementary to the gRNA within the Cas:gRNA ribonucleoprotein complex. A DNA nanoclew may be coated, e.g., coated with PEI to induce endosomal escape.

iTOP

[0322] In some embodiments, the delivery vehicles comprise iTOP. iTOP refers to a combination of small molecules drives the highly efficient intracellular delivery of native proteins, independent of any transduction peptide. iTOP may be used for induced transduction by osmocytosis and propanebetaine, using NaCl-mediated hyperosmolality together with a transduction compound (propanebetaine) to trigger macropinocytotic uptake into cells of extracellular macromolecules. Examples of iTOP methods and reagents include those described in D'Astolfo DS, Pagliero RJ, Pras A, et al. (2015). Cell 161:674–690.

Streptolysin O (SLO)

[0323] The delivery vehicles may be streptolysin O (SLO). SLO is a toxin produced by Group A streptococci that works by creating pores in mammalian cell membranes. SLO may act in a reversible manner, which allows for the delivery of proteins (e.g., up to 100 kDa) to the cytosol of cells without compromising overall viability. Examples of SLO include those described in Sierig G, et al. (2003). Infect Immun 71:446–55; Walev I, et al. (2001). Proc Natl Acad Sci U S A 98:3185–90; Teng KW, et al. (2017). Elife 6:e25460.

Supercharged Proteins

[0324] In some embodiments, the delivery vehicle can be a supercharged protein. As used herein “Supercharged proteins” are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge. Non-limiting, exemplary

supercharged proteins can be any of those set forth in Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110–10112.

Targeted Delivery

[0325] In some embodiments, the delivery vehicle can allow for targeted delivery to a specific cell, tissue, organ, or system. In such embodiments, the delivery vehicle can include one or more targeting moieties that can direct targeted delivery of the cargo(s). In an embodiment, the delivery vehicle comprises a targeting moiety, such as active targeting of a lipid entity of the invention, e.g., lipid particle or nanoparticle or liposome or lipid bilayer of the invention comprising a targeting moiety for active targeting.

[0326] With regard to targeting moieties, mention is made of Deshpande et al, “Current trends in the use of liposomes for tumor targeting,” *Nanomedicine (Lond)*. 8(9), doi:10.2217/nmm.13.118 (2013), and the documents it cites, all of which are incorporated herein by reference and the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein. Mention is also made of International Patent Publication No. WO 2016/027264, and the documents it cites, all of which are incorporated herein by reference, the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein. And mention is made of Lorenzer et al, “Going beyond the liver: Progress and challenges of targeted delivery of siRNA therapeutics,” *Journal of Controlled Release*, 203: 1–15 (2015), and the documents it cites, all of which are incorporated herein by reference, the teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein.

[0327] An actively targeting lipid particle or nanoparticle or liposome or lipid bilayer delivery system (generally as to embodiments of the invention, “lipid entity of the invention” delivery systems) are prepared by conjugating targeting moieties, including small molecule ligands, peptides and monoclonal antibodies, on the lipid or liposomal surface; for example, certain receptors, such as folate and transferrin (Tf) receptors (TfR), are overexpressed on many cancer cells and have been used to make liposomes tumor cell specific. Liposomes that accumulate in the tumor microenvironment can be subsequently endocytosed into the cells by interacting with specific cell surface receptors. To efficiently target liposomes to cells, such as cancer cells, it is useful that the targeting moiety have an affinity for a cell surface receptor and to link the targeting moiety in sufficient quantities to have optimum affinity for the cell surface

receptors; and determining these embodiments are within the ambit of the skilled artisan. In the field of active targeting, there are a number of cell-, e.g., tumor-, specific targeting ligands.

[0328] Also, as to active targeting, with regard to targeting cell surface receptors such as cancer cell surface receptors, targeting ligands on liposomes can provide attachment of liposomes to cells, e.g., vascular cells, via a noninternalizing epitope; and this can increase the extracellular concentration of that which is being delivered, thereby increasing the amount delivered to the target cells. A strategy to target cell surface receptors, such as cell surface receptors on cancer cells, such as overexpressed cell surface receptors on cancer cells, is to use receptor-specific ligands or antibodies. Many cancer cell types display upregulation of tumor-specific receptors. For example, TfRs and folate receptors (FRs) are greatly overexpressed by many tumor cell types in response to their increased metabolic demand. Folic acid can be used as a targeting ligand for specialized delivery owing to its ease of conjugation to nanocarriers, its high affinity for FRs and the relatively low frequency of FRs, in normal tissues as compared with their overexpression in activated macrophages and cancer cells, e.g., certain ovarian, breast, lung, colon, kidney and brain tumors. Overexpression of FR on macrophages is an indication of inflammatory diseases, such as psoriasis, Crohn's disease, rheumatoid arthritis and atherosclerosis; accordingly, folate-mediated targeting of the invention can also be used for studying, addressing or treating inflammatory disorders, as well as cancers. Folate-linked lipid particles or nanoparticles or liposomes or lipid bilayers of the invention ("lipid entity of the invention") deliver their cargo intracellularly through receptor-mediated endocytosis. Intracellular trafficking can be directed to acidic compartments that facilitate cargo release, and, most importantly, release of the cargo can be altered or delayed until it reaches the cytoplasm or vicinity of target organelles. Delivery of cargo using a lipid entity of the invention having a targeting moiety, such as a folate-linked lipid entity of the invention, can be superior to nontargeted lipid entity of the invention. The attachment of folate directly to the lipid head groups may not be favorable for intracellular delivery of folate-conjugated lipid entity of the invention, since they may not bind as efficiently to cells as folate attached to the lipid entity of the invention surface by a spacer, which may can enter cancer cells more efficiently. A lipid entity of the invention coupled to folate can be used for the delivery of complexes of lipid, e.g., liposome, e.g., anionic liposome and virus or capsid or envelope or virus outer protein, such as those herein discussed such as adenovirus or AAV. Tf is a monomeric serum glycoprotein of approximately 80 KDa involved in the transport of iron throughout the body. Tf binds to the

TfR and translocates into cells via receptor-mediated endocytosis. The expression of TfR can be higher in certain cells, such as tumor cells (as compared with normal cells and is associated with the increased iron demand in rapidly proliferating cancer cells. Accordingly, the invention comprehends a TfR-targeted lipid entity of the invention, e.g., as to liver cells, liver cancer, breast cells such as breast cancer cells, colon such as colon cancer cells, ovarian cells such as ovarian cancer cells, head, neck, and lung cells, such as head, neck, and non-small-cell lung cancer cells, cells of the mouth such as oral tumor cells.

[0329] Also, as to active targeting, a lipid entity of the invention can be multifunctional, i.e., employ more than one targeting moiety such as CPP, along with Tf; a bifunctional system; e.g., a combination of Tf and poly-L-arginine which can provide transport across the endothelium of the blood–brain barrier. EGFR is a tyrosine kinase receptor belonging to the ErbB family of receptors that mediates cell growth, differentiation and repair in cells, especially non-cancerous cells, but EGF is overexpressed in certain cells such as many solid tumors, including colorectal, non-small-cell lung cancer, squamous cell carcinoma of the ovary, kidney, head, pancreas, neck and prostate, and especially breast cancer. The invention comprehends EGFR-targeted monoclonal antibody(ies) linked to a lipid entity of the invention. HER-2 is often overexpressed in patients with breast cancer, and is also associated with lung, bladder, prostate, brain, and stomach cancers. HER-2, encoded by the ERBB2 gene. The invention comprehends a HER-2-targeting lipid entity of the invention, e.g., an anti-HER-2-antibody(or binding fragment thereof)-lipid entity of the invention, a HER-2-targeting-PEGylated lipid entity of the invention (e.g., having an anti-HER-2-antibody or binding fragment thereof), a HER-2-targeting-maleimide-PEG polymer- lipid entity of the invention (e.g., having an anti-HER-2-antibody or binding fragment thereof). Upon cellular association, the receptor-antibody complex can be internalized by formation of an endosome for delivery to the cytoplasm.

[0330] With respect to receptor-mediated targeting, the skilled artisan takes into consideration ligand/target affinity and the quantity of receptors on the cell surface, and that PEGylation can act as a barrier against interaction with receptors. The use of antibody-lipid entity of the invention targeting can be advantageous. Multivalent presentation of targeting moieties can also increase the uptake and signaling properties of antibody fragments. In practice of the invention, the skilled person takes into account ligand density (e.g., high ligand densities on a lipid entity of the invention may be advantageous for increased binding to target cells). Preventing early by macrophages can be addressed with a sterically stabilized lipid

entity of the invention and linking ligands to the terminus of molecules such as PEG, which is anchored in the lipid entity of the invention (e.g., lipid particle or nanoparticle or liposome or lipid bilayer). The microenvironment of a cell mass such as a tumor microenvironment can be targeted; for instance, it may be advantageous to target cell mass vasculature, such as the tumor vasculature microenvironment. Thus, the invention comprehends targeting VEGF. VEGF and its receptors are well-known proangiogenic molecules and are well-characterized targets for antiangiogenic therapy. Many small-molecule inhibitors of receptor tyrosine kinases, such as VEGFRs or basic FGFRs, have been developed as anticancer agents and the invention comprehends coupling any one or more of these peptides to a lipid entity of the invention, e.g., phage IVO peptide(s) (e.g., via or with a PEG terminus), tumor-homing peptide APRPG (SEQ ID NO:48) such as APRPG-PEG-modified. VCAM, the vascular endothelium plays a key role in the pathogenesis of inflammation, thrombosis, and atherosclerosis. CAMs are involved in inflammatory disorders, including cancer, and are a logical target, E- and P-selectins, VCAM-1 and ICAMs. Can be used to target a lipid entity of the invention., e.g., with PEGylation.

[0331] Matrix metalloproteases (MMPs) belong to the family of zinc-dependent endopeptidases. They are involved in tissue remodeling, tumor invasiveness, resistance to apoptosis and metastasis. There are four MMP inhibitors called TIMP1–4, which determine the balance between tumor growth inhibition and metastasis; a protein involved in the angiogenesis of tumor vessels is MT1-MMP, expressed on newly formed vessels and tumor tissues. The proteolytic activity of MT1-MMP cleaves proteins, such as fibronectin, elastin, collagen, and laminin, at the plasma membrane and activates soluble MMPs, such as MMP-2, which degrades the matrix. An antibody or fragment thereof such as a Fab' fragment can be used in the practice of the invention such as for an antihuman MT1-MMP monoclonal antibody linked to a lipid entity of the invention, e.g., via a spacer such as a PEG spacer. $\alpha\beta$ -integrins or integrins are a group of transmembrane glycoprotein receptors that mediate attachment between a cell and its surrounding tissues or extracellular matrix.

[0332] Integrins contain two distinct chains (heterodimers) called α - and β -subunits. The tumor tissue-specific expression of integrin receptors can be utilized for targeted delivery in the invention, e.g., whereby the targeting moiety can be an RGD peptide such as a cyclic RGD.

[0333] Aptamers are ssDNA or RNA oligonucleotides that impart high affinity and specific recognition of the target molecules by electrostatic interactions, hydrogen bonding and hydrophobic interactions as opposed to the Watson–Crick base pairing, which is typical for the

bonding interactions of oligonucleotides. Aptamers as a targeting moiety can have advantages over antibodies: aptamers can demonstrate higher target antigen recognition as compared with antibodies; aptamers can be more stable and smaller in size as compared with antibodies; aptamers can be easily synthesized and chemically modified for molecular conjugation; and aptamers can be changed in sequence for improved selectivity and can be developed to recognize poorly immunogenic targets. Such moieties, such as an sgc8 aptamer, can be used as a targeting moiety (e.g., via covalent linking to the lipid entity of the invention, e.g., via a spacer, such as a PEG spacer).

[0334] Also, as to active targeting, the invention also comprehends intracellular delivery. Since liposomes follow the endocytic pathway, they are entrapped in the endosomes (pH 6.5–6) and subsequently fuse with lysosomes (pH <5), where they undergo degradation that results in a lower therapeutic potential. The low endosomal pH can be taken advantage of to escape degradation. Fusogenic lipids or peptides, which destabilize the endosomal membrane after the conformational transition/activation at a lowered pH. Amines are protonated at an acidic pH and cause endosomal swelling and rupture by a buffer effect. Unsaturated dioleoylphosphatidylethanolamine (DOPE) readily adopts an inverted hexagonal shape at a low pH, which causes fusion of liposomes to the endosomal membrane. This process destabilizes a lipid entity containing DOPE and releases the cargo into the cytoplasm; fusogenic lipid GALA (SEQ ID NO: 49), cholesteryl-GALA and PEG-GALA may show a highly efficient endosomal release; a pore-forming protein listeriolysin O may provide an endosomal escape mechanism; and, histidine-rich peptides have the ability to fuse with the endosomal membrane, resulting in pore formation, and can buffer the proton pump causing membrane lysis.

[0335] The invention comprehends a lipid entity of the invention modified with CPP(s), for intracellular delivery that may proceed via energy dependent macropinocytosis followed by endosomal escape. The invention further comprehends organelle-specific targeting. A lipid entity of the invention surface-functionalized with the triphenylphosphonium (TPP) moiety or a lipid entity of the invention with a lipophilic cation, rhodamine 123 can be effective in delivery of cargo to mitochondria. DOPE/sphingomyelin/stearyl-octa-arginine can deliver cargos to the mitochondrial interior via membrane fusion. A lipid entity of the invention surface modified with a lysosomotropic ligand, octadecyl rhodamine B can deliver cargo to lysosomes. Ceramides are useful in inducing lysosomal membrane permeabilization; the invention

comprehends intracellular delivery of a lipid entity of the invention having a ceramide. The invention further comprehends a lipid entity of the invention targeting the nucleus, e.g., via a DNA-intercalating moiety. The invention also comprehends multifunctional liposomes for targeting, i.e., attaching more than one functional group to the surface of the lipid entity of the invention, for instance to enhance accumulation in a desired site and/or promotes organelle-specific delivery and/or target a particular type of cell and/or respond to the local stimuli such as temperature (e.g., elevated), pH (e.g., decreased), respond to externally applied stimuli such as a magnetic field, light, energy, heat or ultrasound and/or promote intracellular delivery of the cargo. All of these are considered actively targeting moieties.

[0336] It should be understood that as to each possible targeting or active targeting moiety herein-discussed, there is an embodiment of the invention wherein the delivery system comprises such a targeting or active targeting moiety. Likewise, **Table 4** provides exemplary targeting moieties that can be used in the practice of the invention and as to each an embodiment of the invention provides a delivery system that comprises such a targeting moiety.

[0337] Table 4.

Targeting Moiety	Target Molecule	Target Cell or Tissue
folate	folate receptor	cancer cells
transferrin	transferrin receptor	cancer cells
Antibody CC52	rat CC531	rat colon adenocarcinoma CC531
anti- <i>HER2</i> antibody	<i>HER2</i>	<i>HER2</i> -overexpressing tumors
anti- <i>GD2</i>	<i>GD2</i>	neuroblastoma, melanoma
anti-EGFR	<i>EGFR</i>	tumor cells overexpressing EGFR
pH-dependent fusogenic peptide diINF-7		ovarian carcinoma
anti-VEGFR	VEGF Receptor	tumor vasculature
anti-CD19	CD19 (B cell marker)	leukemia, lymphoma
cell-penetrating peptide		blood-brain barrier
cyclic arginine-glycine-aspartic acid-tyrosine-cysteine peptide (c(RGDyC)-LP) (SEQ ID NO: 50)	avβ3	glioblastoma cells, human umbilical vein endothelial cells, tumor angiogenesis
ASSHN (SEQ ID NO: 51) peptide		endothelial progenitor cells; anti-cancer

PR_b peptide	$\alpha_5\beta_1$ integrin	cancer cells
AG86 peptide	$\alpha_6\beta_4$ integrin	cancer cells
KCCYSL (SEQ ID NO: 52) (P6.1 peptide)	HER-2 receptor	cancer cells
affinity peptide LN (YEVGHRC) (SEQ ID NO: 53)	Aminopeptidase N (APN/CD13)	APN-positive tumor
synthetic somatostatin analogue	Somatostatin receptor 2 (SSTR2)	breast cancer
anti-CD20 monoclonal antibody	B-lymphocytes	B cell lymphoma

[0338] Thus, in an embodiment of the delivery system, the targeting moiety comprises a receptor ligand, such as, for example, hyaluronic acid for CD44 receptor, galactose for hepatocytes, or antibody or fragment thereof such as a binding antibody fragment against a desired surface receptor, and as to each of a targeting moiety comprising a receptor ligand, or an antibody or fragment thereof such as a binding fragment thereof, such as against a desired surface receptor, there is an embodiment of the invention wherein the delivery system comprises a targeting moiety comprising a receptor ligand, or an antibody or fragment thereof such as a binding fragment thereof, such as against a desired surface receptor, or hyaluronic acid for CD44 receptor, galactose for hepatocytes (*see, e.g.*, Surace et al, “Lipoplexes targeting the CD44 hyaluronic acid receptor for efficient transfection of breast cancer cells,” J. Mol Pharm 6(4):1062-73; doi: 10.1021/mp800215d (2009); Sonoke et al, “Galactose-modified cationic liposomes as a liver-targeting delivery system for small interfering RNA,” Biol Pharm Bull. 34(8):1338-42 (2011); Torchilin, “Antibody-modified liposomes for cancer chemotherapy,” Expert Opin. Drug Deliv. 5 (9), 1003-1025 (2008); Manjappa et al, “Antibody derivatization and conjugation strategies: application in preparation of stealth immunoliposome to target chemotherapeutics to tumor,” J. Control. Release 150 (1), 2-22 (2011); Sofou S “Antibody-targeted liposomes in cancer therapy and imaging,” Expert Opin. Drug Deliv. 5 (2): 189-204 (2008); Gao J et al, “Antibody-targeted immunoliposomes for cancer treatment,” Mini. Rev. Med. Chem. 13(14): 2026-2035 (2013); Molavi et al, “Anti-CD30 antibody conjugated liposomal doxorubicin with significantly improved therapeutic efficacy against anaplastic large cell lymphoma,” Biomaterials 34(34):8718-25 (2013), each of which and the documents cited therein are hereby incorporated herein by reference), the

teachings of which can be applied and/or adapted for targeted delivery of one or more CRISPR-Cas molecules described herein.

[0339] Other exemplary targeting moieties are described elsewhere herein, such as epitope tags and the like.

Responsive Delivery

[0340] In some embodiments, the delivery vehicle can allow for responsive delivery of the cargo(s). Responsive delivery, as used in this context herein, refers to delivery of cargo(s) by the delivery vehicle in response to an external stimuli. Examples of suitable stimuli include, without limitation, an energy (light, heat, cold, and the like), a chemical stimuli (e.g. chemical composition, etc.), and a biologic or physiologic stimuli (e.g. environmental pH, osmolarity, salinity, biologic molecule, etc.). In some embodiments, the targeting moiety can be responsive to an external stimuli and facilitate responsive delivery. In other embodiments, responsiveness is determined by a non-targeting moiety component of the delivery vehicle.

[0341] The delivery vehicle can be stimuli-sensitive, e.g., sensitive to an externally applied stimuli, such as magnetic fields, ultrasound or light; and pH-triggering can also be used, e.g., a labile linkage can be used between a hydrophilic moiety such as PEG and a hydrophobic moiety such as a lipid entity of the invention, which is cleaved only upon exposure to the relatively acidic conditions characteristic of the a particular environment or microenvironment such as an endocytic vacuole or the acidotic tumor mass. pH-sensitive copolymers can also be incorporated in embodiments of the invention can provide shielding; diortho esters, vinyl esters, cysteine-cleavable lipopolymers, double esters and hydrazones are a few examples of pH-sensitive bonds that are quite stable at pH 7.5, but are hydrolyzed relatively rapidly at pH 6 and below, e.g., a terminally alkylated copolymer of N-isopropylacrylamide and methacrylic acid that copolymer facilitates destabilization of a lipid entity of the invention and release in compartments with decreased pH value; or, the invention comprehends ionic polymers for generation of a pH-responsive lipid entity of the invention (e.g., poly(methacrylic acid), poly(diethylaminoethyl methacrylate), poly(acrylamide) and poly(acrylic acid)).

[0342] Temperature-triggered delivery is also within the ambit of the invention. Many pathological areas, such as inflamed tissues and tumors, show a distinctive hyperthermia compared with normal tissues. Utilizing this hyperthermia is an attractive strategy in cancer therapy since hyperthermia is associated with increased tumor permeability and enhanced uptake. This technique involves local heating of the site to increase microvascular pore size

and blood flow, which, in turn, can result in an increased extravasation of embodiments of the invention. Temperature-sensitive lipid entity of the invention can be prepared from thermosensitive lipids or polymers with a low critical solution temperature. Above the low critical solution temperature (e.g., at site such as tumor site or inflamed tissue site), the polymer precipitates, disrupting the liposomes to release. Lipids with a specific gel-to-liquid phase transition temperature are used to prepare these lipid entities of the invention; and a lipid for a thermosensitive embodiment can be dipalmitoylphosphatidylcholine. Thermosensitive polymers can also facilitate destabilization followed by release, and a useful thermosensitive polymer is poly (N-isopropylacrylamide). Another temperature triggered system can employ lysolipid temperature-sensitive liposomes.

[0343] The invention also comprehends redox-triggered delivery. The difference in redox potential between normal and inflamed or tumor tissues, and between the intra- and extra-cellular environments has been exploited for delivery, e.g., GSH is a reducing agent abundant in cells, especially in the cytosol, mitochondria and nucleus. The GSH concentrations in blood and extracellular matrix are just one out of 100 to one out of 1000 of the intracellular concentration, respectively. This high redox potential difference caused by GSH, cysteine and other reducing agents can break the reducible bonds, destabilize a lipid entity of the invention and result in release of payload. The disulfide bond can be used as the cleavable/reversible linker in a lipid entity of the invention, because it causes sensitivity to redox owing to the disulfideto-thiol reduction reaction; a lipid entity of the invention can be made reduction sensitive by using two (e.g., two forms of a disulfide-conjugated multifunctional lipid as cleavage of the disulfide bond (e.g., via tris(2-carboxyethyl)phosphine, dithiothreitol, L-cysteine or GSH), can cause removal of the hydrophilic head group of the conjugate and alter the membrane organization leading to release of payload. Calcein release from reduction-sensitive lipid entity of the invention containing a disulfide conjugate can be more useful than a reduction-insensitive embodiment.

[0344] Enzymes can also be used as a trigger to release payload. Enzymes, including MMPs (e.g. MMP2), phospholipase A2, alkaline phosphatase, transglutaminase, or phosphatidylinositol-specific phospholipase C, have been found to be overexpressed in certain tissues, e.g., tumor tissues. In the presence of these enzymes, specially engineered enzyme-sensitive lipid entity of the invention can be disrupted and release the payload. an MMP2-

cleavable octapeptide (Gly-Pro-Leu-Gly-Ile-Ala-Gly-Gln (SEQ ID NO: 54)) can be incorporated into a linker, and can have antibody targeting, e.g., antibody 2C5.

[0345] The invention also comprehends light-or energy-triggered delivery, e.g., the lipid entity of the invention can be light-sensitive, such that light or energy can facilitate structural and conformational changes, which lead to direct interaction of the lipid entity of the invention with the target cells via membrane fusion, photo-isomerism, photofragmentation or photopolymerization; such a moiety therefor can be benzoporphyrin photosensitizer. Ultrasound can be a form of energy to trigger delivery; a lipid entity of the invention with a small quantity of particular gas, including air or perfluorated hydrocarbon can be triggered to release with ultrasound, e.g., low-frequency ultrasound (LFUS). Magnetic delivery: A lipid entity of the invention can be magnetized by incorporation of magnetites, such as Fe₃O₄ or γ -Fe₂O₃, e.g., those that are less than 10 nm in size. Targeted delivery can be then by exposure to a magnetic field.

GENETICALLY MODIFIED CELLS AND ORGANISMS

[0346] The present disclosure further provides cells comprising one or more of the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, and nucleic acid molecules and/or delivery vehicles (e.g., vectors, optionally viral vectors, or delivery particles described herein), and delivery systems disclosed herein. Also provided include cells modified by the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, delivery vehicles (e.g., vectors, optionally viral vectors, or delivery particles, described herein), and delivery systems, and/or using the methods disclosed herein, and cell cultures, tissues, organs, organism comprising such cells or progeny thereof. In certain example embodiments, a cell or population of cells comprises and/or is modified by a CRISPR-Cas system comprising the engineered Type V Cas polypeptide, one or more of the guide RNAs, and optionally one or more of the donor polynucleotides as described herein. In certain example embodiments, a cell or population of cells comprise a composition comprising one or more of the CRISPR-Cas systems.

[0347] In certain example embodiments, a cell or population of cells comprises and/or is modified by one or more nucleic acid molecules encoding the engineered Type V Cas polypeptide and the one or more guide RNAs and optionally the one or more donor polypeptides as described herein, one or more delivery vehicles (e.g., vectors, optionally viral vectors, or delivery particles, described herein) comprising the one or nucleic acid molecules

as described herein, and/or one or more delivery systems comprising the CRISPR-Cas systems, the compositions, and/or the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, a delivery particle, or the like) as disclosed herein.

[0348] In certain example embodiments, a cell or population of cells is modified by administering one or more of the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, delivery vehicles (e.g., vectors, optionally viral vectors, or delivery particles, described herein), and/or delivery systems disclosed herein. In certain example embodiments, one or more target polypeptides within the cell or population of cell is modified by administering one or more of the Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, delivery vehicles (e.g., vectors, optionally viral vectors, or delivery particles, described herein), and/or delivery systems disclosed herein, to the cell or population of cells.

[0349] 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.

[0350] 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 polynucleotide-targeting system are introduced into a host cell such that expression of the elements of the polynucleotide-targeting system direct formation of a polynucleotide-targeting complex at one or more target sites. In one embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0351] In particular embodiments, 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 disclosed 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 disclosed 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 disclosed herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0352] 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 disclosed 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.

[0353] In one 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 disclosed herein at least one tissue type of the plant or non-human animal. In one 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 disclosed herein in at least one tissue type. In one embodiment, the presence of the system components is transient, in that they are degraded over time. In one 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 one 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 one 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 one 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-Type V Cas molecule in the plant or non-human animal.

PHARMACEUTICAL FORMULATIONS

[0354] Also disclosed herein are pharmaceutical formulations that can contain an amount, effective amount, and/or least effective amount, and/or therapeutically effective amount of one or more compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof (which are also referred to as the primary active agent or ingredient disclosed elsewhere herein) described in greater detail elsewhere herein and a pharmaceutically acceptable carrier or excipient. As used herein, unless otherwise indicated, “pharmaceutical formulation” refers to the combination of an active agent, compound, or ingredient with a pharmaceutically acceptable carrier or excipient, making the composition suitable for diagnostic, therapeutic, or preventive use in vitro, in vivo, or ex vivo. As used herein, unless otherwise indicated, “pharmaceutically acceptable carrier or excipient” refers to a carrier or excipient that is useful in preparing a pharmaceutical formulation that is generally safe, non-toxic, and is neither biologically or otherwise undesirable, and includes a carrier or excipient that is acceptable for veterinary use as well as human pharmaceutical use. A “pharmaceutically acceptable carrier or excipient” as used in the specification and claims includes both one and more than one such carrier or excipient. When present, the compound can optionally be present in the pharmaceutical formulation as a pharmaceutically acceptable salt. In some embodiments, the pharmaceutical formulation can include, such as an active ingredient, a Type V Cas system or component thereof described in greater detail elsewhere herein. In some embodiments, the pharmaceutical formulation can include, such as an active ingredient, a Type V Cas polynucleotide described in greater detail elsewhere herein. In some embodiments, the pharmaceutical formulation can include, such as an active ingredient one or more modified cells, such as one or more modified cells described in greater detail elsewhere herein.

[0355] In some embodiments, the active ingredient is present as a pharmaceutically acceptable salt of the active ingredient. As used herein, unless otherwise indicated, “pharmaceutically acceptable salt” refers to any acid or base addition salt whose counter-ions are non-toxic to the subject to which they are administered in pharmaceutical doses of the salts. Suitable salts include, hydrobromide, iodide, nitrate, bisulfate, phosphate, isonicotinate, lactate, salicylate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate,

succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate, camphorsulfonate, naphthalenesulfonate, propionate, malonate, mandelate, malate, phthalate, and pamoate.

[0356] The pharmaceutical formulations disclosed herein can be administered to a subject in need thereof via any suitable method or route to a subject in need thereof. Suitable administration routes can include, but are not limited to auricular (otic), buccal, conjunctival, cutaneous, dental, electro-osmosis, endocervical, endosinusial, endotracheal, enteral, epidural, extra-amniotic, extracorporeal, hemodialysis, infiltration, interstitial, intra-abdominal, intra-amniotic, intra-arterial, intra-articular, intrabiliary, intrabronchial, intrabursal, intracardiac, intracartilaginous, intracaudal, intracavernous, intracavitary, intracerebral, intracisternal, intracorneal, intracoronar (dental), intracoronary, intracorporus cavernosum, intradermal, intradiscal, intraductal, intraduodenal, intradural, intraepidermal, intraesophageal, intragastric, intragingival, intraileal, intralesional, intraluminal, intralymphatic, intramedullary, intrameningeal, intramuscular, intraocular, intraovarian, intrapericardial, intraperitoneal, intrapleural, intraprostatic, intrapulmonary, intrasinal, intraspinal, intrasynovial, intratendinous, intratesticular, intrathecal, intrathoracic, intratubular, intratumor, intratympanic, intrauterine, intravascular, intravenous, intravenous bolus, intravenous drip, intraventricular, intravesical, intravitreal, iontophoresis, irrigation, laryngeal, nasal, nasogastric, occlusive dressing technique, ophthalmic, oral, oropharyngeal, other, parenteral, percutaneous, periarticular, peridural, perineural, periodontal, rectal, respiratory (inhalation), retrobulbar, soft tissue, subarachnoid, subconjunctival, subcutaneous, sublingual, submucosal, topical, transdermal, transmucosal, transplacental, transtracheal, transtympanic, ureteral, urethral, and/or vaginal administration, and/or any combination of the above administration routes, which typically depends on the disease to be treated and/or the active ingredient(s).

[0357] Where appropriate, compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof described in greater detail elsewhere herein can be provided to a subject in need thereof as an ingredient, such as an active ingredient or agent, in a pharmaceutical formulation. As such, also described are pharmaceutical formulations containing one or more of the compounds and salts thereof, or pharmaceutically acceptable salts thereof disclosed herein. Suitable salts include, hydrobromide, iodide, nitrate, bisulfate, phosphate, isonicotinate, lactate, salicylate, acid citrate, tartrate, oleate, tannate, pantothenate,

bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucuronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate, camphorsulfonate, naphthalenesulfonate, propionate, malonate, mandelate, malate, phthalate, and pamoate.

[0358] In some embodiments, the subject in need thereof has or is suspected of having a genetic or epigenetic disease or condition. In some embodiments, the subject in need thereof has or is suspected of having a hematopoietic disease or a symptom thereof. In some embodiments, the subject in need thereof has or is suspected of having, a neurobiological disease or disorder, a psychiatric disease or disorder, a cancer, an autoimmune or immune disease or disorder, a thrombosis disease, a heart disease, a kidney disease, a lung disease, a brain disease, a musculoskeletal disease, a bone disease, a muscle disease, a pancreatic disease, a liver disease, an intestinal disease, a stomach disease, an esophageal disease, an ear disease, an oral disease, a skin disease, a nose or sinus disease, or a blood vessel disease, or any combination thereof. Exemplary diseases are described elsewhere herein. As used herein, unless otherwise indicated, “agent” refers to any substance, compound, molecule, and the like, which can be biologically active or otherwise can induce a biological and/or physiological effect on a subject to which it is administered to. As used herein, unless otherwise indicated, “active agent” or “active ingredient” refers to a substance, compound, or molecule, which is biologically active or otherwise, induces a biological or physiological effect on a subject to which it is administered to. In other words, “active agent” or “active ingredient” refers to a component or components of a composition to which the whole or part of the effect of the composition is attributed. An agent can be a primary active agent, or in other words, the component(s) of a composition to which the whole or part of the effect of the composition is attributed. An agent can be a secondary agent, or in other words, the component(s) of a composition to which an additional part and/or other effect of the composition is attributed.

Pharmaceutically Acceptable Carriers and Secondary Ingredients and Agents

[0359] The pharmaceutical formulation can include a pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers include, but are not limited to, water, salt solutions, alcohols, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxy methylcellulose, and polyvinyl pyrrolidone, which do not deleteriously react with the active composition.

[0360] The pharmaceutical formulations can be sterilized, and if desired, mixed with agents, such as lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances, and the like which do not deleteriously react with the active compound.

[0361] In some embodiments, the pharmaceutical formulation can also include an effective amount of secondary active agents, including but not limited to, biologic agents or molecules including, but not limited to, e.g., polynucleotides, amino acids, peptides, polypeptides, antibodies, aptamers, ribozymes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics, analgesics, antispasmodics, anti-inflammatories, anti-histamines, anti-infectives, chemotherapeutics, and combinations thereof.

Effective Amounts

[0362] In some embodiments, the amount of the primary active agent and/or optional secondary agent can be an effective amount, least effective amount, and/or therapeutically effective amount. As used herein, unless otherwise indicated, “effective amount” refers to the amount of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieve one or more therapeutic effects or desired effect. As used herein, unless otherwise indicated, “least effective” amount refers to the lowest amount of the primary and/or optional secondary agent that achieves the one or more therapeutic or other desired effects. As used herein, unless otherwise indicated, “therapeutically effective amount” refers to the amount of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieves one or more therapeutic effects. In some embodiments, the one or more therapeutic effects are to modify one or more polynucleotides.

[0363] The effective amount, least effective amount, and/or therapeutically effective amount of the primary and optional secondary active agent described elsewhere herein contained in the pharmaceutical formulation can be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 pg, ng, µg, mg, or g or be any numerical value or subrange within any of these ranges.

[0364] In some embodiments, the effective amount, least effective amount, and/or therapeutically effective amount can be an effective concentration, least effective concentration, and/or therapeutically effective concentration, which can each be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 pM, nM, μ M, mM, or M or be any numerical value or subrange within any of these ranges.

[0365] In other embodiments, the effective amount, least effective amount, and/or therapeutically effective amount of the primary and optional secondary active agent be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960, 970, 980, 990, 1000 IU or be any numerical value or subrange within any of these ranges.

[0366] In some embodiments, the primary and/or the optional secondary active agent present in the pharmaceutical formulation can be any non-zero amount ranging from about 0 to 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.9, to 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, 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, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w,

v/v, or w/v of the pharmaceutical formulation or be any numerical value or subrange within any of these ranges.

[0367] In some embodiments where a cell or cell population is present in the pharmaceutical formulation (e.g., as a primary and/or or secondary active agent), the effective amount of cells can be any amount ranging from about 1 or 2 cells to $1 \times 10^1/\text{mL}$, $1 \times 10^{20}/\text{mL}$ or more, such as about $1 \times 10^1/\text{mL}$, $1 \times 10^2/\text{mL}$, $1 \times 10^3/\text{mL}$, $1 \times 10^4/\text{mL}$, $1 \times 10^5/\text{mL}$, $1 \times 10^6/\text{mL}$, $1 \times 10^7/\text{mL}$, $1 \times 10^8/\text{mL}$, $1 \times 10^9/\text{mL}$, $1 \times 10^{10}/\text{mL}$, $1 \times 10^{11}/\text{mL}$, $1 \times 10^{12}/\text{mL}$, $1 \times 10^{13}/\text{mL}$, $1 \times 10^{14}/\text{mL}$, $1 \times 10^{15}/\text{mL}$, $1 \times 10^{16}/\text{mL}$, $1 \times 10^{17}/\text{mL}$, $1 \times 10^{18}/\text{mL}$, $1 \times 10^{19}/\text{mL}$, to/or about $1 \times 10^{20}/\text{mL}$ or any numerical value or subrange within any of these ranges.

[0368] In some embodiments, the amount or effective amount, particularly where an infective particle is being delivered (e.g., a virus particle having the primary or secondary agent as a cargo), the effective amount of virus particles can be expressed as a titer (plaque forming units per unit of volume) or as a MOI (multiplicity of infection). In some embodiments, the effective amount can be about 1×10^1 particles per pL, nL, μL , mL, or L to 1×10^{20} particles per pL, nL, μL , mL, or L or more, such as about 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , to/or about 1×10^{20} particles per pL, nL, μL , mL, or L. In some embodiments, the effective titer can be about 1×10^1 transforming units per pL, nL, μL , mL, or L to 1×10^{20} transforming units per pL, nL, μL , mL, or L or more, such as about 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , to/or about 1×10^{20} transforming units per pL, nL, μL , mL, or L or any numerical value or subrange within these ranges. In some embodiments, the MOI of the pharmaceutical formulation can range from about 0.1 to 10 or more, such as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 10 or more or any numerical value or subrange within these ranges.

[0369] In some embodiments, the amount or effective amount of the one or more of the active agent(s) disclosed herein contained in the pharmaceutical formulation can range from about 1 pg/kg to about 10 mg/kg based upon the bodyweight of the subject in need thereof or

average bodyweight of the specific patient population to which the pharmaceutical formulation can be administered.

[0370] In embodiments where there is a secondary agent contained in the pharmaceutical formulation, the effective amount of the secondary active agent will vary depending on the secondary agent, the primary agent, the administration route, subject age, disease, stage of disease, among other things, which will be one of ordinary skill in the art.

[0371] When optionally present in the pharmaceutical formulation, the secondary active agent can be included in the pharmaceutical formulation or can exist as a stand-alone compound or pharmaceutical formulation that can be administered contemporaneously or sequentially with the compound, derivative thereof, or pharmaceutical formulation thereof.

[0372] In some embodiments, the effective amount of the secondary active agent, when optionally present, is any non-zero amount ranging from about 0 to 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, 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, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total active agents present in the pharmaceutical formulation or any numerical value or subrange within these ranges. In additional embodiments, the effective amount of the secondary active agent is any non-zero amount ranging from about 0 to 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, 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, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total pharmaceutical formulation or any numerical value or subrange within these ranges.

Dosage Forms

[0373] In some embodiments, the pharmaceutical formulations disclosed herein can be provided in a dosage form. The dosage form can be administered to a subject in need thereof. The dosage form can be effective generate specific concentration, such as an effective concentration, at a given site in the subject in need thereof. As used herein, unless otherwise indicated, “dose,” “unit dose,” or “dosage” can refer to physically discrete units suitable for use in a subject, each unit containing a predetermined quantity of the primary active agent, and

optionally present secondary active ingredient, and/or a pharmaceutical formulation thereof calculated to produce the desired response or responses in association with its administration. In some embodiments, the given site is proximal to the administration site. In some embodiments, the given site is distal to the administration site. In some cases, the dosage form contains a greater amount of one or more of the active ingredients present in the pharmaceutical formulation than the final intended amount needed to reach a specific region or location within the subject to account for loss of the active components such as via first and second pass metabolism.

[0374] The dosage forms can be adapted for administration by any appropriate route. Appropriate routes include, but are not limited to, oral (including buccal or sublingual), rectal, intraocular, inhaled, intranasal, topical (including buccal, sublingual, or transdermal), vaginal, parenteral, subcutaneous, intramuscular, intravenous, internasal, and intradermal. Other appropriate routes are described elsewhere herein. Such formulations can be prepared by any method known in the art.

[0375] Dosage forms adapted for oral administration can discrete dosage units such as capsules, pellets or tablets, powders or granules, solutions, or suspensions in aqueous or non-aqueous liquids; edible foams or whips, or in oil-in-water liquid emulsions or water-in-oil liquid emulsions. In some embodiments, the pharmaceutical formulations adapted for oral administration also include one or more agents which flavor, preserve, color, or help disperse the pharmaceutical formulation. Dosage forms prepared for oral administration can also be in the form of a liquid solution that can be delivered as a foam, spray, or liquid solution. The oral dosage form can be administered to a subject in need thereof. Where appropriate, the dosage forms disclosed herein can be microencapsulated.

[0376] The dosage form can also be prepared to prolong or sustain the release of any ingredient. In some embodiments, compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof disclosed herein can be the ingredient whose release is delayed. In some embodiments the primary active agent is the ingredient whose release is delayed. In some embodiments, an optional secondary agent can be the ingredient whose release is delayed. Suitable methods for delaying the release of an ingredient include, but are not limited to, coating or embedding the ingredients in material in polymers, wax, gels, and the like. Delayed release dosage formulations can be prepared as described in standard references such as "Pharmaceutical dosage form tablets," eds. Liberman et. al. (New York, Marcel

Dekker, Inc., 1989), "Remington - The science and practice of pharmacy", 20th ed., Lippincott Williams & Wilkins, Baltimore, MD, 2000, and "Pharmaceutical dosage forms and drug delivery systems", 6th Edition, Ansel et al., (Media, PA: Williams and Wilkins, 1995). These references provide information on excipients, materials, equipment, and processes for preparing tablets and capsules and delayed release dosage forms of tablets and pellets, capsules, and granules. The delayed release can be anywhere from about an hour to about 3 months or more.

[0377] Examples of suitable coating materials include, but are not limited to, cellulose polymers such as cellulose acetate phthalate, hydroxypropyl cellulose, hydroxypropyl methylcellulose, hydroxypropyl methylcellulose phthalate, and hydroxypropyl methylcellulose acetate succinate; polyvinyl acetate phthalate, acrylic acid polymers and copolymers, and methacrylic resins that are commercially available under the trade name EUDRAGIT® (Roth Pharma, Westerstadt, Germany), zein, shellac, and polysaccharides.

[0378] Coatings may be formed with a different ratio of water-soluble polymer, water insoluble polymers, and/or pH dependent polymers, with or without water insoluble/water soluble non-polymeric excipient, to produce the desired release profile. The coating is either performed on the dosage form (matrix or simple) which includes, but is not limited to, tablets (compressed with or without coated beads), capsules (with or without coated beads), beads, particle compositions, "ingredient as is" formulated as, but not limited to, suspension form or as a sprinkle dosage form.

[0379] Where appropriate, the dosage forms disclosed herein can be a liposome. In these embodiments, primary active ingredient(s), and/or optional secondary active ingredient(s), and/or pharmaceutically acceptable salt thereof where appropriate are incorporated into a liposome. In embodiments where the dosage form is a liposome, the pharmaceutical formulation is thus a liposomal formulation. The liposomal formulation can be administered to a subject in need thereof.

[0380] Dosage forms adapted for topical administration can be formulated as ointments, creams, suspensions, lotions, powders, solutions, pastes, gels, sprays, aerosols, or oils. In some embodiments for treatments of the eye or other external tissues, for example the mouth or the skin, the pharmaceutical formulations are applied as a topical ointment or cream. When formulated in an ointment, a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be formulated with a

paraffinic or water-miscible ointment base. In other embodiments, the primary and/or secondary active ingredient can be formulated in a cream with an oil-in-water cream base or a water-in-oil base. Dosage forms adapted for topical administration in the mouth include lozenges, pastilles, and mouth washes.

[0381] Dosage forms adapted for nasal or inhalation administration include aerosols, solutions, suspension drops, gels, or dry powders. In some embodiments, a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be in a dosage form adapted for inhalation is in a particle-size-reduced form that is obtained or obtainable by micronization. In some embodiments, the particle size of the size reduced (e.g., micronized) compound or salt or solvate thereof, is defined by a D_{50} value of about 0.5 to about 10 microns as measured by an appropriate method known in the art. Dosage forms adapted for administration by inhalation also include particle dusts or mists. Suitable dosage forms wherein the carrier or excipient is a liquid for administration as a nasal spray or drops include aqueous or oil solutions/suspensions of an active (primary and/or secondary) ingredient, which may be generated by various types of metered dose pressurized aerosols, nebulizers, or insufflators. The nasal/inhalation formulations can be administered to a subject in need thereof.

[0382] In some embodiments, the dosage forms are aerosol formulations suitable for administration by inhalation. In some of these embodiments, the aerosol formulation contains a solution or fine suspension of a primary active ingredient, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate and a pharmaceutically acceptable aqueous or non-aqueous solvent. Aerosol formulations can be presented in single or multi-dose quantities in sterile form in a sealed container. For some of these embodiments, the sealed container is a single dose or multi-dose nasal or an aerosol dispenser fitted with a metering valve (e.g., metered dose inhaler), which is intended for disposal once the contents of the container have been exhausted.

[0383] Where the aerosol dosage form is contained in an aerosol dispenser, the dispenser contains a suitable propellant under pressure, such as compressed air, carbon dioxide, or an organic propellant, including but not limited to a hydrofluorocarbon. The aerosol formulation dosage forms in other embodiments are contained in a pump-atomizer. The pressurized aerosol formulation can also contain a solution or a suspension of a primary active ingredient, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof. In further

embodiments, the aerosol formulation also contains co-solvents and/or modifiers incorporated to improve, for example, the stability and/or taste and/or fine particle mass characteristics (amount and/or profile) of the formulation. Administration of the aerosol formulation can be once daily or several times daily, for example 2, 3, 4, or 8 times daily, in which 1, 2, 3 or more doses are delivered each time. The aerosol formulations can be administered to a subject in need thereof.

[0384] For some dosage forms suitable and/or adapted for inhaled administration, the pharmaceutical formulation is a dry powder inhalable-formulations. In addition to a primary active agent, optional secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate, such a dosage form can contain a powder base such as lactose, glucose, trehalose, mannitol, and/or starch. In some of these embodiments, a primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate is in a particle-size reduced form. In further embodiments, a performance modifier, such as L-leucine or another amino acid, cellobiose octaacetate, and/or metals salts of stearic acid, such as magnesium or calcium stearate. In some embodiments, the aerosol formulations are arranged so that each metered dose of aerosol contains a predetermined amount of an active ingredient, such as the one or more of the compositions, compounds, vector(s), molecules, cells, and combinations thereof disclosed herein.

[0385] Dosage forms adapted for vaginal administration can be presented as pessaries, tampons, creams, gels, pastes, foams, or spray formulations. Dosage forms adapted for rectal administration include suppositories or enemas. The vaginal formulations can be administered to a subject in need thereof.

[0386] Dosage forms adapted for parenteral administration and/or adapted for injection can include aqueous and/or non-aqueous sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, solutes that render the composition isotonic with the blood of the subject, and aqueous and non-aqueous sterile suspensions, which can include suspending agents and thickening agents. The dosage forms adapted for parenteral administration can be presented in a single-unit dose or multi-unit dose containers, including but not limited to sealed ampoules or vials. The doses can be lyophilized and re-suspended in a sterile carrier to reconstitute the dose prior to administration. Extemporaneous injection solutions and suspensions can be prepared in some embodiments, from sterile powders, granules, and tablets. The parenteral formulations can be administered to a subject in need thereof.

[0387] For some embodiments, the dosage form contains a predetermined amount of a primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate per unit dose. In an embodiment, the predetermined amount of primary active agent, secondary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate can be an effective amount, a least effect amount, and/or a therapeutically effective amount. In other embodiments, the predetermined amount of a primary active agent, secondary active agent, and/or pharmaceutically acceptable salt thereof where appropriate, can be an appropriate fraction of the effective amount of the active ingredient.

Co-Therapies and Combination Therapies

[0388] In some embodiments, the pharmaceutical formulation(s) disclosed herein are part of a combination treatment or combination therapy. The combination treatment can include the pharmaceutical formulation disclosed herein and an additional treatment modality. The additional treatment modality can be a chemotherapeutic, a biological therapeutic, surgery, radiation, diet modulation, environmental modulation, a physical activity modulation, and combinations thereof.

[0389] In some embodiments, the co-therapy or combination therapy can additionally include but not limited to, polynucleotides, amino acids, peptides, polypeptides, antibodies, aptamers, ribozymes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics, analgesics, antispasmodics, anti-inflammatories, anti-histamines, anti-infectives, chemotherapeutics, and combinations thereof.

Administration of the Pharmaceutical Formulations

[0390] The pharmaceutical formulations or dosage forms thereof disclosed herein can be administered one or more times hourly, daily, monthly, or yearly (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or more times hourly, daily, monthly, or yearly). In some embodiments, the pharmaceutical formulations or dosage forms thereof disclosed herein can be administered continuously over a period of time ranging from minutes to hours to days. Devices and dosages forms are known in the art and disclosed herein that are effective to provide continuous administration of the pharmaceutical formulations disclosed herein. In some embodiments, the first one or a few initial amount(s) administered can be a higher dose than subsequent doses. This is typically referred to in the art as a loading dose or doses and a maintenance dose, respectively. In some embodiments, the pharmaceutical formulations can

be administered such that the doses over time are tapered (increased or decreased) overtime so as to wean a subject gradually off of a pharmaceutical formulation or gradually introduce a subject to the pharmaceutical formulation.

[0391] As previously discussed, the pharmaceutical formulation can contain a predetermined amount of a primary active agent, secondary active agent, and/or pharmaceutically acceptable salt thereof where appropriate. In some of these embodiments, the predetermined amount can be an appropriate fraction of the effective amount of the active ingredient. Such unit doses may therefore be administered once or more than once a day, month, or year (e.g., 1, 2, 3, 4, 5, 6, or more times per day, month, or year). Such pharmaceutical formulations may be prepared by any of the methods well known in the art.

[0392] Where co-therapies or multiple pharmaceutical formulations are to be delivered to a subject, the different therapies or formulations can be administered sequentially or simultaneously. Sequential administration is administration where an appreciable amount of time occurs between administrations, such as more than about 15, 20, 30, 45, 60 minutes or more. The time between administrations in sequential administration can be on the order of hours, days, months, or even years, depending on the active agent present in each administration. Simultaneous administration refers to administration of two or more formulations at the same time or substantially at the same time (e.g., within seconds or just a few minutes apart), where the intent is that the formulations be administered together at the same time.

APPLICATIONS AND METHODS OF USE IN GENERAL

[0393] The Type V Cas polypeptide, the CRISRP-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment 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 polynucleotides, tracing of target polynucleotides, isolation of target polynucleotides, visualization of target polynucleotides, etc.

[0394] Aspects of the invention thus also encompass methods and uses of the Type V Cas polypeptide, the CRISRP-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment 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.

[0395] Typically, in the context of a polynucleotide-targeting system, formation of a polynucleotide-targeting complex (comprising a guide molecule hybridized to a target sequence and complexed with one or more polynucleotide-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, unless otherwise indicated, 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).

Targeting and Modifying Target Polynucleotides

[0396] In one embodiment, the present disclosure provides a method of targeting a polynucleotide, comprising contacting a sample (such as a cell, a population of cells, a tissue, an organ, or an organism) that comprises a target polynucleotide with the Type V Cas polypeptide, the CRISRP-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment herein. 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 target polynucleotide is a disease-associated target sequence.

[0397] In one embodiment, the present disclosure provides a method of modifying target polynucleotides comprising delivering the Type V Cas polypeptide, the CRISRP-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment herein 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 molecule into the target polynucleotide.

[0398] 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.

[0399] 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 TAM (targeted adjacent motif); that is, a short sequence recognized by the complex. The precise sequence and length requirements for the TAM differ depending on the Type V Cas polypeptide used, but TAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). TAM specificity can be determined, for example according to the experimental setup described in Figure 8. In one embodiment, the TAM sequence comprises TCA. In embodiments, the TAM sequence is TCAN, wherein N may comprise any nucleotide. In one embodiment the TAM sequence comprises TCAG or TCAT. A skilled person will be able to identify further TAM sequences for use with a given Type V Cas polypeptide. Further, engineering of the TAM Interacting (PI) domain may allow programming of TAM specificity, improve target site recognition fidelity, and increase the versatility of the Type V Cas polypeptide, genome engineering platform. Type V Cas polypeptide may be engineered to alter their TAM specificity, for example as described in Nishimasu et al. Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1. *Mol Cell*. 67(1):139-147.e2. (2017); Tóth et al. “Improved LbCas12a variants with

altered PAM specificities further broaden the genome targeting range of Cas12a nucleases.” Nucleic Acids Res. (2020) 48(7):3722-3733; Kim et al. “Enhancement of target specificity of CRISPR-Cas12a by using a chimeric DNA-RNA guide.” Nucleic Acids Res. (2020) 48(15):8601-8616; Chen et al. “A Cas12a ortholog with stringent PAM recognition followed by low off-target editing rates for genome editing.” Genome Biol (2020) 21, 78.

[0400] 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.

[0401] Aspects of the invention relate to a method of targeting a polynucleotide, comprising contacting a sample that comprises the target polynucleotide with the Type V Cas polypeptide, the CRISPR-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment herein. In one embodiment, a target polynucleotide is contacted with at least two different compositions, CRISPR-Cas systems, or Type V Cas polypeptides. In further embodiments, the two different Type V Cas polypeptides have different target polynucleotide specificities, or degrees of specificity. In one embodiment, the two different Type V Cas polypeptides have a different TAM specificity.

[0402] Also envisaged are methods of targeting a polynucleotide, comprising contacting a sample that comprises the target polynucleotide with the Type V Cas polypeptide, the CRISPR-Cas systems, the compositions, the delivery vehicles (e.g., a nucleic acid molecule, a plasmid, mRNA, a vector, optionally a viral vector, and/or a delivery particle), and/or the delivery systems as described in any embodiment herein wherein contacting results in modification of

a gene product or modification of the amount or expression of a gene product. In one embodiment, the expression of the targeted gene product is increased by the method. In one 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%, p at least 90%, at least 95%, 100%. In one 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 one 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 one 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.

[0403] In one embodiment, one or more polynucleotide molecules, vectors, or vector systems driving expression of one or more elements of a polynucleotide-targeting system or delivery systems comprising one or more elements of the polynucleotide-targeting system are introduced into a host cell such that expression of the elements of the polynucleotide-targeting system direct formation of a polynucleotide-targeting complex at one or more target sites. In one embodiment of the invention the host cell may be a eukaryotic cell, a prokaryotic cell, or a plant cell.

[0404] In particular embodiments, 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 disclosed 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 disclosed 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 disclosed herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0405] 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.

[0406] In one 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 one embodiment, the presence of the compositions is transient, in that they are degraded over time. In one 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 one 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 one 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 one 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.

[0407] In one aspect, the invention provides methods for using one or more elements of a polynucleotide-targeting system. The polynucleotide-targeting complex of the invention

provides an effective means for modifying a target DNA or RNA (single or double stranded, linear or super-coiled). The polynucleotide-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 polynucleotide-targeting complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary polynucleotide-targeting complex comprises a DNA or RNA-targeting effector protein complexed with a guide molecule hybridized to a target sequence within the target locus of interest.

[0408] In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method may comprise modifying a target polynucleotide using a polynucleotide-targeting complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. In an embodiment, the polynucleotide-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, e.g.*, 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 polynucleotide-targeting complex, the break is repaired via homologous recombination with a 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 an RNA in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a polynucleotide-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 polynucleotide-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 polynucleotide-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 polynucleotide-targeting complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target RNA can be an 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).

[0409] In one embodiment, the method may comprise allowing compositions 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 polynucleotide-targeting complex comprises a polynucleotide-targeting effector protein complexed with a guide molecule 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 polynucleotide-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 polynucleotide-targeting complex comprises a polynucleotide-targeting effector protein complexed with a guide molecule. 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. Polynucleotide identifiers are non-coding polynucleotides that may be used to

identify a particular article. Example polynucleotide identifiers, such as DNA watermarks, are described in Heider and Barnekow. "DNA watermarks: A proof of concept" BMC Molecular Biology 9:40 (2008). The polynucleotide identifiers may also be a polynucleotide 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 polynucleotide. A polynucleotide 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 polynucleotide barcodes can be attached, or "tagged," to a target molecule and/or target polynucleotide. 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 polynucleotide). Target polynucleotides can be labeled with multiple polynucleotide barcodes in combinatorial fashion, such as a polynucleotide barcode concatemer. Typically, a polynucleotide barcode is used to identify target molecules and/or target polynucleotides 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 polynucleotide can be associated with multiple polynucleotide barcodes to provide information about all of these features (and more). Methods of generating polynucleotide-barcode are disclosed, for example, in International Patent Application Publication No. WO/2014/047561.

Homologous Recombination

[0410] In an embodiment, compositions induce a double strand break for the purpose of inducing HDR-mediated correction. In a further embodiment, two or more guide molecules complexing with the Type V Cas polypeptide or an ortholog or homolog thereof, may be used to induce multiplexed breaks for purpose of inducing HDR-mediated correction.

[0411] A "recombination template polynucleotide" (also referred to herein as a "donor polynucleotide" or "insert polynucleotide"), as that term is used herein, refers to a polynucleotide sequence which can be used in conjunction with compositions disclosed herein to alter the structure of a target position. In an embodiment, the target polynucleotide is modified to have some or all of the sequence of the recombination template polynucleotide,

typically at or near cleavage site(s). In an embodiment, the recombination template polynucleotide is single stranded. In an alternate embodiment, the recombination template polynucleotide is double stranded. In an embodiment, the recombination template polynucleotide is DNA, e.g., double stranded DNA. In an alternate embodiment, the recombination template polynucleotide is single stranded DNA.

[0412] 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 polynucleotide-targeting effector protein as a part of a polynucleotide-targeting complex.

[0413] A recombination template may be a component of another vector as disclosed 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.

[0414] In an embodiment, the recombination template polynucleotide alters the structure of the target position by participating in homologous recombination. In an embodiment, the recombination template polynucleotide alters the sequence of the target position. In an embodiment, the recombination template polynucleotide results in the incorporation of a modified, or non-naturally occurring base into the target polynucleotide.

[0415] The recombination template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the recombination template polynucleotide may include sequence that corresponds to a site on the target sequence that is cleaved by a Type V Cas polypeptide mediated cleavage event. In an embodiment, the recombination template polynucleotide may include sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Type V Cas polypeptide mediated event and

a second site on the target sequence that is cleaved in a second Type V Cas polypeptide mediated event.

[0416] In one embodiment, the recombination template polynucleotide 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 one embodiment, the recombination template polynucleotide 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.

[0417] A recombination template polynucleotide 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 polynucleotide 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.

[0418] The recombination template polynucleotide 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 polynucleotide may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 110+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 180+/- 10, 190+/- 10, 200+/- 10, 210+/-10, or 220+/- 10 nucleotides in length. In an embodiment, the recombination template polynucleotide may be 30+/-20, 40+/-20, 50+/-20, 60+/-20, 70+/- 20, 80+/-20, 90+/- 20, 100+/-20, 110+/-20, 120+/-20, 130+/-20, 140+/-20, 150+/-20, 160+/-20, 170+/-20, 180+/-

20, 190+/-20, 200+/-20, 210+/-20, or 220+/-20 nucleotides in length. In an embodiment, the recombination template polynucleotide 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.

[0419] A recombination template polynucleotide 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.

[0420] In one 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.

[0421] In one embodiment, recombination template polynucleotides 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.

[0422] Unlike Type V Cas polypeptide-mediated gene knockout, which permanently eliminates expression by mutating the gene at the DNA level, Type V Cas polypeptide knockdown allows for temporary reduction of gene expression through the use of artificial transcription factors. Mutating key residues in both DNA cleavage domains of the Type V Cas polypeptide, results in the generation of a catalytically inactive Type V Cas polypeptide. A catalytically inactive Type V Cas polypeptide complexes with a guide molecule and localizes to the DNA sequence specified by that guide molecule's targeting domain, however, it does not cleave the target DNA. Fusion of the inactive Type V Cas polypeptide protein to an effector

domain, e.g., a transcription repression domain, enables recruitment of the effector to any DNA site specified by the guide molecule. In one embodiment, Type V Cas 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 Type V Cas polypeptide can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene.

[0423] In an embodiment, a guide molecule can be targeted to known transcription response elements (e.g., promoters, enhancers, etc.), 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.

[0424] 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.

Non-Homologous End-Joining

[0425] In one embodiment, nuclease-induced non-homologous end-joining (NHEJ) can be used to target gene-specific knockouts. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein. The indel mutations generated by NHEJ are unpredictable in nature;

however, at a given break site certain indel sequences are favored and are overrepresented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; most commonly in the 1-50 bp range, but they can easily be greater than 50 bp, e.g., they can easily reach greater than about 100-200 bp. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

[0426] Because NHEJ is a mutagenic process, it may also be used to delete small sequence motifs as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

[0427] Both double strand cleaving Type V Cas polypeptide, or an ortholog or homolog thereof, and single strand, or nickase, Type V Cas polypeptide, or an ortholog or homolog thereof, molecules can be used in the methods and compositions disclosed herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

[0428] In an embodiment, in which a guide molecule and Type V Cas polypeptide, or an ortholog or homolog thereof, generate a double strand break for the purpose of inducing NHEJ-mediated indels, an RNA component molecule may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

[0429] In an embodiment, in which two guide molecules complexing with Type V Cas polypeptide, or an ortholog or homolog thereof, e.g., Type V Cas polypeptide nickases induce

two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide molecules may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

[0430] In some examples, the systems herein may introduce one or more indels via NHEJ pathway and insert sequence from a combination template via HDR.

Samples in General

[0431] It will be appreciated that in many applications of the compositions, formulations, and systems described herein utilize samples, such as biological samples, that can be obtained from a human or non-human animal subject, plant, prokaryote, and/or environment (e.g., air, soil, and water).

[0432] 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.

[0433] Any suitable sample collection method may be employed. Such techniques are generally known to the skilled artisan and/or are disclosed herein.

[0434] In some embodiments, samples are processed in one or more steps after initial collection. This processing can, for example, purify, clean, filter, isolate, one or more components in the sample for downstream analysis, long term storage, short term storage, and/or the like. Samples can be aliquoted in to replicates.

[0435] 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.

Exemplary Applications and Methods of Use

[0436] The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell *in vivo*, *ex vivo* or *in vitro* and, may be conducted in a

manner alters the cell such that once modified the progeny or cell line of the Type V Cas 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.

[0437] In one embodiment, one or more vectors disclosed herein are used to produce a non-human transgenic animal or transgenic plant. In one embodiment, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic animals and plants are known in the art, and generally begin with a method of cell transfection, such as disclosed herein.

Use of Orthogonal Catalytically Inactive Type V Cas Polypeptides

[0438] In particular embodiments, the Type V Cas polypeptide nickase is used in combination with an orthogonal catalytically inactive Type V Cas polypeptide to increase efficiency of said nickase. More particularly, the orthogonal catalytically inactive Type V Cas polypeptide is characterized by a different TAM recognition site than the Type V Cas nickase used in the AD-functionalized composition and the corresponding guide molecule sequence is selected to bind to a target sequence proximal to that of the nickase of the functionalized Type V Cas polypeptide. The orthogonal catalytically inactive Type V Cas polypeptide as used in the context of the present invention does not form part of the functionalized composition but merely functions to increase the efficiency of said nickase and is used in combination with a standard guide molecule as described in the art for said Type V Cas polypeptide. In particular embodiments, said orthogonal catalytically inactive Type V Cas polypeptide is a dead Type V Cas polypeptide, i.e., comprising one or more mutations which abolishes the nuclease activity of said Type V Cas polypeptide. In particular embodiments, the catalytically inactive orthogonal Type V Cas polypeptide is provided with two or more guide molecules which are capable of hybridizing to target sequences which are proximal to the target sequence of the nickase. In particular embodiments, at least two guide molecules are used to target said catalytically inactive Type V Cas polypeptide, of which at least one guide molecule is capable of hybridizing to a target sequence 5' of the target sequence of the nickase and at least one guide molecule is capable of hybridizing to a target sequence 3' of the target sequence of the nickase of the functionalized composition, whereby said one or more target sequences may be

on the same or the opposite DNA strand as the target sequence of the Type V Cas nickase. In particular embodiments, the guide sequences of the one or more guide molecules of the orthogonal catalytically inactive Type V Cas polypeptide are selected such that the target sequences are proximal to that of the guide molecule for the targeting of the functionalized composition, e.g. for the targeting of the nickase. In particular embodiments, the one or more target sequences of the orthogonal catalytically inactive Type V Cas polypeptide are each separated from the target sequence of the nickase by more than 5 but less than 450 base pairs. Optimal distances between the target sequences of the guide molecules for use with the orthogonal catalytically inactive Type V Cas polypeptide and the target sequence of the functionalized composition can be determined by the skilled person. In particular embodiments, the catalytically inactive orthogonal Type V Cas polypeptide has been modified to alter its TAM specificity as described elsewhere herein. In particular embodiments, the Type V Cas polypeptide nickase is a nickase which, by itself has limited activity in human cells, but which, in combination with an inactive orthogonal Type V Cas polypeptide and one or more corresponding proximal guide molecules ensures the required nickase activity.

Models of Diseases and Conditions

[0439] In an aspect, the invention provides a method of modeling a disease associated with a genomic locus in a eukaryotic organism or a non-human organism comprising manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus comprising delivering a non- naturally occurring or engineered composition comprising a viral vector system comprising one or more viral vectors operably encoding a composition for expression thereof, wherein the composition comprises particle delivery system or the delivery system or the virus particle of any one of the above embodiments or the cell of any one of the above embodiment.

[0440] In one aspect, the invention provides a method of generating a model eukaryotic cell that can include one or more a mutated disease genes and/or infectious microorganisms. In one embodiment, a disease gene is any gene associated an increase in the risk of having or developing a disease. In one embodiment, the method includes (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors comprise a composition, system, and/or component thereof and/or a vector or vector system that is capable of driving expression of a composition, system, and/or component thereof including, but not limited to: a guide molecule sequence, one or more Type V Cas polypeptides, and combinations thereof and (b)

allowing a composition, system, or complex to bind to one or more target polynucleotides, e.g., to effect cleavage, nicking, or other modification of the target polynucleotide within said disease gene, wherein the composition, system, or complex is composed of one or more Type V Cas polypeptide complexed with (1) one or more guide molecule sequences that is/are hybridized to the target sequence(s) within the target polynucleotide(s), and optionally (2) the guide molecule scaffold sequence(s), thereby generating a model eukaryotic cell comprising one or more mutated disease gene(s). Thus, in one embodiment the composition and system, contains nucleic acid molecules for and drives expression of one or more of: a Type V Cas polypeptide, a guide molecule sequence and/or a homologous recombination template and/or a stabilizing ligand if the Type V Cas polypeptide has a destabilization domain. In one embodiment, said cleavage comprises cleaving one or two strands at the location of the target sequence by the Type V Cas polypeptide. In one embodiment, nicking comprises nicking one or two strands at the location of the target sequence by the Type V Cas polypeptide. In one embodiment, said cleavage or nicking results in modified transcription of a target polynucleotide. In one embodiment, modification results in decreased transcription of the target polynucleotide. In one embodiment, the method further comprises repairing said cleaved or nicked target polynucleotide by homologous recombination with a recombination template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In one embodiment, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0441] The disease modeled can be any disease with a genetic or epigenetic component. In one embodiment, the disease modeled can be any as discussed elsewhere herein.

Models of Genetic and Epigenetic Conditions

[0442] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. As used herein, unless otherwise indicated, “disease” refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more polynucleotide sequences associated with a disease, or a plant, animal, or cell in which the expression of one or more polynucleotide sequences associated with a disease are altered. Such a polynucleotide sequence may encode a disease associated protein sequence or may be a

disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism, or cell can be a non-human subject, patient, organism, or cell. Thus, the invention provides a plant, animal, or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[0443] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0444] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0445] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a Type V Cas polypeptide, and a conserved nucleotide sequence linked to a guide/spacer sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0446] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the complex of the invention on a cellular function of interest. For example, a cellular function model may be used to study the effect of

a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0447] Several disease models have been specifically investigated. These include de novo autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred but serve to show the broad applicability of the invention across genes and corresponding models. An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0448] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, polynucleotide contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989) or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted polynucleotide sample is then detected by amplification procedures or conventional hybridization assays (e.g., Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0449] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E. coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[0450] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoechst, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0451] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Patent No. 5,210,015.

[0452] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe polynucleotide, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense polynucleotides. Conversely, where the nucleotide probe is a sense polynucleotide, the target polynucleotide is selected to be complementary to sequences of the sense polynucleotide.

[0453] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. (*See, e.g.,* Sambrook, et al.,

(1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Patent No. 5,445,934.

[0454] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical, or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, β -galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0455] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphorimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[0456] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent:protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[0457] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or

indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agent:polypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and, hence, generating a detectable signal.

[0458] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[0459] The amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[0460] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassay, ELISA (enzyme linked immunoradiometric assays), “sandwich” immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0461] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example,

anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Anti-phosphotyrosine antibodies are particularly useful in detecting proteins that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2 α). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

Modification of a Cell or Organism

[0462] The present disclosure further provides cells comprising one or more components of the systems herein, e.g., the Type V Cas polypeptide and/or guide molecule(s). Also provided include cells modified by the systems and methods herein, and cell cultures, tissues, organs, organism comprising such cells or progeny thereof. The invention comprehends a method of modifying a cell or organism. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The mammalian cell may be a non-human primate, bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish, or shrimp. The cell may also be a plant cell. The plant cell may be of a crop plant such as cassava, corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree, or vegetable. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol, or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

Therapeutic Uses and Methods of Treatment

[0463] The systems, compositions, and formulations disclosed herein can be used for diagnosing, prognosing, treating and/or preventing a disease, condition, disorder, or a symptom thereof. As such, also provided herein are methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject. Generally, the methods of diagnosing, prognosing, treating, and/or preventing a disease, state, or condition in or of a subject can include modifying a polynucleotide in a subject or cell thereof using a composition, system, or component thereof disclosed herein and/or include detecting a diseased or healthy polynucleotide in a subject or cell thereof using a composition, system, or component thereof

disclosed herein. In one embodiment, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism (e.g., bacterial or virus) within a subject or cell thereof. In one embodiment, the method of treatment or prevention can include using a composition, system, or component thereof to modify a polynucleotide of an infectious organism or symbiotic organism within a subject. The composition, system, and components thereof can be used to develop models of diseases, states, or conditions. The composition, system, and components thereof can be used to detect a disease state or correction thereof, such as by a method of treatment or prevention disclosed herein. The composition, system, and components thereof can be used to screen and select cells that can be used, for example, as treatments or preventions disclosed herein. The composition, system, and components thereof can be used to develop biologically active agents that can be used to modify one or more biologic functions or activities in a subject or a cell thereof.

[0464] In general, the method can include delivering a composition, system, and/or component thereof to a subject or cell thereof, or to an infectious or symbiotic organism by a suitable delivery technique and/or composition. Once administered the components can operate as described elsewhere herein to elicit a polynucleotide modification event. In some aspects, the polynucleotide modification event can occur at the genomic, epigenomic, and/or transcriptomic level. DNA and/or RNA cleavage, gene activation, and/or gene deactivation can occur. Additional features, uses, and advantages are described in greater detail below. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. In addition to treating and/or preventing a disease in a subject, the compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling *in vivo* (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[0465] The composition, system, and components thereof described elsewhere herein can be used to treat and/or prevent a disease, such as a genetic and/or epigenetic disease, in a subject. The composition, system, and components thereof described elsewhere herein can be

used to treat and/or prevent genetic infectious diseases in a subject, such as bacterial infections, viral infections, fungal infections, parasite infections, and combinations thereof. The composition, system, and components thereof described elsewhere herein can be used to modify the composition or profile of a microbiome in a subject, which can in turn modify the health status of the subject. The composition, system, disclosed herein can be used to modify cells *ex vivo*, which can then be administered to the subject whereby the modified cells can treat or prevent a disease or symptom thereof. This is also referred to in some contexts as adoptive therapy. The composition, system, disclosed herein can be used to treat mitochondrial diseases, where the mitochondrial disease etiology involves a mutation in the mitochondrial DNA.

[0466] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding one or more components of the composition, system, or complex or any of polynucleotides or vectors disclosed herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. The repair template may be a recombination template herein. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors disclosed herein, wherein said polynucleotide or vector encodes or comprises one or more components of composition, system, complex or component thereof comprising multiple Type V Cas polypeptides. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term ‘subject’ may be replaced by the phrase “cell or cell culture.”

[0467] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the Type V Cas polypeptide(s), advantageously encoding, and expressing *in vivo* the remaining portions of the composition, system, (e.g., RNA). A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression by transforming the subject with the Type V Cas polypeptide(s) advantageously encoding and expressing *in vivo* the remaining portions of the composition, system, (e.g., guide molecule); advantageously in one embodiment the Type V Cas polypeptide is a catalytically

inactive Type V Cas polypeptide and includes one or more associated functional domains. Where any treatment is occurring *ex vivo*, for example in a cell culture, then it will be appreciated that the term ‘subject’ may be replaced by the phrase “cell or cell culture.”

[0468] One or more components of the composition and system disclosed herein can be included in a composition, such as a pharmaceutical composition, and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or disclosed herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral guide molecule selection) and concentration of guide molecule (e.g., dependent on whether multiple guide molecules are used) may be advantageous for eliciting an improved effect.

[0469] Thus, also disclosed herein are methods of inducing one or more polynucleotide modifications in a eukaryotic or prokaryotic cell or component thereof (e.g., a mitochondria) of a subject, infectious organism, and/or organism of the microbiome of the subject. The modification can include the introduction, deletion, or substitution of one or more nucleotides at a target sequence of a polynucleotide of one or more cell(s). The modification can occur *in vitro*, *ex vivo*, *in situ*, or *in vivo*.

[0470] In one embodiment, the method of treating or inhibiting a condition or a disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non-human organism can include manipulation of a target sequence within a coding, non-coding or regulatory element of said genomic locus in a target sequence in a subject or a non-human subject in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition or disease is susceptible to treatment or inhibition by manipulation of the target sequence including providing treatment comprising delivering a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment.

[0471] Also provided herein is the use of the particle delivery system or the delivery system or the virus particle of any one of the above embodiments or the cell of any one of the above embodiments in *ex vivo* or *in vivo* gene or genome editing; or for use in *in vitro*, *ex vivo* or *in vivo* gene therapy. Also provided herein are particle delivery systems, non-viral delivery

systems, and/or the virus particle of any one of the above embodiments or the cell of any one of the above embodiments used in the manufacture of a medicament for *in vitro*, *ex vivo* or *in vivo* gene or genome editing or for use in *in vitro*, *ex vivo* or *in vivo* gene therapy or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus associated with a disease or in a method of treating or inhibiting a condition or disease caused by one or more mutations in a genomic locus in a eukaryotic organism or a non- human organism.

[0472] In one embodiment, polynucleotide modification can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said polynucleotide of said cell(s). The modification can include the introduction, deletion, or substitution of at least 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence. The modification can include the introduction, deletion, or substitution of at least 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s). The modification can include the introduction, deletion, or substitution of at least 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2700, 2800, 2900, 3000, 3100, 3200, 3300, 3400, 3500, 3600, 3700, 3800, 3900, 4000, 4100, 4200, 4300, 4400, 4500, 4600, 4700, 4800, 4900, 5000, 5100, 5200, 5300, 5400, 5500, 5600, 5700, 5800, 5900, 6000, 6100, 6200, 6300, 6400, 6500, 6600, 6700, 6800, 6900, 7000, 7100, 7200, 7300, 7400, 7500, 7600, 7700, 7800, 7900, 8000, 8100, 8200, 8300, 8400, 8500, 8600, 8700, 8800, 8900, 9000, 9100, 9200, 9300, 9400, 9500, 9600, 9700, 9800, or 9900 to 10000 nucleotides at each target sequence of said cell(s).

[0473] In one embodiment, the modifications can include the introduction, deletion, or substitution of nucleotides at each target sequence of said cell(s) via guide molecules (e.g., guide RNA(s)), such as those mediated by a composition, system, or a component thereof

described elsewhere herein. In one embodiment, the modifications can include the introduction, deletion, or substitution of nucleotides at a target or random sequence of said cell(s) via a composition, system, or technique.

[0474] In one embodiment, the composition, system, or component thereof can promote Non-Homologous End-Joining (NHEJ). In one embodiment, modification of a polynucleotide by a composition, system, or a component thereof, such as a diseased polynucleotide, can include NHEJ. In one embodiment, promotion of this repair pathway by the composition, system, or a component thereof can be used to target gene or polynucleotide specific knock-outs and/or knock-ins. In one embodiment, promotion of this repair pathway by the composition, system, or a component thereof can be used to generate NHEJ-mediated indels. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. The indel can range in size from 1-50 or more base pairs. In one embodiment the indel can be 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, 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, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288,

289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, or 500 base pairs or more. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences.

[0475] In one embodiment, composition, system, mediated NHEJ can be used in the method to delete small sequence motifs. In one embodiment, composition, system, mediated NHEJ can be used in the method to generate NHEJ-mediate indels that can be targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp). In an embodiment, in which a guide molecule and Type V Cas polypeptide generate a double strand break for the purpose of inducing NHEJ-mediated indels, a guide molecule may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position). In an embodiment, in which two component RNAs complexing with one or more nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two

component RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

[0476] For minimization of toxicity and off-target effect, it may be important to control the concentration of Type V Cas polypeptide mRNA and component RNA delivered. Optimal concentrations of Type V Cas polypeptide mRNA and component RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, nickase mRNA (for example a mutated Type V Cas) can be delivered with a pair of guide molecules targeting a site of interest.

[0477] Typically, in the context of an endogenous Type V Cas polypeptide, formation of a Type V Cas polypeptide or complex (comprising a polynucleotide component sequence hybridized to a target sequence and complexed with one or more Type V Cas polypeptides) results in cleavage, nicking, and/or another modification of one or both 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.

[0478] In one embodiment, a method of modifying a target polynucleotide in a cell to treat or prevent a disease can include allowing a composition, system, or component thereof to bind to the target polynucleotide, e.g., to effect cleavage, nicking, or other modification as the composition, system, is capable of said target polynucleotide, thereby modifying the target polynucleotide, wherein the composition, system, or component thereof, complex with a guide molecule sequence, and hybridize said guide molecule sequence to a target sequence within the target polynucleotide, wherein said guide molecule sequence is optionally linked to a guide molecule scaffold sequence. In some of these embodiments, the composition, system, or component thereof can be or include a Type V Cas polypeptide complexed with a guide molecule sequence. In one embodiment, modification can include cleaving or nicking one or two strands at the location of the target sequence by one or more components of the composition, system, or component thereof.

[0479] The cleavage, nicking, or other modification capable of being performed by the composition, system, can modify transcription of a target polynucleotide. In one embodiment, modification of transcription can include decreasing transcription of a target polynucleotide. In one embodiment, modification can include increasing transcription of a target polynucleotide. In one embodiment, the method includes repairing said cleaved target polynucleotide by

homologous recombination with a recombination template polynucleotide, wherein said repair results in a modification such as, but not limited to, an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In one embodiment, said modification results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In one embodiment, the modification imparted by the composition, system, or component thereof provides a transcript and/or protein that can correct a disease or a symptom thereof, including but not limited to, any of those described in greater detail elsewhere herein.

[0480] In one embodiment, the method of treating or preventing a disease can include delivering one or more vectors or vector systems to a cell, such as a eukaryotic or prokaryotic cell, wherein one or more vectors or vector systems include the composition, system, or component thereof. In one embodiment, the vector(s) or vector system(s) can be a viral vector or vector system, such as an AAV or lentiviral vector system, which are described in greater detail elsewhere herein. In one embodiment, the method of treating or preventing a disease can include delivering one or more viral particles, such as an AAV or lentiviral particle, containing the composition, system, or component thereof. In one embodiment, the viral particle has a tissue specific tropism. In one embodiment, the viral particle has a liver, muscle, eye, heart, pancreas, kidney, neuron, epithelial cell, endothelial cell, astrocyte, glial cell, immune cell, or red blood cell specific tropism.

[0481] It will be understood that the composition and system, according to the invention as disclosed herein, such as the composition and system, for use in the methods according to the invention as disclosed herein, may be suitably used for any type of application known for composition, system, preferably in eukaryotes. In certain aspects, the application is therapeutic, preferably therapeutic in a eukaryote organism, such as including but not limited to animals (including human), plants, algae, fungi (including yeasts), etc. Alternatively, or in addition, in certain aspects, the application may involve accomplishing or inducing one or more particular traits or characteristics, such as genotypic and/or phenotypic traits or characteristics, as also described elsewhere herein.

Exemplary Disease Treatments

[0482] The following discussion provides non-limiting exemplary uses of the systems and compositions of the present invention in the context of non-limiting exemplary diseases.

Treating Diseases of the Circulatory System

[0483] In one embodiment, the composition, system, and/or component thereof disclosed herein can be used to treat and/or prevent a circulatory system disease. Exemplary disease is provided, for example, in **Tables 5A and 5B**. In one embodiment the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) can be used to deliver the composition, system, and/or component thereof disclosed herein to the blood. In one embodiment, the circulatory system disease can be treated by using a lentivirus to deliver the composition, system, disclosed herein to modify hematopoietic stem cells (HSCs) *in vivo* or *ex vivo* (*See, e.g.*, Drakopoulou, “Review Article, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β -Thalassemia,” Stem Cells International, Volume 2011, Article ID 987980, 10 pages, doi:10.4061/2011/987980, which can be adapted for use with the composition, system, herein in view of the description herein). In one embodiment, the circulatory system disorder can be treated by correcting HSCs as to the disease using a composition, system, herein or a component thereof, wherein the composition, system, optionally includes a suitable HDR repair template (*See, e.g.*, Cavazzana, “Outcomes of Gene Therapy for β -Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced *Ex Vivo* with a Lentiviral β A-T87Q-Globin Vector.”; Cavazzana-Calvo, “Transfusion independence and HMGA2 activation after gene therapy of human β -thalassaemia”, Nature 467, 318–322 (16 September 2010) doi:10.1038/nature09328; Nienhuis, “Development of Gene Therapy for Thalassemia, Cold Spring Harbor Perspectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, a lentiviral vector containing an engineered β -globin gene (β A-T87Q); and Xie et al., “Seamless gene correction of β -thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback” Genome Research gr.173427.114 (2014) www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press; Watts, “Hematopoietic Stem Cell Expansion and Gene Therapy” Cytotherapy 13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011), which can be adapted for use with the composition, system, herein in view of the description herein). In one embodiment, iPSCs can be modified using a composition, system, disclosed herein to correct a disease polynucleotide associated with a circulatory disease. In this regard, the teachings of Xu et al. (Sci Rep. 2015 Jul 9;5:12065. doi: 10.1038/srep12065) and Song et al. (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5)

with respect to modifying iPSCs can be adapted for use in view of the description herein with the composition, system, disclosed herein.

[0484] The term “Hematopoietic Stem Cell” or “HSC” refers broadly those cells considered to be an HSC, e.g., blood cells that give rise to all the other blood cells and are derived from mesoderm; located in the red bone marrow, which is contained in the core of most bones. HSCs of the invention include cells having a phenotype of hematopoietic stem cells, identified by small size, lack of lineage (lin) markers, and markers that belong to the cluster of differentiation series, like: CD34, CD38, CD90, CD133, CD105, CD45, and also c-kit, - the receptor for stem cell factor. Hematopoietic stem cells are negative for the markers that are used for detection of lineage commitment, and are, thus, called Lin-; and, during their purification by FACS, a number of up to 14 different mature blood-lineage markers, e.g., CD13 & CD33 for myeloid, CD71 for erythroid, CD19 for B cells, CD61 for megakaryocytic, etc. for humans; and, B220 (murine CD45) for B cells, Mac-1 (CD11b/CD18) for monocytes, Gr-1 for Granulocytes, Ter119 for erythroid cells, Il7Ra, CD3, CD4, CD5, CD8 for T cells, etc. Mouse HSC markers: CD34^{lo/-}, SCA-1⁺, Thy1.1^{+/lo}, CD38⁺, C-kit⁺, lin⁻, and Human HSC markers: CD34⁺, CD59⁺, Thy1/CD90⁺, CD38^{lo/-}, C-kit/CD117⁺, and lin⁻. HSCs are identified by markers. Hence in embodiments discussed herein, the HSCs can be CD34⁺ cells. HSCs can also be hematopoietic stem cells that are CD34⁻/CD38⁻. Stem cells that may lack c-kit on the cell surface that are considered in the art as HSCs are within the ambit of the invention, as well as CD133⁺ cells likewise considered HSCs in the art.

[0485] In one embodiment, the treatment or prevention for treating a circulatory system or blood disease can include modifying a human cord blood cell with any modification disclosed herein. In one embodiment, the treatment or prevention for treating a circulatory system or blood disease can include modifying a granulocyte colony-stimulating factor-mobilized peripheral blood cell (mPB) with any modification disclosed herein. In one embodiment, the human cord blood cell or mPB can be CD34⁺. In one embodiment, the cord blood cell(s) or mPB cell(s) modified can be autologous. In one embodiment, the cord blood cell(s) or mPB cell(s) can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, disclosed herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g., Cartier, “MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-

Linked Adrenoleukodystrophy,” Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein. The modified cord blood cell(s) or mPB cell(s) can be optionally expanded *in vitro*. The modified cord blood cell(s) or mPB cell(s) can be derived to a subject in need thereof using any suitable delivery technique.

[0486] The compositions may be engineered to target genetic locus or loci in HSCs. In one embodiment, the Type V Cas polypeptide(s) can be codon-optimized for a eukaryotic cell and especially a mammalian cell, e.g., a human cell, for instance, HSC, or iPSC and guide molecule targeting a locus or loci in HSC, such as circulatory disease, can be prepared. These may be delivered via particles. The particles may be formed by the Type V Cas polypeptide and the guide molecule being admixed. The guide molecule and Type V Cas polypeptide mixture can be, for example, admixed with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein, and alcohol, whereby particles containing the guide molecule and Type V Cas polypeptide may be formed. The invention comprehends so making particles and particles from such a method as well as uses thereof. Particles suitable delivery of the composition in the context of blood or circulatory system or HSC delivery to the blood or circulatory system are described in greater detail elsewhere herein.

[0487] In one embodiment, after *ex vivo* modification the HSCs or iPCS can be expanded prior to administration to the subject. Expansion of HSCs can be via any suitable method such as that described by, Lee, “Improved *ex vivo* expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4.” Blood. 2013 May 16;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 2013 Mar 21.

[0488] In one embodiment, the HSCs or iPSCs modified can be autologous. In one embodiment, the HSCs or iPSCs can be allogenic. In addition to the modification of the disease gene(s), allogenic cells can be further modified using the composition, system, disclosed herein to reduce the immunogenicity of the cells when delivered to the recipient. Such techniques are described elsewhere herein and e.g., Cartier, “MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-Linked Adrenoleukodystrophy,” Brain Pathology 20 (2010) 857–862, which can be adapted for use with the composition, system, herein.

Treating Neurological Diseases

[0489] In one embodiment, the compositions, systems, disclosed herein can be used to treat diseases of the brain and CNS. Delivery options for the brain include encapsulation of Type V Cas polypeptide and guide molecule in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing Type V Cas polypeptide and guide molecule. For instance, Xia CF and Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." *Mol Pharm.* 2009 May-Jun;6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and *in vivo*, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed *in vivo* following an intravenous administration of the targeted siRNA, the teachings of which can be adapted for use with the compositions, systems, herein. In other embodiments, an artificial virus can be generated for CNS and/or brain delivery. *See, e.g.,* Zhang et al. (*Mol Ther.* 2003 Jan;7(1):11-8.)), the teachings of which can be adapted for use with the compositions, systems, herein.

Treating Hearing Diseases

[0490] In one embodiment the composition and system disclosed herein can be used to treat a hearing disease or hearing loss in one or both ears. Deafness is often caused by lost or damaged hair cells that cannot relay signals to auditory neurons. In such cases, cochlear implants may be used to respond to sound and transmit electrical signals to the nerve cells. But these neurons often degenerate and retract from the cochlea as fewer growth factors are released by impaired hair cells.

[0491] In one embodiment, the composition, system, or modified cells can be delivered to one or both ears for treating or preventing hearing disease or loss by any suitable method or technique. Suitable methods and techniques include, but are not limited to, those set forth in US Patent Publication No. 20120328580 describes injection of a pharmaceutical composition into the ear (e.g., auricular administration), such as into the luminae of the cochlea (e.g., the Scala media, Sc vestibulae, and Sc tympani), e.g., using a syringe, e.g., a single-dose syringe.

For example, one or more of the compounds disclosed herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner ear; administration in situ, via a catheter or pump (*See, e.g.*, McKenna et al., (U.S. Patent Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639); administration in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear (*See, e.g.*, U.S. Patent Publication No. 2007/0093878, which provides an exemplary cochlear implant suitable for delivery of the compositions, systems, disclosed herein to the ear). Such methods are routinely used in the art, for example, for the administration of steroids and antibiotics into human ears. Injection can be, for example, through the round window of the ear or through the cochlear capsule. Other inner ear administration methods are known in the art (*See, e.g.*, Salt and Plontke, *Drug Discovery Today*, 10:1299-1306, 2005). In one embodiment, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient during a surgical procedure. In one embodiment, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient without the need for a surgical procedure.

[0492] In general, the cell therapy methods described in US Patent Publication No. 20120328580 can be used to promote complete or partial differentiation of a cell to or towards a mature cell type of the inner ear (e.g., a hair cell) *in vitro*. Cells resulting from such methods can then be transplanted or implanted into a patient in need of such treatment. The cell culture methods required to practice these methods, including methods for identifying and selecting suitable cell types, methods for promoting complete or partial differentiation of selected cells, methods for identifying complete or partially differentiated cell types, and methods for implanting complete or partially differentiated cells are described below.

[0493] Cells suitable for use in the present invention include, but are not limited to, cells that are capable of differentiating completely or partially into a mature cell of the inner ear, e.g., a hair cell (e.g., an inner and/or outer hair cell), when contacted, e.g., *in vitro*, with one or more of the compounds disclosed herein. Exemplary cells that are capable of differentiating into a hair cell include, but are not limited to stem cells (e.g., inner ear stem cells, adult stem cells, bone marrow derived stem cells, embryonic stem cells, mesenchymal stem cells, skin stem cells, iPS cells, and fat derived stem cells), progenitor cells (e.g., inner ear progenitor cells), support cells (e.g., Deiters' cells, pillar cells, inner phalangeal cells, tectal cells and Hensen's cells), and/or germ cells. The use of stem cells for the replacement of inner ear sensory

cells is described in Li et al., (U.S. Patent Publication No. 2005/0287127) and Li et al., (U.S. Patent Application No. 11/953,797). The use of bone marrow derived stem cells for the replacement of inner ear sensory cells is described in Edge et al., PCT/US2007/084654. iPS cells are described, e.g., at Takahashi et al., *Cell*, Volume 131, Issue 5, Pages 861-872 (2007); Takahashi and Yamanaka, *Cell* 126, 663-76 (2006); Okita et al., *Nature* 448, 260-262 (2007); Yu, J. et al., *Science* 318(5858):1917-1920 (2007); Nakagawa et al., *Nat. Biotechnol.* 26:101-106 (2008); and Zaehres and Scholer, *Cell* 131(5):834-835 (2007). Such suitable cells can be identified by analyzing (e.g., qualitatively or quantitatively) the presence of one or more tissue specific genes. For example, gene expression can be detected by detecting the protein product of one or more tissue-specific genes. Protein detection techniques involve staining proteins (e.g., using cell extracts or whole cells) using antibodies against the appropriate antigen. In this case, the appropriate antigen is the protein product of the tissue-specific gene expression. Although, in principle, a first antibody (i.e., the antibody that binds the antigen) can be labeled, it is more common (and improves the visualization) to use a second antibody directed against the first (e.g., an anti-IgG). This second antibody is conjugated either with fluorochromes, or appropriate enzymes for colorimetric reactions, or gold beads (for electron microscopy), or with the biotin-avidin system, so that the location of the primary antibody, and thus the antigen, can be recognized.

[0494] The composition and system may be delivered to the ear by direct application of pharmaceutical composition to the outer ear, with compositions modified from US Patent Publication No. 20110142917. In one embodiment the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[0495] In one embodiment, the compositions, systems, or components thereof and/or vectors or vector systems can be delivered to ear via a transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the polynucleotide-targeting system of the present invention (*See, e.g.,* Qi et al., *Gene Therapy* (2013), 1-9). About 40 μ l of 10mM RNA may be contemplated as the dosage for administration to the ear.

[0496] According to Rejali et al. (*Hear Res.* 2007 Jun;228(1-2):180-7), cochlear implant function can be improved by good preservation of the spiral ganglion neurons, which are the target of electrical stimulation by the implant and brain derived neurotrophic factor (BDNF) has previously been shown to enhance spiral ganglion survival in experimentally deafened ears.

Rejali et al. tested a modified design of the cochlear implant electrode that includes a coating of fibroblast cells transduced by a viral vector with a BDNF gene insert. To accomplish this type of *ex vivo* gene transfer, Rejali et al. transduced guinea pig fibroblasts with an adenovirus with a BDNF gene cassette insert and determined that these cells secreted BDNF and then attached BDNF-secreting cells to the cochlear implant electrode via an agarose gel and implanted the electrode in the scala tympani. Rejali et al. determined that the BDNF expressing electrodes were able to preserve significantly more spiral ganglion neurons in the basal turns of the cochlea after 48 days of implantation when compared to control electrodes and demonstrated the feasibility of combining cochlear implant therapy with *ex vivo* gene transfer for enhancing spiral ganglion neuron survival. Such a system may be applied to the polynucleotide-targeting system of the present invention for delivery to the ear.

[0497] In one embodiment, the system set forth in Mukherjee et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) can be adapted for transtympanic administration of the composition, system, or component thereof to the ear. In one embodiment, a dosage of about 2 mg to about 4 mg of Type V Cas polypeptide for administration to a human.

[0498] In one embodiment, the system set forth in [Jung et al. (Molecular Therapy, vol. 21 no. 4, 834–841 Apr. 2013) can be adapted for vestibular epithelial delivery of the composition, system, or component thereof to the ear. In one embodiment, a dosage of about 1 to about 30 mg of Type V Cas polypeptide for administration to a human.

Treating Diseases in Non-Dividing Cells

[0499] In one embodiment, the gene or transcript to be corrected is in a non-dividing cell. Exemplary non-dividing cells are muscle cells or neurons. Non-dividing (especially non-dividing, fully differentiated) cell types present issues for gene targeting or genome engineering, for example because homologous recombination (HR) is generally suppressed in the G1 cell-cycle phase. However, while studying the mechanisms by which cells control normal DNA repair systems, Durocher discovered a previously unknown switch that keeps HR “off” in non-dividing cells and devised a strategy to toggle this switch back on. Orthwein et al. (Daniel Durocher’s lab at the Mount Sinai Hospital in Ottawa, Canada) recently reported (Nature 16142, published online 9 Dec 2015) have shown that the suppression of HR can be lifted and gene targeting successfully concluded in both kidney (293T) and osteosarcoma (U2OS) cells. Tumor suppressors, BRCA1, PALB2 and BRAC2 are known to promote DNA DSB repair by HR. They found that formation of a complex of BRCA1 with PALB2 - BRAC2

is governed by a ubiquitin site on PALB2, such that action on the site by an E3 ubiquitin ligase. This E3 ubiquitin ligase is composed of KEAP1 (a PALB2 -interacting protein) in complex with cullin-3 (CUL3)–RBX1. PALB2 ubiquitylation suppresses its interaction with BRCA1 and is counteracted by the deubiquitylase USP11, which is itself under cell cycle control. Restoration of the BRCA1–PALB2 interaction combined with the activation of DNA-end resection is sufficient to induce homologous recombination in G1, as measured by a number of methods including a Cas polypeptide nuclease-based gene-targeting assay directed at USP11 or KEAP1 (expressed from a pX459 vector). However, when the BRCA1–PALB2 interaction was restored in resection-competent G1 cells using either KEAP1 depletion or expression of the PALB2-KR mutant, a robust increase in gene-targeting events was detected. These teachings can be adapted for and/or applied to the compositions, systems, disclosed herein.

[0500] Thus, reactivation of HR in cells, especially non-dividing, fully differentiated cell types is preferred. In one embodiment, promotion of the BRCA1–PALB2 interaction is preferred. In one embodiment, the target cell is a non-dividing cell. In one embodiment, the target cell is a neuron or muscle cell. In one embodiment, the target cell is targeted *in vivo*. In one embodiment, the cell is in G1 and HR is suppressed. In one embodiment, use of KEAP1 depletion, for example inhibition of expression of KEAP1 activity, is preferred. KEAP1 depletion may be achieved through siRNA, for example as shown in Orthwein et al. Alternatively, expression of the PALB2-KR mutant (lacking all eight Lys residues in the BRCA1-interaction domain) is preferred, either in combination with KEAP1 depletion or alone. PALB2-KR interacts with BRCA1 irrespective of cell cycle position. Thus, promotion or restoration of the BRCA1-PALB2 interaction, especially in G1 cells, is preferred. In one embodiment, especially where the target cells are non-dividing, or where removal and return (*ex vivo* gene targeting) is problematic, for example neuron or muscle cells. KEAP1 siRNA is available from ThermoFischer. In one embodiment, a BRCA1–PALB2 complex may be delivered to the G1 cell. In one embodiment, PALB2 deubiquitylation may be promoted for example by increased expression of the deubiquitylase USP11, so it is envisaged that a construct may be provided to promote or up-regulate expression or activity of the deubiquitylase USP11.

Treating Diseases of the Eye

[0501] In one embodiment, the disease to be treated is a disease that affects the eyes. Thus, In one embodiment, the composition, system, or component thereof disclosed herein is delivered to one or both eyes.

[0502] The composition, system, can be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[0503] In one embodiment, the condition to be treated or targeted is an eye disorder. In one embodiment, the eye disorder may include glaucoma. In one embodiment, the eye disorder includes a retinal degenerative disease. In one embodiment, the retinal degenerative disease is selected from Stargardt disease, Bardet-Biedl Syndrome, Best disease, Blue Cone Monochromacy, Choroideremia, Cone-rod dystrophy, Congenital Stationary Night Blindness, Enhanced S-Cone Syndrome, Juvenile X-Linked Retinoschisis, Leber Congenital Amaurosis, Malattia Leventinese, Norrie Disease or X-linked Familial Exudative Vitreoretinopathy, Pattern Dystrophy, Sorsby Dystrophy, Usher Syndrome, Retinitis Pigmentosa, Achromatopsia or Macular dystrophies or degeneration, Retinitis Pigmentosa, Achromatopsia, and age related macular degeneration. In one embodiment, the retinal degenerative disease is Leber Congenital Amaurosis (LCA) or Retinitis Pigmentosa. Other exemplary eye diseases are described in greater detail elsewhere herein.

[0504] In one embodiment, the composition, system, is delivered to the eye, optionally via intravitreal injection or subretinal injection. Intraocular injections may be performed with the aid of an operating microscope. For subretinal and intravitreal injections, eyes may be prolapsed by gentle digital pressure and fundi visualized using a contact lens system consisting of a drop of a coupling medium solution on the cornea covered with a glass microscope slide coverslip. For subretinal injections, the tip of a 10-mm 34-gauge needle, mounted on a 5- μ l Hamilton syringe may be advanced under direct visualization through the superior equatorial sclera tangentially towards the posterior pole until the aperture of the needle was visible in the subretinal space. Then, 2 μ l of vector suspension may be injected to produce a superior bullous retinal detachment, thus confirming subretinal vector administration. This approach creates a self-sealing sclerotomy allowing the vector suspension to be retained in the subretinal space until it is absorbed by the RPE, usually within 48 h of the procedure. This procedure may be repeated in the inferior hemisphere to produce an inferior retinal detachment. This technique

results in the exposure of approximately 70% of neurosensory retina and RPE to the vector suspension. For intravitreal injections, the needle tip may be advanced through the sclera 1 mm posterior to the corneoscleral limbus and 2 µl of vector suspension injected into the vitreous cavity. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 µl of vector suspension may be injected. These vectors may be injected at titers of either $1.0\text{--}1.4 \times 10^{10}$ or $1.0\text{--}1.4 \times 10^9$ transducing units (TU)/ml.

[0505] In one embodiment, for administration to the eye, lentiviral vectors. In one embodiment, the lentiviral vector is an equine infectious anemia virus (EIAV) vector. Exemplary EIAV vectors for eye delivery are described in Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845; Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012), which can be adapted for use with the composition, system, disclosed herein. In one embodiment, the dosage can be 1.1×10^5 transducing units per eye (TU/eye) in a total volume of 100 µl.

[0506] Other viral vectors can also be used for delivery to the eye, such as AAV vectors, such as those described in Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006), Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642–649 Apr. 2011; Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)), which can be adapted for use with the composition, system, disclosed herein. In one embodiment, the dose can range from about 10^6 to $10^{9.5}$ particle units. In the context of the Millington-Ward AAV vectors, a dose of about 2×10^{11} to about 6×10^{13} virus particles can be administered. In the context of Dalkara vectors, a dose of about 1×10^{15} to about 1×10^{16} vg/ml administered to a human.

[0507] In one embodiment, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering composition, system, to the eye. In this system, a single intravitreal administration of 3 µg of sd-rxRNA results in sequence-specific reduction of PPIB mRNA levels for 14 days. The sd-rxRNA® system may be applied to the polynucleotide-targeting system of the present invention, contemplating a dose of about 3 to 20 mg of composition administered to a human.

[0508] In other embodiments, the methods of US Patent Publication No. 20130183282, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the polynucleotide-targeting system of the present invention.

[0509] In other embodiments, the methods of US Patent Publication No. 20130202678 for treating retinopathies and sight-threatening ophthalmologic disorders relating to delivering of the Puf-A gene (which is expressed in retinal ganglion and pigmented cells of eye tissues and displays a unique anti-apoptotic activity) to the sub-retinal or intravitreal space in the eye may be used or adapted. In particular, desirable targets are *zgc:193933*, *prdm1a*, *spata2*, *tex10*, *rbb4*, *ddx3*, *zp2.2*, *Blimp-1* and *HtrA2*, all of which may be targeted by the composition, system, of the present invention.

[0510] Wu (Cell Stem Cell,13:659–62, 2013) designed a guide RNA that led Cas9 to a single base pair mutation that causes cataracts in mice, where it induced DNA cleavage. Then using either the other wild-type allele or oligos given to the zygotes repair mechanisms corrected the sequence of the broken allele and corrected the cataract-causing genetic defect in mutant mouse. This approach can be adapted to and/or applied to the Type V Cas compositions, systems, disclosed herein.

[0511] US Patent Publication No. 20120159653, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeneration (MD), the teachings of which can be applied to and/or adapted for the Type V Cas compositions, systems, disclosed herein.

[0512] One aspect of US Patent Publication No. 20120159653 relates to editing of any chromosomal sequences that encode proteins associated with MD which may be applied to the polynucleotide-targeting system of the present invention.

Treating Muscle Diseases and Cardiovascular Diseases

[0513] In one embodiment, the composition, system can be used to treat and/or prevent a muscle disease and associated circulatory or cardiovascular disease or disorder. The present invention also contemplates delivering the composition, system, disclosed herein, e.g., Type V Cas effector protein systems, to the heart. For the heart, a myocardium tropic adeno-associated virus (AAVM) is preferred, in particular AAVM41 which showed preferential gene transfer in the heart (See, e.g., Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). Administration may be systemic or local. A dosage of about $1\text{-}10 \times 10^{14}$ vector genomes is contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492: 376 and

Somasuntharam et al. (2013) *Biomaterials* 34: 7790, the teachings of which can be adapted for and/or applied to the compositions, systems, disclosed herein.

[0514] For example, US Patent Publication No. 20110023139, the teachings of which can be adapted for and/or applied to the compositions, systems, disclosed herein describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with cardiovascular disease. Cardiovascular diseases generally include high blood pressure, heart attacks, heart failure, and stroke and TIA. Any chromosomal sequence involved in cardiovascular disease, or the protein encoded by any chromosomal sequence involved in cardiovascular disease may be utilized in the methods described in this disclosure. The cardiovascular-related proteins are typically selected based on an experimental association of the cardiovascular-related protein to the development of cardiovascular disease. For example, the production rate or circulating concentration of a cardiovascular-related protein may be elevated or depressed in a population having a cardiovascular disorder relative to a population lacking the cardiovascular disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the cardiovascular-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[0515] The compositions, systems, herein can be used for treating diseases of the muscular system. The present invention also contemplates delivering the composition, system, disclosed herein, effector protein systems, to muscle(s).

[0516] In one embodiment, the muscle disease to be treated is a muscle dystrophy such as DMD. In one embodiment, the composition, system, such as a system capable of RNA modification, disclosed herein can be used to achieve exon skipping to achieve correction of the diseased gene. As used herein, unless otherwise indicated, the term “exon skipping” refers to the modification of pre-mRNA splicing by the targeting of splice donor and/or acceptor sites within a pre-mRNA with one or more complementary antisense oligonucleotide(s) (AONs). By blocking access of a spliceosome to one or more splice donor or acceptor site, an AON may prevent a splicing reaction thereby causing the deletion of one or more exons from a fully-processed mRNA. Exon skipping may be achieved in the nucleus during the maturation process

of pre-mRNAs. In some examples, exon skipping may include the masking of key sequences involved in the splicing of targeted exons by using a composition, system, disclosed herein capable of RNA modification. In one embodiment, exon skipping can be achieved in dystrophin mRNA. In one embodiment, the composition, system, can induce exon skipping at exon 1, 2, 3, 4, 5, 6, 7, 8, 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, 45, 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, or any combination thereof of the dystrophin mRNA. In one embodiment, the composition, system, can induce exon skipping at exon 43, 44, 50, 51, 52, 55, or any combination thereof of the dystrophin mRNA. Mutations in these exons, can also be corrected using non-exon skipping polynucleotide modification methods.

[0517] In one embodiment, for treatment of a muscle disease, the method of Bortolanza et al. (Molecular Therapy vol. 19 no. 11, 2055–2064 Nov. 2011) may be applied to an AAV expressing Type V Cas polypeptide and injected into humans at a dosage of about 2×10^{15} or 2×10^{16} vg of vector. The teachings of Bortolanza et al., can be adapted for and/or applied to the compositions, systems, disclosed herein.

[0518] In one embodiment, the method of Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881–887 May 2010) may be applied to an AAV expressing Type V Cas polypeptide and injected into humans, for example, at a dosage of about 10^{14} to about 10^{15} vg of vector. The teachings of Dumonceaux disclosed herein can be adapted for and/or applied to the compositions, systems, disclosed herein.

[0519] In one embodiment, the method of Kinouchi et al. (Gene Therapy (2008) 15, 1126–1130) may be applied to compositions disclosed herein and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 μ M solution into the muscle.

[0520] In one embodiment, the method of Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) can be adapted for and/or applied to the compositions, systems, herein and injected at a dose of about 15 to about 50 mg into the great saphenous vein of a human.

[0521] In one embodiment, the method comprises treating a sickle cell related disease, e.g., sickle cell trait, sickle cell disease such as sickle cell anemia, β -thalassaemia. For example, the method and system may be used to modify the genome of the sickle cell, e.g., by correcting one or more mutations of the β -globin gene. In the case of β -thalassaemia, sickle cell anemia can be corrected by modifying HSCs with the systems. The system allows the specific editing

of the cell's genome by cutting its DNA and then letting it repair itself. The Type V Cas polypeptide is inserted and directed by a guide molecule to the mutated point and then it cuts the DNA at that point. Simultaneously, a healthy version of the sequence is inserted. This sequence is used by the cell's own repair system to fix the induced cut. In this way, the Type V Cas polypeptide allows the correction of the mutation in the previously obtained stem cells. The methods and systems may be used to correct HSCs as to sickle cell anemia using a system that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β -globin, advantageously non-sickling β -globin); specifically, the guide molecule can target mutation that give rise to sickle cell anemia, and the HDR can provide coding for proper expression of β -globin. A guide molecule that targets the mutation-and- Type V Cas polypeptide containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of β -globin; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. The HDR template can provide for the HSC to express an engineered β -globin gene (e.g., β A-T87Q), or β -globin.

Treating Diseases of the Liver and Kidney

[0522] In one embodiment, the composition, system, or component thereof disclosed herein can be used to treat a disease of the kidney or liver. Thus, in one embodiment, delivery of the composition or component thereof disclosed herein is to the liver or kidney.

[0523] Delivery strategies to induce cellular uptake of the therapeutic nucleic acid molecule include physical force or vector systems such as viral-, lipid- or complex-based delivery, or nanocarriers. From the initial applications with less possible clinical relevance, when nucleic acid molecules were addressed to renal cells with hydrodynamic high-pressure injection systemically, a wide range of gene therapeutic viral and non-viral carriers have been applied already to target posttranscriptional events in different animal kidney disease models *in vivo* (Csaba Révész and Péter Hamar (2011). *Delivery Methods to Target RNAs in the Kidney*, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-in-the-kidney). Delivery methods to the kidney may include those in Yuan et al. (Am J Physiol Renal Physiol 295: F605–F617, 2008). The method of Yuang et al. may be applied to the composition of the present invention contemplating a 1-2 g subcutaneous

injection of polypeptide nuclease conjugated with cholesterol to a human for delivery to the kidneys. In one embodiment, the method of Molitoris et al. (J Am Soc Nephrol 20: 1754–1764, 2009) can be adapted to the composition and a cumulative dose of 12- 20 mg/kg to a human can be used for delivery to the proximal tubule cells of the kidneys. In one embodiment, the methods of Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) can be adapted to the compositions and a dose of up to 25 mg/kg can be delivered via i.v. administration. In one embodiment, the method of Shimizu et al. (J Am Soc Nephrol 21: 622–633, 2010) can be adapted to the compositions and a dose of about of 10-20 μ mol compositions complexed with nanocarriers in about 1-2 liters of a physiologic fluid for i.p. administration can be used.

[0524] Other various delivery vehicles can be used to deliver the composition, system to the kidney such as viral, hydrodynamic, lipid, polymer nanoparticles, aptamers and various combinations thereof (*See, e.g.*, Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269); Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888); Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980); Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289); Q. Zhang et al., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13); Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331); Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103); Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693); Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157); Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8 pp. (1754-1764); Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226); Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101); Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251); Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108); Shimizu et al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633); Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737); Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212); Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637); Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178).

[0525] In one embodiment, delivery is to liver cells. In one embodiment, the liver cell is a hepatocyte. Delivery of the composition and system herein may be via viral vectors, especially

AAV (and in particular AAV2/6) vectors. These can be administered by intravenous injection. A preferred target for the liver, whether *in vitro* or *in vivo*, is the albumin gene. This is a so-called ‘safe harbor’ as albumin is expressed at very high levels and so some reduction in the production of albumin following successful gene editing is tolerated. It is also preferred as the high levels of expression seen from the albumin promoter/enhancer allows for useful levels of correct or transgene production (from the inserted recombination template) to be achieved even if only a small fraction of hepatocytes is edited. *See* sites identified by Wechsler et al. (reported at the 57th Annual Meeting and Exposition of the American Society of Hematology - abstract available online at ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015) which can be adapted for use with the compositions, systems, herein.

[0526] Exemplary liver and kidney diseases that can be treated and/or prevented are described elsewhere herein.

Treating Epithelial and Lung Diseases

[0527] In one embodiment, the disease treated or prevented by the composition and system disclosed herein can be a lung or epithelial disease. The compositions and systems disclosed herein can be used for treating epithelial and/or lung diseases. The present invention also contemplates delivering the composition, system, disclosed herein, to one or both lungs.

[0528] In one embodiment, as viral vector can be used to deliver the composition, system, or component thereof to the lungs. In one embodiment, the AAV is an AAV-1, AAV-2, AAV-5, AAV-6, and/or AAV-9 for delivery to the lungs. (*See, e.g.*, Li et al., *Molecular Therapy*, vol. 17 no. 12, 2067-2077 Dec 2009). In one embodiment, the MOI can vary from 1×10^3 to 4×10^5 vector genomes/cell. In one embodiment, the delivery vector can be an RSV vector as in Zamora et al. (*Am J Respir Crit Care Med* Vol 183. pp 531–538, 2011. The method of Zamora et al. may be applied to the polynucleotide-targeting system of the present invention and an aerosolized composition, for example with a dosage of 0.6 mg/kg, may be contemplated for the present invention.

[0529] Subjects treated for a lung disease may for example receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a

promoter for Type V Cas, U6 or H1 promoter for guide molecule). A preferred arrangement is to use a CFTRdelta508 targeting guide molecule, a repair template for deltaF508 mutation and a codon optimized composition, with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs.

Treating Diseases of the Skin

[0530] The compositions and systems disclosed herein can be used for the treatment of skin diseases. The present invention also contemplates delivering the composition and system, disclosed herein, to the skin.

[0531] In one embodiment, delivery to the skin (intra dermal delivery) of the composition, system, or component thereof can be via one or more microneedles or microneedle containing device. For example, In one embodiment the device and methods of Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) can be used and/or adapted to deliver the composition, system, disclosed herein, for example, at a dosage of up to 300 µl of 0.1 mg/ml compositions to the skin.

[0532] In one embodiment, the methods and techniques of Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442–446 Feb. 2010) can be used and/or adapted for delivery of a compositions disclosed herein to the skin.

[0533] In one embodiment, the methods and techniques of Zheng et al. (PNAS, July 24, 2012, vol. 109, no. 30, 11975–11980) can be used and/or adapted for nanoparticle delivery of a compositions disclosed herein to the skin. In one embodiment, as dosage of about 25 nM applied in a single application can achieve gene knockdown in the skin.

Treating Cancer

[0534] The compositions, systems, disclosed herein can be used for the treatment of cancer. The present invention also contemplates delivering the composition, system, disclosed herein, to a cancer cell. Also, as is described elsewhere herein the compositions, systems, can be used to modify an immune cell, such as a CAR or CAR T cell, which can then in turn be used to treat and/or prevent cancer. This is also described in International Patent Publication No. WO 2015/161276, the disclosure of which is hereby incorporated by reference and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

[0535] Target genes suitable for the treatment or prophylaxis of cancer can include those set forth in **Tables 5 and 6**. In one embodiment, target genes for cancer treatment and

prevention can also include those described in International Patent Publication No. WO 2015/048577 the disclosure of which is hereby incorporated by reference and can be adapted for use with the engineered Type V Cas proteins, CRISPR-Cas systems, compositions, nucleic acid molecules, cargos/delivery vehicles, delivery systems, and methods disclosed herein.

Adoptive Cell Therapy

[0536] The compositions, systems, and components thereof disclosed herein can be used to modify cells for an adoptive cell therapy. In an aspect of the invention, methods and compositions which involve editing a target polynucleotide sequence, or modulating expression of a target polynucleotide sequence, and applications thereof in connection with cancer immunotherapy are comprehended by adapting the composition, system, of the present invention. In some examples, the compositions, systems, and methods may be used to modify a stem cell (e.g., induced pluripotent cell) to derive modified natural killer cells, gamma delta T cells, and alpha beta T cells, which can be used for the adoptive cell therapy. In certain examples, the compositions, systems, and methods may be used to modify modified natural killer cells, gamma delta T cells, and alpha beta T cells.

[0537] As used herein, unless otherwise indicated, “ACT”, “adoptive cell therapy” and “adoptive cell transfer” may be used interchangeably. In one embodiment, Adoptive cell therapy (ACT) can refer to the transfer of cells to a patient with the goal of transferring the functionality and characteristics into the new host by engraftment of the cells (*See, e.g.,* Mettananda et al., Editing an α -globin enhancer in primary human hematopoietic stem cells as a treatment for β -thalassemia, Nat Commun. 2017 Sep 4;8(1):424). As used herein, unless otherwise indicated, the term "engraft" or "engraftment" refers to the process of cell incorporation into a tissue of interest in vivo through contact with existing cells of the tissue. Adoptive cell therapy (ACT) can refer to the transfer of cells, most commonly immune-derived cells, back into the same patient or into a new recipient host with the goal of transferring the immunologic functionality and characteristics into the new host. If possible, use of autologous cells helps the recipient by minimizing GVHD issues. The adoptive transfer of autologous tumor infiltrating lymphocytes (TIL) (Zacharakis et al., (2018) Nat Med. 2018 Jun;24(6):724-730; Besser et al., (2010) Clin. Cancer Res 16 (9) 2646–55; Dudley et al., (2002) Science 298 (5594): 850–4; and Dudley et al., (2005) Journal of Clinical Oncology 23 (10): 2346–57.) or genetically re-directed peripheral blood mononuclear cells (Johnson et al., (2009) Blood 114 (3): 535–46; and Morgan et al., (2006) Science 314(5796) 126-9) has been used to successfully

treat patients with advanced solid tumors, including melanoma, metastatic breast cancer and colorectal carcinoma, as well as patients with CD19-expressing hematologic malignancies (Kalos et al., (2011) *Science Translational Medicine* 3 (95): 95ra73). In one embodiment, allogenic cells immune cells are transferred (*See, e.g.*, Ren et al., (2017) *Clin Cancer Res* 23 (9) 2255-2266). As described further herein, allogenic cells can be edited to reduce alloreactivity and prevent graft-versus-host disease. Thus, use of allogenic cells allows for cells to be obtained from healthy donors and prepared for use in patients as opposed to preparing autologous cells from a patient after diagnosis.

[0538] Aspects of the invention involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens or tumor specific neoantigens (*See, e.g.*, Maus et al., 2014, *Adoptive Immunotherapy for Cancer or Viruses*, *Annual Review of Immunology*, Vol. 32: 189-225; Rosenberg and Restifo, 2015, *Adoptive cell transfer as personalized immunotherapy for human cancer*, *Science* Vol. 348 no. 6230 pp. 62-68; Restifo et al., 2015, *Adoptive immunotherapy for cancer: harnessing the T cell response*. *Nat. Rev. Immunol.* 12(4): 269-281; and Jenson and Riddell, 2014, *Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells*. *Immunol Rev.* 257(1): 127-144; and Rajasagi et al., 2014, *Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia*. *Blood*. 2014 Jul 17;124(3):453-62).

[0539] In one embodiment, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: MR1 (*See, e.g.*, Crowther, et al., 2020, *Genome-wide CRISPR-Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1*, *Nature Immunology* volume 21, pages178-185), B cell maturation antigen (BCMA) (*See, e.g.*, Friedman et al., *Effective Targeting of Multiple BCMA-Expressing Hematological Malignancies by Anti-BCMA CAR T Cells*, *Hum Gene Ther.* 2018 Mar 8; Berdeja JG, et al. *Durable clinical responses in heavily pretreated patients with relapsed/refractory multiple myeloma: updated results from a multicenter study of bb2121 anti-Bcma CAR T cell therapy*. *Blood*. 2017;130:740; and Mouhieddine and Ghobrial, *Immunotherapy in Multiple Myeloma: The Era of CAR T Cell Therapy*, *Hematologist*, May-June 2018, Volume 15, issue 3); PSA (prostate-specific antigen); prostate-specific membrane antigen (PSMA); PSCA (Prostate stem cell antigen); Tyrosine-protein kinase transmembrane receptor ROR1; fibroblast activation protein (FAP); Tumor-

associated glycoprotein 72 (TAG72); Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); Mesothelin; Human Epidermal growth factor Receptor 2 (ERBB2 (Her2/neu)); Prostase; Prostatic acid phosphatase (PAP); elongation factor 2 mutant (ELF2M); Insulin-like growth factor 1 receptor (IGF-1R); gplOO; BCR-ABL (breakpoint cluster region-Abelson); tyrosinase; New York esophageal squamous cell carcinoma 1 (NY-ESO-1); κ -light chain, LAGE (L antigen); MAGE (melanoma antigen); Melanoma-associated antigen 1 (MAGE-A1); MAGE A3; MAGE A6; legumain; Human papillomavirus (HPV) E6; HPV E7; prostatein; survivin; PCTA1 (Galectin 8); Melan-A/MART-1; Ras mutant; TRP-1 (tyrosinase related protein 1, or gp75); Tyrosinase-related Protein 2 (TRP2); TRP-2/INT2 (TRP-2/intron 2); RAGE (renal antigen); receptor for advanced glycation end products 1 (RAGE1); Renal ubiquitous 1, 2 (RU1, RU2); intestinal carboxyl esterase (iCE); Heat shock protein 70-2 (HSP70-2) mutant; thyroid stimulating hormone receptor (TSHR); CD123; CD171; CD19; CD20; CD22; CD26; CD30; CD33; CD44v7/8 (cluster of differentiation 44, exons 7/8); CD53; CD92; CD100; CD148; CD150; CD200; CD261; CD262; CD362; CS-1 (CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1); ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); Tn antigen (Tn Ag); Fms-Like Tyrosine Kinase 3 (FLT3); CD38; CD138; CD44v6; B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2); Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); stage-specific embryonic antigen-4 (SSEA-4); Mucin 1, cell surface associated (MUC1); mucin 16 (MUC16); epidermal growth factor receptor (EGFR); epidermal growth factor receptor variant III (EGFRvIII); neural cell adhesion molecule (NCAM); carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); ephrin type-A receptor 2 (EphA2); Ephrin B2; Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); TGS5; high molecular weight-melanoma-associated antigen (HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor alpha; Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); G protein-coupled receptor class C group 5, member D (GPRC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide

portion of globoH glycosphingolipid (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); CT (cancer/testis (antigen)); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; p53; p53 mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP); ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; Cyclin D1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS); Squamous Cell Carcinoma Antigen Recognized By T Cells-1 or 3 (SART1, SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint-1, -2, -3 or -4 (SSX1, SSX2, SSX3, SSX4); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRL5); mouse double minute 2 homolog (MDM2); livin; alphafetoprotein (AFP); transmembrane activator and CAML Interactor (TACI); B-cell activating factor receptor (BAFF-R); V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS); immunoglobulin lambda-like polypeptide 1 (IGLL1); 707-AP (707 alanine proline); ART-4 (adenocarcinoma antigen recognized by T4 cells); BAGE (B antigen; b-catenin/m, b-catenin/mutated); CAMEL (CTL-recognized antigen on melanoma); CAP1 (carcinoembryonic antigen peptide 1); CASP-8 (caspase-8); CDC27m (cell-division cycle 27 mutated); CDK4/m (cyclin-dependent kinase 4 mutated); Cyp-B (cyclophilin B); DAM (differentiation antigen melanoma); EGP-2 (epithelial

glycoprotein 2); EGP-40 (epithelial glycoprotein 40); ErbB2, 3, 4 (erythroblastic leukemia viral oncogene homolog-2, -3, 4); FBP (folate binding protein); fAChR (Fetal acetylcholine receptor); G250 (glycoprotein 250); GAGE (G antigen); GnT-V (N-acetylglucosaminyltransferase V); HAGE (helicose antigen); ULA-A (human leukocyte antigen-A); HST2 (human signet ring tumor 2); KIAA0205; KDR (kinase insert domain receptor); LDLR/FUT (low density lipid receptor/GDP L-fucose: b-D-galactosidase 2-a-L fucosyltransferase); L1CAM (L1 cell adhesion molecule); MC1R (melanocortin 1 receptor); Myosin/m (myosin mutated); MUM-1, -2, -3 (melanoma ubiquitous mutated 1, 2, 3); NA88-A (NA cDNA clone of patient M88); KG2D (Natural killer group 2, member D) ligands; oncofetal antigen (h5T4); p190 minor bcr-abl (protein of 190KD bcr-abl); Pml/RARa (promyelocytic leukemia/retinoic acid receptor a); PRAME (preferentially expressed antigen of melanoma); SAGE (sarcoma antigen); TEL/AML1 (translocation Ets-family leukemia/acute myeloid leukemia 1); TPI/m (triosephosphate isomerase mutated); CD70; and any combination thereof. This approach can be adapted to and/or applied to the Type V Cas compositions, systems, disclosed herein.

[0540] In one embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-specific antigen (TSA).

[0541] In one embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a neoantigen.

[0542] In one embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-associated antigen (TAA).

[0543] In one embodiment, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a universal tumor antigen. In certain preferred embodiments, the universal tumor antigen is selected from the group consisting of: a human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (D1), and any combinations thereof.

[0544] In one embodiment, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: CD19, BCMA, CD70, CLL-1, MAGE A3, MAGE A6, HPV E6, HPV E7, WT1, CD22, CD171, ROR1, MUC16, and SSX2. In certain preferred embodiments, the antigen may be CD19. For example, CD19 may be targeted in hematologic malignancies, such as in lymphomas, more particularly in B-cell lymphomas, such as without limitation in diffuse large B-cell lymphoma, primary mediastinal b-cell lymphoma, transformed follicular lymphoma, marginal zone lymphoma, mantle cell lymphoma, acute lymphoblastic leukemia including adult and pediatric ALL, non-Hodgkin lymphoma, indolent non-Hodgkin lymphoma, or chronic lymphocytic leukemia. For example, BCMA may be targeted in multiple myeloma or plasma cell leukemia (*See, e.g.*, 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic Chimeric Antigen Receptor T Cells Targeting B Cell Maturation Antigen). For example, CLL1 may be targeted in acute myeloid leukemia. For example, MAGE A3, MAGE A6, SSX2, and/or KRAS may be targeted in solid tumors. For example, HPV E6 and/or HPV E7 may be targeted in cervical cancer or head and neck cancer. For example, WT1 may be targeted in acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), chronic myeloid leukemia (CML), non-small cell lung cancer, breast, pancreatic, ovarian, or colorectal cancers, or mesothelioma. For example, CD22 may be targeted in B cell malignancies, including non-Hodgkin lymphoma, diffuse large B-cell lymphoma, or acute lymphoblastic leukemia. For example, CD171 may be targeted in neuroblastoma, glioblastoma, or lung, pancreatic, or ovarian cancers. For example, ROR1 may be targeted in ROR1+ malignancies, including non-small cell lung cancer, triple negative breast cancer, pancreatic cancer, prostate cancer, ALL, chronic lymphocytic leukemia, or mantle cell lymphoma. For example, MUC16 may be targeted in MUC16ecto+ epithelial ovarian, fallopian tube or primary peritoneal cancer. For example, CD70 may be targeted in both hematologic malignancies as well as in solid cancers such as renal cell carcinoma (RCC), gliomas (e.g., GBM), and head and neck cancers (HNSCC). CD70 is expressed in both hematologic malignancies as well as in solid cancers, while its expression in normal tissues is restricted to a subset of lymphoid cell types (*See, e.g.*, 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic CRISPR Engineered Anti-CD70 CAR-T Cells Demonstrate Potent Preclinical Activity Against Both Solid and Hematological Cancer Cells).

[0545] Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (*See* U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[0546] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (*See* U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO 9215322).

[0547] In general, CARs are comprised of an extracellular domain, a transmembrane domain, and an intracellular domain, wherein the extracellular domain comprises an antigen-binding domain that is specific for a predetermined target. While the antigen-binding domain of a CAR is often an antibody or antibody fragment (e.g., a single chain variable fragment, scFv), the binding domain is not particularly limited so long as it results in specific recognition of a target. For example, In one embodiment, the antigen-binding domain may comprise a receptor, such that the CAR is capable of binding to the ligand of the receptor. Alternatively, the antigen-binding domain may comprise a ligand, such that the CAR is capable of binding the endogenous receptor of that ligand.

[0548] The antigen-binding domain of a CAR is generally separated from the transmembrane domain by a hinge or spacer. The spacer is also not particularly limited, and it is designed to provide the CAR with flexibility. For example, a spacer domain may comprise a portion of a human Fc domain, including a portion of the CH3 domain, or the hinge region of any immunoglobulin, such as IgA, IgD, IgE, IgG, or IgM, or variants thereof. Furthermore, the hinge region may be modified so as to prevent off-target binding by FcRs or other potential interfering objects. For example, the hinge may comprise an IgG4 Fc domain with or without a S228P, L235E, and/or N297Q mutation (according to Kabat numbering) in order to decrease binding to FcRs. Additional spacers/hinges include, but are not limited to, CD4, CD8, and CD28 hinge regions.

[0549] The transmembrane domain of a CAR may be derived either from a natural or from a synthetic source. Where the source is natural, the domain may be derived from any membrane bound or transmembrane protein. Transmembrane regions of particular use in this disclosure may be derived from CD8, CD28, CD3, CD45, CD4, CD5, CDS, CD9, CD 16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD137, CD 154, TCR. Alternatively, the transmembrane domain may be synthetic, in which case it will comprise predominantly hydrophobic residues such as leucine and valine. Preferably a triplet of phenylalanine, tryptophan and valine will be found at each end of a synthetic transmembrane domain. Optionally, a short oligo- or polypeptide linker, preferably between 2 and 10 amino acids in length may form the linkage between the transmembrane domain and the cytoplasmic signaling domain of the CAR. A glycine-serine doublet provides a particularly suitable linker.

[0550] Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8 α hinge domain and a CD8 α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3 ζ or FcR γ (scFv-CD3 ζ or scFv-FcR γ ; *See* U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3 ζ ; *See* U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3 ζ -chain, CD97, GDI 1a-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, CD2, CD7, LIGHT, LFA-1, NKG2C, B7-H3, CD30, CD40, PD-1, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3 ζ or scFv-CD28-OX40-CD3 ζ ; *See* U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO 2014/134165; PCT Publication No. WO 2012/079000). In one embodiment, the primary signaling domain comprises a functional signaling domain of a protein selected from the group consisting of CD3 zeta, CD3 gamma, CD3 delta, CD3 epsilon, common FcR gamma (FCERIG), FcR beta (Fc Epsilon R1b), CD79a, CD79b, Fc gamma RIIa, DAP10, and DAP12. In certain preferred embodiments, the primary signaling domain comprises a functional signaling domain of CD3 ζ or FcR γ . In one embodiment, the one or more costimulatory signaling domains comprise a

functional signaling domain of a protein selected, each independently, from the group consisting of: CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, CDS, ICAM-1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), CD160, CD19, CD4, CD8 alpha, CD8 beta, IL2R beta, IL2R gamma, IL7R alpha, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CD11d, ITGAE, CD103, ITGAL, CD11a, LFA-1, ITGAM, CD11b, ITGAX, CD11c, ITGB1, CD29, ITGB2, CD18, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Lyl08), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, NKp44, NKp30, NKp46, and NKG2D. In one embodiment, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: 4-1BB, CD27, and CD28. In one embodiment, a chimeric antigen receptor may have the design as described in U.S. Patent No. 7,446,190, comprising an intracellular domain of CD3 ζ chain (such as amino acid residues 52-163 of the human CD3 zeta chain, as shown in SEQ ID NO: 14 of US 7,446,190), a signaling region from CD28 and an antigen-binding element (or portion or domain; such as scFv). The CD28 portion, when between the zeta chain portion and the antigen-binding element, may suitably include the transmembrane and signaling domains of CD28 (such as amino acid residues 114-220 of SEQ ID NO: 10, full sequence shown in SEQ ID NO: 6 of US 7,446,190; these can include the following portion of CD28 as set forth in Genbank identifier NM_006139. Alternatively, when the zeta sequence lies between the CD28 sequence and the antigen-binding element, intracellular domain of CD28 can be used alone (such as amino sequence set forth in SEQ ID NO: 9 of US 7,446,190). Hence, certain embodiments employ a CAR comprising (a) a zeta chain portion comprising the intracellular domain of human CD3 ζ chain, (b) a costimulatory signaling region, and (c) an antigen-binding element (or portion or domain), wherein the costimulatory signaling region comprises the amino acid sequence encoded by SEQ ID NO: 6 of US 7,446,190.

[0551] Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native $\alpha\beta$ TCR, for example by antigen on professional antigen-presenting cells, with attendant

costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects.

[0552] By means of an example and without limitation, Kochenderfer et al., (2009) *J Immunother.* 32 (7): 689-702 described anti-CD19 chimeric antigen receptors (CAR). FMC63-28Z CAR contained a single chain variable region moiety (scFv) recognizing CD19 derived from the FMC63 mouse hybridoma (described in Nicholson et al., (1997) *Molecular Immunology* 34: 1157–1165), a portion of the human CD28 molecule, and the intracellular component of the human TCR- ζ molecule. FMC63-CD828BBZ CAR contained the FMC63 scFv, the hinge and transmembrane regions of the CD8 molecule, the cytoplasmic portions of CD28 and 4-1BB, and the cytoplasmic component of the TCR- ζ molecule. The exact sequence of the CD28 molecule included in the FMC63-28Z CAR corresponded to Genbank identifier NM_006139; the sequence included all amino acids starting with the amino acid sequence IEVMYPPPY (SEQ ID NO: 55). and continuing all the way to the carboxy-terminus of the protein. To encode the anti-CD19 scFv component of the vector, the authors designed a DNA sequence which was based on a portion of a previously published CAR (Cooper et al., (2003) *Blood* 101: 1637–1644). This sequence encoded the following components in frame from the 5' end to the 3' end: an XhoI site, the human granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor α -chain signal sequence, the FMC63 light chain variable region (as in Nicholson et al., supra), a linker peptide (as in Cooper et al., supra), the FMC63 heavy chain variable region (as in Nicholson et al., supra), and a NotI site. A plasmid encoding this sequence was digested with XhoI and NotI. To form the MSGV-FMC63-28Z retroviral vector, the XhoI and NotI-digested fragment encoding the FMC63 scFv was ligated into a second XhoI and NotI-digested fragment that encoded the MSGV retroviral backbone (as in Hughes et al., (2005) *Human Gene Therapy* 16: 457–472) as well as part of the extracellular portion of human CD28, the entire transmembrane and cytoplasmic portion of human CD28, and the cytoplasmic portion of the human TCR- ζ molecule (as in Maher et al., 2002) *Nature Biotechnology* 20: 70–75). The FMC63-28Z CAR is included in the KTE-C19 (axicabtagene ciloleucel) anti-CD19 CAR-T therapy product in development by Kite Pharma, Inc. for the treatment of inter alia patients with relapsed/refractory aggressive B-cell non-Hodgkin lymphoma (NHL). Accordingly, in one embodiment, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may express the FMC63-28Z CAR as described by

Kochenderfer et al. (supra). Hence, in one embodiment, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element (or portion or domain, such as scFv) that specifically binds to an antigen, an intracellular signaling domain comprising an intracellular domain of a CD3 ζ chain, and a costimulatory signaling region comprising a signaling domain of CD28. Preferably, the CD28 amino acid sequence is as set forth in Genbank identifier NM_006139 (sequence version 1, 2 or 3) starting with the amino acid sequence IEVMYPPPY (SEQ ID NO: 55) and continuing all the way to the carboxy-terminus of the protein. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the anti-CD19 scFv as described by Kochenderfer et al. (supra).

[0553] Additional anti-CD19 CARs are further described in International Patent Publication No. WO 2015/187528. More particularly Example 1 and Table 1 of WO2015187528, incorporated by reference herein, demonstrate the generation of anti-CD19 CARs based on a fully human anti-CD19 monoclonal antibody (47G4, as described in US20100104509) and murine anti-CD19 monoclonal antibody (as described in Nicholson et al. and explained above). Various combinations of a signal sequence (human CD8-alpha or GM-CSF receptor), extracellular and transmembrane regions (human CD8-alpha) and intracellular T-cell signaling domains (CD28-CD3 ζ ; 4-1BB-CD3 ζ ; CD27-CD3 ζ ; CD28-CD27-CD3 ζ ; 4-1BB-CD27-CD3 ζ ; CD27-4-1BB-CD3 ζ ; CD28-CD27-Fc ϵ RI gamma chain; or CD28-Fc ϵ RI gamma chain) were disclosed. Hence, in one embodiment, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element that specifically binds to an antigen, an extracellular and transmembrane region as set forth in Table 1 of WO2015187528 and an intracellular T-cell signaling domain as set forth in Table 1 of International Application No. WO 2015/187528. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the mouse or human anti-CD19 scFv as described in Example 1 of WO 2015/187528. In one embodiment, the CAR comprises, consists essentially of or consists of an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, or SEQ ID NO: 13 as set forth in Table 1 of WO2015187528.

[0554] By means of an example and without limitation, chimeric antigen receptor that recognizes the CD70 antigen is described in WO2012058460A2 (*See also*, Park et al., CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma, *Oral Oncol.* 2018 Mar;78:145-150; and Jin et al., CD70, a novel target of CAR T-cell therapy for gliomas, *Neuro Oncol.* 2018 Jan 10;20(1):55-65). CD70 is expressed by diffuse large B-cell and follicular lymphoma and also by the malignant cells of Hodgkins lymphoma, Waldenstrom's macroglobulinemia and multiple myeloma, and by HTLV-1- and EBV-associated malignancies. (Agathangelou et al. *Am.J.Pathol.* 1995;147: 1152-1160; Hunter et al., *Blood* 2004; 104:4881. 26; Lens et al., *J Immunol.* 2005;174:6212-6219; Baba et al., *J Virol.* 2008;82:3843-3852.) In addition, CD70 is expressed by non-hematological malignancies such as renal cell carcinoma and glioblastoma. (Junker et al., *J Urol.* 2005;173:2150-2153; Chahlavi et al., *Cancer Res* 2005;65:5428-5438) Physiologically, CD70 expression is transient and restricted to a subset of highly activated T, B, and dendritic cells.

[0555] By means of an example and without limitation, chimeric antigen receptor that recognizes BCMA has been described (*See, e.g.*, US20160046724A1; WO2016014789A2; WO2017211900A1; WO2015158671A1; US20180085444A1; WO2018028647A1; US20170283504A1; and WO2013154760A1).

[0556] In one embodiment, the immune cell may, in addition to a CAR or exogenous TCR as disclosed herein, further comprise a chimeric inhibitory receptor (inhibitory CAR) that specifically binds to a second target antigen and is capable of inducing an inhibitory or immunosuppressive or repressive signal to the cell upon recognition of the second target antigen. In one embodiment, the chimeric inhibitory receptor comprises an extracellular antigen-binding element (or portion or domain) configured to specifically bind to a target antigen, a transmembrane domain, and an intracellular immunosuppressive or repressive signaling domain. In one embodiment, the second target antigen is an antigen that is not expressed on the surface of a cancer cell or infected cell or the expression of which is downregulated on a cancer cell or an infected cell. In one embodiment, the second target antigen is an MHC-class I molecule. In one embodiment, the intracellular signaling domain comprises a functional signaling portion of an immune checkpoint molecule, such as for example PD-1 or CTLA4. Advantageously, the inclusion of such inhibitory CAR reduces the chance of the engineered immune cells attacking non-target (e.g., non-cancer) tissues.

[0557] Alternatively, T-cells expressing CARs may be further modified to reduce or eliminate expression of endogenous TCRs in order to reduce off-target effects. Reduction or elimination of endogenous TCRs can reduce off-target effects and increase the effectiveness of the T cells (U.S. 9,181,527). T cells stably lacking expression of a functional TCR may be produced using a variety of approaches. T cells internalize, sort, and degrade the entire T cell receptor as a complex, with a half-life of about 10 hours in resting T cells and 3 hours in stimulated T cells (von Essen, M. et al. 2004. J. Immunol. 173:384-393). Proper functioning of the TCR complex requires the proper stoichiometric ratio of the proteins that compose the TCR complex. TCR function also requires two functioning TCR zeta proteins with ITAM motifs. The activation of the TCR upon engagement of its MHC-peptide ligand requires the engagement of several TCRs on the same T cell, which all must signal properly. Thus, if a TCR complex is destabilized with proteins that do not associate properly or cannot signal optimally, the T cell will not become activated sufficiently to begin a cellular response.

[0558] Accordingly, in one embodiment, TCR expression may be eliminated using RNA interference (e.g., sgRNA molecule, siRNA, miRNA, etc.), Type V Cas polypeptide, or other methods that target the nucleic acids encoding specific TCRs (e.g., TCR- α and TCR- β) and/or CD3 chains in primary T cells. By blocking expression of one or more of these proteins, the T cell will no longer produce one or more of the key components of the TCR complex, thereby destabilizing the TCR complex and preventing cell surface expression of a functional TCR.

[0559] In some instances, CAR may also comprise a switch mechanism for controlling expression and/or activation of the CAR. For example, a CAR may comprise an extracellular, transmembrane, and intracellular domain, in which the extracellular domain comprises a target-specific binding element that comprises a label, binding domain, or tag that is specific for a molecule other than the target antigen that is expressed on or by a target cell. In such embodiments, the specificity of the CAR is provided by a second construct that comprises a target antigen binding domain (e.g., an scFv or a bispecific antibody that is specific for both the target antigen and the label or tag on the CAR) and a domain that is recognized by or binds to the label, binding domain, or tag on the CAR. *See, e.g.,* WO 2013/044225, WO 2016/000304, WO 2015/057834, WO 2015/057852, WO 2016/070061, US 9,233,125, US 2016/0129109. In this way, a T-cell that expresses the CAR can be administered to a subject, but the CAR cannot bind its target antigen until the second composition comprising an antigen-specific binding domain is administered.

[0560] Alternative switch mechanisms include CARs that require multimerization in order to activate their signaling function (*See, e.g.*, US Patent Publication Nos. US 2015/0368342, US 2016/0175359, US 2015/0368360) and/or an exogenous signal, such as a small molecule drug (US 2016/0166613, Yung et al., *Science*, 2015), in order to elicit a T-cell response. Some CARs may also comprise a “suicide switch” to induce cell death of the CAR T-cells following treatment (Buddee et al., *PLoS One*, 2013) or to downregulate expression of the CAR following binding to the target antigen (International Patent Publication No. WO 2016/011210).

[0561] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids, or transposons, such as a Sleeping Beauty transposon (*See* U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3 ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[0562] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ -irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR⁺ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon- γ). CAR T cells of this kind may for example be used in animal models, for example to treat tumor xenografts.

[0563] In one embodiment, ACT includes co-transferring CD4⁺ Th1 cells and CD8⁺ CTLs to induce a synergistic antitumor response (*See, e.g.*, Li et al., *Adoptive cell therapy with CD4⁺ T helper 1 cells and CD8⁺ cytotoxic T cells enhances complete rejection of an established*

tumor, leading to generation of endogenous memory responses to non-targeted tumor epitopes. Clin Transl Immunology. 2017 Oct; 6(10): e160).

[0564] In one embodiment, Th17 cells are transferred to a subject in need thereof. Th17 cells have been reported to directly eradicate melanoma tumors in mice to a greater extent than Th1 cells (Muranski P, et al., Tumor-specific Th17-polarized cells eradicate large established melanoma. Blood. 2008 Jul 15; 112(2):362-73; and Martin-Orozco N, et al., T helper 17 cells promote cytotoxic T cell activation in tumor immunity. Immunity. 2009 Nov 20; 31(5):787-98). Those studies involved an adoptive T cell transfer (ACT) therapy approach, which takes advantage of CD4⁺ T cells that express a TCR recognizing tyrosinase tumor antigen. Exploitation of the TCR leads to rapid expansion of Th17 populations to large numbers *ex vivo* for reinfusion into the autologous tumor-bearing hosts.

[0565] In one embodiment, ACT may include autologous iPSC-based vaccines, such as irradiated iPSCs in autologous anti-tumor vaccines (*See, e.g.*, Kooreman, Nigel G. et al., Autologous iPSC-Based Vaccines Elicit Anti-tumor Responses *In Vivo*, Cell Stem Cell 22, 1–13, 2018, doi.org/10.1016/j.stem.2018.01.016).

[0566] Unlike T-cell receptors (TCRs) that are MHC restricted, CARs can potentially bind any cell surface-expressed antigen and can thus be more universally used to treat patients (*See* Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267). In one embodiment, in the absence of endogenous T-cell infiltrate (e.g., due to aberrant antigen processing and presentation), which precludes the use of TIL therapy and immune checkpoint blockade, the transfer of CAR T-cells may be used to treat patients (*See, e.g.*, Hinrichs CS, Rosenberg SA. Exploiting the curative potential of adoptive T-cell therapy for cancer. Immunol Rev (2014) 257(1):56–71. doi:10.1111/imr.12132).

[0567] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoresponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction).

[0568] In one embodiment, the treatment can be administered after lymphodepleting pretreatment in the form of chemotherapy (typically a combination of cyclophosphamide and

fludarabine) or radiation therapy. Initial studies in ACT had short lived responses and the transferred cells did not persist *in vivo* for very long (Houot et al., T-cell-based immunotherapy: adoptive cell transfer and checkpoint inhibition. *Cancer Immunol Res* (2015) 3(10):1115–22; and Kamta et al., Advancing Cancer Therapy with Present and Emerging Immuno-Oncology Approaches. *Front. Oncol.* (2017) 7:64). Immune suppressor cells like Tregs and MDSCs may attenuate the activity of transferred cells by outcompeting them for the necessary cytokines. Not being bound by a theory lymphodepleting pretreatment may eliminate the suppressor cells allowing the TILs to persist.

[0569] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment (e.g., glucocorticoid treatment). The cells or population of cells may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. In one embodiment, the immunosuppressive treatment provides for the selection and expansion of the immunoresponsive T cells within the patient.

[0570] In one embodiment, the treatment can be administered before primary treatment (e.g., surgery or radiation therapy) to shrink a tumor before the primary treatment. In another embodiment, the treatment can be administered after primary treatment to remove any remaining cancer cells.

[0571] In one embodiment, immunometabolic barriers can be targeted therapeutically prior to and/or during ACT to enhance responses to ACT or CAR T-cell therapy and to support endogenous immunity (*See, e.g.,* Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, *Front. Immunol.*, 03 April 2017, doi.org/10.3389/fimmu.2017.00267).

[0572] The administration of cells or population of cells, such as immune system cells or cell populations, such as more particularly immunoresponsive cells or cell populations, as disclosed herein may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation, or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, intrathecally, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the disclosed CARs may be delivered or administered into a cavity formed by the resection of tumor tissue (i.e., intracavity delivery) or

directly into a tumor prior to resection (i.e., intratumoral delivery). In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[0573] The administration of the cells or population of cells can consist of the administration of 10⁴- 10⁹ cells per kg body weight, preferably 10⁵ to 10⁶ cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies may for example involve administration of from 10⁶ to 10⁹ cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administered in one or more doses. In another embodiment, the effective amount of cells are administered as a single dose. In another embodiment, the effective amount of cells are administered as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administered will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[0574] In another embodiment, the effective amount of cells or composition comprising those cells are administered parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[0575] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. *Front. Pharmacol.* 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (*See* U.S. Patent Publication No. 20130071414; International Patent Publication WO

2011/146862; International Patent Publication WO 2014/011987; International Patent Publication WO 2013/040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[0576] In a further refinement of adoptive therapies, genome editing may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (*See* Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853; Ren et al., 2017, Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition, Clin Cancer Res. 2017 May 1;23(9):2255-2266. doi: 10.1158/1078-0432.CCR-16-1300. Epub 2016 Nov 4; Qasim et al., 2017, Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells, Sci Transl Med. 2017 Jan 25;9(374); Legut, et al., 2018, CRISPR-mediated TCR replacement generates superior anticancer transgenic T cells. Blood, 131(3), 311-322; and Georgiadis et al., Long Terminal Repeat CRISPR-CAR-Coupled "Universal" T Cells Mediate Potent Anti-leukemic Effects, Molecular Therapy, In Press, Corrected Proof, Available online 6 March 2018). Cells may be edited using any CRISPR-Cas system and method of use thereof as disclosed herein. The composition and systems may be delivered to an immune cell by any method disclosed herein. In preferred embodiments, cells are edited *ex vivo* and transferred to a subject in need thereof. Immunoresponsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed for example to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell (e.g. TRAC locus); to eliminate potential alloreactive T-cell receptors (TCR) or to prevent inappropriate pairing between endogenous and exogenous TCR chains, such as to knock-out or knock-down expression of an endogenous TCR in a cell; to disrupt the target of a chemotherapeutic agent in a cell; to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell; to knock-out or knock-down expression of other gene or genes in a cell, the reduced expression or lack of expression of which can enhance the efficacy of adoptive therapies using the cell; to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR; to knock-out or knock-down expression of one or more MHC constituent proteins in a cell; to activate a T cell; to modulate cells such that the cells are resistant to exhaustion or dysfunction; and/or increase

the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (See International Patent Publication Nos. WO 2013/176915, WO 2014/059173, WO 2014/172606, WO 2014/184744, and WO 2014/191128).

[0577] In one embodiment, editing may result in inactivation of a gene. By inactivating a gene, it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art. In one embodiment, homology directed repair (HDR) is used to concurrently inactivate a gene (e.g., TRAC) and insert an endogenous TCR or CAR into the inactivated locus.

[0578] Hence, in one embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell. Conventionally, nucleic acid molecules encoding CARs or TCRs are transfected or transduced to cells using randomly integrating vectors, which, depending on the site of integration, may lead to clonal expansion, oncogenic transformation, variegated transgene expression and/or transcriptional silencing of the transgene. Directing of transgene(s) to a specific locus in a cell can minimize or avoid such risks and advantageously provide for uniform expression of the transgene(s) by the cells. Without limitation, suitable 'safe harbor' loci for directed transgene integration include CCR5 or AAVS1. Homology-directed repair (HDR) strategies are known and described elsewhere in this specification allowing to insert transgenes into desired loci (e.g., TRAC locus).

[0579] Further suitable loci for insertion of transgenes, in particular CAR or exogenous TCR transgenes, include without limitation loci comprising genes coding for constituents of endogenous T-cell receptor, such as T-cell receptor alpha locus (TRA) or T-cell receptor beta locus (TRB), for example T-cell receptor alpha constant (TRAC) locus, T-cell receptor beta constant 1 (TRBC1) locus or T-cell receptor beta constant 2 (TRBC2) locus. Advantageously,

insertion of a transgene into such locus can simultaneously achieve expression of the transgene, potentially controlled by the endogenous promoter, and knock-out expression of the endogenous TCR. This approach has been exemplified in Eyquem et al., (2017) Nature 543: 113-117, wherein the authors used CRISPR/Cas9 gene editing to knock-in a DNA molecule encoding a CD19-specific CAR into the TRAC locus downstream of the endogenous promoter; the CAR-T cells obtained by CRISPR were significantly superior in terms of reduced tonic CAR signaling and exhaustion. This approach can be adapted to and/or applied to the Type V Cas compositions, systems, disclosed herein.

[0580] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β , which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the potential development of graft versus host disease (GVHD). The inactivation of TCR α or TCR β can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[0581] Hence, in one embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous TCR in a cell. For example, NHEJ-based or HDR-based gene editing approaches can be employed to disrupt the endogenous TCR alpha and/or beta chain genes. For example, gene editing system or systems, such as Type V Cas system or systems, can be designed to target a sequence found within the TCR beta chain conserved between the beta 1 and beta 2 constant region genes (TRBC1 and TRBC2) and/or to target the constant region of the TCR alpha chain (TRAC) gene.

[0582] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α -chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid, or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[0583] In one embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell. Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In one embodiment, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[0584] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? *Biochem Soc Trans.* 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. *Front. Immunol.* 6:418).

[0585] International Patent Publication No. WO 2014/172606 relates to the use of MT1 and/or MT2 inhibitors to increase proliferation and/or activity of exhausted CD8⁺ T-cells and to decrease CD8⁺ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8⁺ immune cells). In one embodiment, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[0586] In one embodiment, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBR1, TGFBR2, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1, TIM-3, CEACAM-1, CEACAM-3, or CEACAM-5. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[0587] By means of an example and without limitation, International Patent Publication No. WO 2016/196388 concerns an engineered T cell comprising (a) a genetically engineered antigen receptor that specifically binds to an antigen, which receptor may be a CAR; and (b) a disrupted gene encoding a PD-L1, an agent for disruption of a gene encoding a PD-L1, and/or

disruption of a gene encoding PD-L1, wherein the disruption of the gene may be mediated by a gene editing nuclease, a zinc finger nuclease (ZFN), CRISPR/Cas9 and/or TALEN. WO2015142675 relates to immune effector cells comprising a CAR in combination with an agent (such as the composition or system herein) that increases the efficacy of the immune effector cells in the treatment of cancer, wherein the agent may inhibit an immune inhibitory molecule, such as PD1, PD-L1, CTLA-4, TIM-3, LAG-3, VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, TGFR beta, CEACAM-1, CEACAM-3, or CEACAM-5. Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas9 mRNA and gRNAs targeting endogenous TCR, β -2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1. This approach can be adapted to and/or applied to the Type V Cas compositions, systems, disclosed herein.

[0588] In one embodiment, cells may be engineered to express a CAR, wherein expression and/or function of methylcytosine dioxygenase genes (TET1, TET2 and/or TET3) in the cells has been reduced or eliminated, (such as the composition or system herein) (for example, as described in WO201704916).

[0589] In one embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR, thereby reducing the likelihood of targeting of the engineered cells. In one embodiment, the targeted antigen may be one or more antigen selected from the group consisting of CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, CD362, human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (D1), B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), and B-cell activating factor receptor (BAFF-R) (for example, as described in International Patent Publication Nos. WO 2016/011210 and WO 2017/011804).

[0590] In one embodiment, editing of cells, particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to

knock-out or knock-down expression of one or more MHC constituent proteins, such as one or more HLA proteins and/or beta-2 microglobulin (B2M), in a cell, whereby rejection of non-autologous (e.g., allogeneic) cells by the recipient's immune system can be reduced or avoided. In preferred embodiments, one or more HLA class I proteins, such as HLA-A, B and/or C, and/or B2M may be knocked-out or knocked-down. Preferably, B2M may be knocked-out or knocked-down. By means of an example, Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas mRNA and gRNAs targeting endogenous TCR, β -2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[0591] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCR α , PD1 and TCR β , CTLA-4 and TCR α , CTLA-4 and TCR β , LAG3 and TCR α , LAG3 and TCR β , Tim3 and TCR α , Tim3 and TCR β , BTLA and TCR α , BTLA and TCR β , BY55 and TCR α , BY55 and TCR β , TIGIT and TCR α , TIGIT and TCR β , B7H5 and TCR α , B7H5 and TCR β , LAIR1 and TCR α , LAIR1 and TCR β , SIGLEC10 and TCR α , SIGLEC10 and TCR β , 2B4 and TCR α , 2B4 and TCR β , B2M and TCR α , B2M and TCR β .

[0592] In one embodiment, a cell may be multiplied edited (multiplex genome editing) as taught herein to (1) knock-out or knock-down expression of an endogenous TCR (for example, TRBC1, TRBC2 and/or TRAC), (2) knock-out or knock-down expression of an immune checkpoint protein or receptor (for example PD1, PD-L1 and/or CTLA4); and (3) knock-out or knock-down expression of one or more MHC constituent proteins (for example, HLA-A, B and/or C, and/or B2M, preferably B2M).

[0593] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patent Nos. 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded *in vitro* or *in vivo*.

[0594] Immune cells may be obtained using any method known in the art. In one embodiment, allogeneic T cells may be obtained from healthy subjects. In one embodiment T cells that have infiltrated a tumor are isolated. T cells may be removed during surgery. T cells may be isolated after removal of tumor tissue by biopsy. T cells may be isolated by any means known in the art. In one embodiment, T cells are obtained by apheresis. In one embodiment,

the method may comprise obtaining a bulk population of T cells from a tumor sample by any suitable method known in the art. For example, a bulk population of T cells can be obtained from a tumor sample by dissociating the tumor sample into a cell suspension from which specific cell populations can be selected. Suitable methods of obtaining a bulk population of T cells may include, but are not limited to, any one or more of mechanically dissociating (e.g., mincing) the tumor, enzymatically dissociating (e.g., digesting) the tumor, and aspiration (e.g., as with a needle).

[0595] The bulk population of T cells obtained from a tumor sample may comprise any suitable type of T cell. Preferably, the bulk population of T cells obtained from a tumor sample comprises tumor infiltrating lymphocytes (TILs).

[0596] The tumor sample may be obtained from any mammal. Unless stated otherwise, as used herein, unless otherwise indicated, the term "mammal" refers to any mammal including, but not limited to, mammals of the order Logomorpha, such as rabbits; the order Carnivora, including Felines (cats) and Canines (dogs); the order Artiodactyla, including Bovines (cows) and Swines (pigs); or of the order Perssodactyla, including Equines (horses). The mammals may be non-human primates, e.g., of the order Primates, Ceboids, or Simoids (monkeys) or of the order Anthropoids (humans and apes). In one embodiment, the mammal may be a mammal of the order Rodentia, such as mice and hamsters. Preferably, the mammal is a non-human primate or a human. An especially preferred mammal is the human.

[0597] T cells can be obtained from a number of sources, including peripheral blood mononuclear cells (PBMC), bone marrow, lymph node tissue, spleen tissue, and tumors. In one embodiment of the present invention, T cells can be obtained from a unit of blood collected from a subject using any number of techniques known to the skilled artisan, such as Ficoll separation. In one preferred embodiment, cells from the circulating blood of an individual are obtained by apheresis or leukapheresis. The apheresis product typically contains lymphocytes, including T cells, monocytes, granulocytes, B cells, other nucleated white blood cells, red blood cells, and platelets. In one embodiment, the cells collected by apheresis may be washed to remove the plasma fraction and to place the cells in an appropriate buffer or media for subsequent processing steps. In one embodiment of the invention, the cells are washed with phosphate buffered saline (PBS). In an alternative embodiment, the wash solution lacks calcium and may lack magnesium or may lack many if not all divalent cations. Initial activation steps in the absence of calcium lead to magnified activation. As those of ordinary skill in the

art would readily appreciate a washing step may be accomplished by methods known to those in the art, such as by using a semi-automated “flow-through” centrifuge (for example, the Cobe 2991 cell processor) according to the manufacturer's instructions. After washing, the cells may be resuspended in a variety of biocompatible buffers, such as, for example, Ca-free, Mg-free PBS. Alternatively, the undesirable components of the apheresis sample may be removed, and the cells directly resuspended in culture media.

[0598] In another embodiment, T cells are isolated from peripheral blood lymphocytes by lysing the red blood cells and depleting the monocytes, for example, by centrifugation through a PERCOLL™ gradient. A specific subpopulation of T cells, such as CD28⁺, CD4⁺, CD45RA⁺, and CD45RO⁺ T cells, can be further isolated by positive or negative selection techniques. For example, in one preferred embodiment, T cells are isolated by incubation with anti-CD3/anti-CD28 (i.e., 3×28)-conjugated beads, such as DYNABEADS® M-450 CD3/CD28 T, or XCYTE DYNABEADS™ for a time period sufficient for positive selection of the desired T cells. In one embodiment, the time period is about 30 minutes. In a further embodiment, the time period ranges from 30 minutes to 36 hours or longer and all integer values there between. In a further embodiment, the time period is at least 1, 2, 3, 4, 5, or 6 hours. In yet another preferred embodiment, the time period is 10 to 24 hours. In one preferred embodiment, the incubation time period is 24 hours. For isolation of T cells from patients with leukemia, use of longer incubation times, such as 24 hours, can increase cell yield. Longer incubation times may be used to isolate T cells in any situation where there are few T cells as compared to other cell types, such in isolating tumor infiltrating lymphocytes (TIL) from tumor tissue or from immunocompromised individuals. Further, use of longer incubation times can increase the efficiency of capture of CD8⁺ T cells.

[0599] Enrichment of a T cell population by negative selection can be accomplished with a combination of antibodies directed to surface markers unique to the negatively selected cells. A preferred method is cell sorting and/or selection via negative magnetic immunoadherence or flow cytometry that uses a cocktail of monoclonal antibodies directed to cell surface markers present on the cells negatively selected. For example, to enrich for CD4⁺ cells by negative selection, a monoclonal antibody cocktail typically includes antibodies to CD14, CD20, CD11b, CD16, HLA-DR, and CD8.

[0600] Further, monocyte populations (e.g., CD14⁺ cells) may be depleted from blood preparations by a variety of methodologies, including anti-CD14 coated beads or columns, or

utilization of the phagocytotic activity of these cells to facilitate removal. Accordingly, in one embodiment, the invention uses paramagnetic particles of a size sufficient to be engulfed by phagocytotic monocytes. In one embodiment, the paramagnetic particles are commercially available beads, for example, those produced by Life Technologies under the trade name Dynabeads™. In one embodiment, other non-specific cells are removed by coating the paramagnetic particles with “irrelevant” proteins (e.g., serum proteins or antibodies). Irrelevant proteins and antibodies include those proteins and antibodies or fragments thereof that do not specifically target the T cells to be isolated. In one embodiment, the irrelevant beads include beads coated with sheep anti-mouse antibodies, goat anti-mouse antibodies, and human serum albumin.

[0601] In brief, such depletion of monocytes is performed by preincubating T cells isolated from whole blood, apheresed peripheral blood, or tumors with one or more varieties of irrelevant or non-antibody coupled paramagnetic particles at any amount that allows for removal of monocytes (approximately a 20:1 bead:cell ratio) for about 30 minutes to 2 hours at 22 to 37 degrees C., followed by magnetic removal of cells which have attached to or engulfed the paramagnetic particles. Such separation can be performed using standard methods available in the art. For example, any magnetic separation methodology may be used including a variety of which are commercially available, (e.g., DYNAL® Magnetic Particle Concentrator (DYNAL MPC®)). Assurance of requisite depletion can be monitored by a variety of methodologies known to those of ordinary skill in the art, including flow cytometric analysis of CD14 positive cells, before and after depletion.

[0602] For isolation of a desired population of cells by positive or negative selection, the concentration of cells and surface (e.g., particles such as beads) can be varied. In one embodiment, it may be desirable to significantly decrease the volume in which beads and cells are mixed together (i.e., increase the concentration of cells), to ensure maximum contact of cells and beads. For example, in one embodiment, a concentration of 2 billion cells/ml is used. In one embodiment, a concentration of 1 billion cells/ml is used. In a further embodiment, greater than 100 million cells/ml is used. In a further embodiment, a concentration of cells of 10, 15, 20, 25, 30, 35, 40, 45, or 50 million cells/ml is used. In yet another embodiment, a concentration of cells from 75, 80, 85, 90, 95, or 100 million cells/ml is used. In further embodiments, concentrations of 125 or 150 million cells/ml can be used. Using high concentrations can result in increased cell yield, cell activation, and cell expansion. Further,

use of high cell concentrations allows more efficient capture of cells that may weakly express target antigens of interest, such as CD28-negative T cells, or from samples where there are many tumor cells present (i.e., leukemic blood, tumor tissue, etc.). Such populations of cells may have therapeutic value and would be desirable to obtain. For example, using high concentration of cells allows more efficient selection of CD8⁺ T cells that normally have weaker CD28 expression.

[0603] In a related embodiment, it may be desirable to use lower concentrations of cells. By significantly diluting the mixture of T cells and surface (e.g., particles such as beads), interactions between the particles and cells are minimized. This selects for cells that express high amounts of desired antigens to be bound to the particles. For example, CD4⁺ T cells express higher levels of CD28 and are more efficiently captured than CD8⁺ T cells in dilute concentrations. In one embodiment, the concentration of cells used is $5 \times 10^6/\text{ml}$. In other embodiments, the concentration used can be from about $1 \times 10^5/\text{ml}$ to $1 \times 10^6/\text{ml}$, and any integer value in between.

[0604] T cells can also be frozen. Wishing not to be bound by theory, the freeze and subsequent thaw step provides a more uniform product by removing granulocytes and to some extent monocytes in the cell population. After a washing step to remove plasma and platelets, the cells may be suspended in a freezing solution. While many freezing solutions and parameters are known in the art and will be useful in this context, one method involves using PBS containing 20% DMSO and 8% human serum albumin, or other suitable cell freezing media, the cells then are frozen to -80°C at a rate of 1° per minute and stored in the vapor phase of a liquid nitrogen storage tank. Other methods of controlled freezing may be used as well as uncontrolled freezing immediately at -20°C . or in liquid nitrogen.

[0605] T cells for use in the present invention may also be antigen-specific T cells. For example, tumor-specific T cells can be used. In one embodiment, antigen-specific T cells can be isolated from a patient of interest, such as a patient afflicted with a cancer or an infectious disease. In one embodiment, neoepitopes are determined for a subject, and T cells specific to these antigens are isolated. Antigen-specific cells for use in expansion may also be generated *in vitro* using any number of methods known in the art, for example, as described in U.S. Patent Publication No. US 20040224402 entitled, Generation and Isolation of Antigen-Specific T Cells, or in U.S. Pat. No. 6,040,177. Antigen-specific cells for use in the present invention may also be generated using any number of methods known in the art, for example, as described in

Current Protocols in Immunology, or Current Protocols in Cell Biology, both published by John Wiley & Sons, Inc., Boston, Mass.

[0606] In a related embodiment, it may be desirable to sort or otherwise positively select (e.g., via magnetic selection) the antigen specific cells prior to or following one or two rounds of expansion. Sorting or positively selecting antigen-specific cells can be carried out using peptide-MHC tetramers (Altman, et al., Science. 1996 Oct. 4; 274(5284):94-6). In another embodiment, the adaptable tetramer technology approach is used (Andersen et al., 2012 Nat Protoc. 7:891-902). Tetramers are limited by the need to utilize predicted binding peptides based on prior hypotheses, and the restriction to specific HLAs. Peptide-MHC tetramers can be generated using techniques known in the art and can be made with any MHC molecule of interest and any antigen of interest as disclosed herein. Specific epitopes to be used in this context can be identified using numerous assays known in the art. For example, the ability of a polypeptide to bind to MHC class I may be evaluated indirectly by monitoring the ability to promote incorporation of ¹²⁵I labeled β 2-microglobulin (β 2m) into MHC class I/ β 2m/peptide heterotrimeric complexes (*See* Parker et al., J. Immunol. 152:163, 1994).

[0607] In one embodiment cells are directly labeled with an epitope-specific reagent for isolation by flow cytometry followed by characterization of phenotype and TCRs. In one embodiment, T cells are isolated by contacting with T cell specific antibodies. Sorting of antigen-specific T cells, or generally any cells of the present invention, can be carried out using any of a variety of commercially available cell sorters, including, but not limited to, MoFlo sorter (DakoCytomation, Fort Collins, Colo.), FACS Aria™, FACSArray™, FACSVantage™, BD™ LSR II, and FACSCalibur™ (BD Biosciences, San Jose, Calif.).

[0608] In a preferred embodiment, the method comprises selecting cells that also express CD3. The method may comprise specifically selecting the cells in any suitable manner. Preferably, the selecting is carried out using flow cytometry. The flow cytometry may be carried out using any suitable method known in the art. The flow cytometry may employ any suitable antibodies and stains. Preferably, the antibody is chosen such that it specifically recognizes and binds to the particular biomarker being selected. For example, the specific selection of CD3, CD8, TIM-3, LAG-3, 4-1BB, or PD-1 may be carried out using anti-CD3, anti-CD8, anti-TIM-3, anti-LAG-3, anti-4-1BB, or anti-PD-1 antibodies, respectively. The antibody or antibodies may be conjugated to a bead (e.g., a magnetic bead) or to a fluorochrome. Preferably, the flow cytometry is fluorescence-activated cell sorting (FACS).

TCRs expressed on T cells can be selected based on reactivity to autologous tumors. Additionally, T cells that are reactive to tumors can be selected for based on markers using the methods described in patent publication Nos. WO2014133567 and WO2014133568, herein incorporated by reference in their entirety. Additionally, activated T cells can be selected for based on surface expression of CD107a.

[0609] In one embodiment of the invention, the method further comprises expanding the numbers of T cells in the enriched cell population. Such methods are described in U.S. Patent No. 8,637,307 and is herein incorporated by reference in its entirety. The numbers of T cells may be increased at least about 3-fold (or 4-, 5-, 6-, 7-, 8-, or 9-fold), more preferably at least about 10-fold (or 20-, 30-, 40-, 50-, 60-, 70-, 80-, or 90-fold), more preferably at least about 100-fold, more preferably at least about 1,000-fold, or most preferably at least about 100,000-fold. The numbers of T cells may be expanded using any suitable method known in the art. Exemplary methods of expanding the numbers of cells are described in patent publication No. WO 2003/057171, U.S. Patent No. 8,034,334, and U.S. Patent Publication No. 2012/0244133, each of which is incorporated herein by reference in their entireties and can be adapted for use with the engineered Type V Cas proteins disclosed herein.

[0610] In one embodiment, *ex vivo* T cell expansion can be performed by isolation of T cells and subsequent stimulation or activation followed by further expansion. In one embodiment of the invention, the T cells may be stimulated or activated by a single agent. In another embodiment, T cells are stimulated or activated with two agents, one that induces a primary signal and a second that is a co-stimulatory signal. Ligands useful for stimulating a single signal or stimulating a primary signal and an accessory molecule that stimulates a second signal may be used in soluble form. Ligands may be attached to the surface of a cell, to an Engineered Multivalent Signaling Platform (EMSP), or immobilized on a surface. In a preferred embodiment, both primary and secondary agents are co-immobilized on a surface, for example a bead or a cell. In one embodiment, the molecule providing the primary activation signal may be a CD3 ligand, and the co-stimulatory molecule may be a CD28 ligand or 4-1BB ligand.

[0611] In one embodiment, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in International Patent Publication No. WO 2015/120096, by a method comprising enriching a population of lymphocytes obtained from a donor subject; stimulating the population of lymphocytes with one or more T-cell stimulating agents to

produce a population of activated T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using a single cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells for a predetermined time to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. In one embodiment, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in WO 2015/120096, by a method comprising: obtaining a population of lymphocytes; stimulating the population of lymphocytes with one or more stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using at least one cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. The predetermined time for expanding the population of transduced T cells may be 3 days. The time from enriching the population of lymphocytes to producing the engineered T cells may be 6 days. The closed system may be a closed bag system. Further provided is population of T cells comprising a CAR or an exogenous TCR obtainable or obtained by said method, and a pharmaceutical composition comprising such cells.

[0612] In one embodiment, T cell maturation or differentiation *in vitro* may be delayed or inhibited by the method as described in International Patent Publication No. WO 2017/070395, comprising contacting one or more T cells from a subject in need of a T cell therapy with an AKT inhibitor (such as, e.g., one or a combination of two or more AKT inhibitors disclosed in claim 8 of WO2017070395) and at least one of exogenous Interleukin-7 (IL-7) and exogenous Interleukin-15 (IL-15), wherein the resulting T cells exhibit delayed maturation or differentiation, and/or wherein the resulting T cells exhibit improved T cell function (such as, e.g., increased T cell proliferation; increased cytokine production; and/or increased cytolytic activity) relative to a T cell function of a T cell cultured in the absence of an AKT inhibitor.

[0613] In one embodiment, a patient in need of a T cell therapy may be conditioned by a method as described in International Patent Publication No. WO 2016/191756 comprising administering to the patient a dose of cyclophosphamide between 200 mg/m²/day and 2000 mg/m²/day and a dose of fludarabine between 20 mg/m²/day and 900 mg/m²/day.

Further Exemplary Diseases and Disease Treatments

[0614] The following provides additional exemplary diseases for which the systems and compositions of the present invention can be used to treat and/or prevent, such as by a method disclosed herein.

Genetic Diseases and Diseases with a Genetic and/or Epigenetic Aspect

[0615] The compositions, systems, or components thereof can be used to treat and/or prevent a genetic disease or a disease with a genetic and/or epigenetic aspect. The genes and conditions exemplified herein are not exhaustive. In one embodiment, a method of treating and/or preventing a genetic disease can include administering a composition, system, and/or one or more components thereof to a subject, where the composition, system, and/or one or more components thereof is capable of modifying one or more copies of one or more genes associated with the genetic disease or a disease with a genetic and/or epigenetic aspect in one or more cells of the subject. In one embodiment, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can eliminate a genetic disease or a symptom thereof in the subject. In one embodiment, modifying one or more copies of one or more genes associated with a genetic disease or a disease with a genetic and/or epigenetic aspect in the subject can decrease the severity of a genetic disease or a symptom thereof in the subject. In one embodiment, the compositions, systems, or components thereof can modify one or more genes or polynucleotides associated with one or more diseases, including genetic diseases and/or those having a genetic aspect and/or epigenetic aspect, including but not limited to, any one or more set forth in **Table 5A**. It will be appreciated that those diseases and associated genes listed herein are non-exhaustive and non-limiting. Further some genes play roles in the development of multiple diseases.

[0616] Table 5A.

Exemplary Genetic and Other Diseases and Associated Genes			
Disease Name	Primary Tissues or System Affected	Additional Tissues/Systems Affected	Genes
Achondroplasia	Bone and Muscle		fibroblast growth factor receptor 3 (FGFR3)
Achromatopsia	eye		CNGA3, CNGB3, GNAT2, PDE6C, PDE6H, ACHM2, ACHM3,
Acute Renal Injury	kidney		NFkappaB, AATF, p85alpha, FAS, Apoptosis cascade elements (e.g. FASR, Caspase 2, 3, 4, 6, 7, 8, 9, 10, AKT, TNF alpha, IGF1, IGF1R, RIPK1), p53
Age Related Macular Degeneration	eye		Abcr; CCL2; CC2; CP (ceruloplasmin); Timp3; cathepsinD; VLDLR, CCR2
AIDS	Immune System		KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1
Albinism (including oculocutaneous albinism (types 1-7) and ocular albinism)	Skin, hair, eyes,		TYR, OCA2, TYRP1, and SLC45A2, SLC24A5 and C10orf11
Alkaptonuria	Metabolism of amino acids	Tissues/organs where homogentisic acid accumulates, particularly cartilage (joints), heart valves, kidneys	HGD
alpha-1 antitrypsin deficiency (AATD or A1AD)	Lung	Liver, skin, vascular system, kidneys, GI	SERPINA1, those set forth in WO2017165862, PiZ allele
ALS	CNS		SOD1; ALS2; ALS3; ALS5; ALS7; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c); DPP6; NEFH, PTGS1, SLC1A2, TNFRSF10B, PRPH, HSP90AA1, CRIA2, IFNG, AMPA2 S100B, FGF2, AOX1, CS, TXN, RAPHJ1, MAP3K5, NBEAL1, GPX1, ICA1L, RAC1, MAPT, ITPR2, ALS2CR4, GLS, ALS2CR8, CNTFR, ALS2CR11, FOLH1, FAM117B, P4HB, CNTF, SQSTM1, STRADB, NAIP, NLR, YWHAQ, SLC33A1, TRAK2, SCA1, NIF3L1, NIF3, PARD3B, COX8A, CDK15, HECW1, HECT, C2, WW 15, NOS1, MET, SOD2, HSPB1, NEFL, CTSB, ANG, HSPA8, RNase A, VAPB, VAMP, SNCA, alpha HGF, CAT, ACTB, NEFM, TH, BCL2, FAS, CASP3, CLU, SMN1, G6PD,

			BAX, HSF1, RNF19A, JUN, ALS2CR12, HSPA5, MAPK14, APEX1, TXNRD1, NOS2, TIMP1, CASP9, XIAP, GLG1, EPO, VEGFA, ELN, GDNF, NFE2L2, SLC6A3, HSPA4, APOE, PSMB8, DCTN2, TIMP3, KIFAP3, SLC1A1, SMN2, CCNC, STUB1, ALS2, PRDX6, SYP, CABIN1, CASP1, GART, CDK5, ATXN3, RTN4, C1QB, VEGFC, HTT, PARK7, XDH, GFAP, MAP2, CYCS, FCGR3B, CCS, UBL5, MMP9m SLC18A3, TRPM7, HSPB2, AKT1, DEERL1, CCL2, NGRN, GSR, TPPP3, APAF1, BTBD10, GLUD1, CXCR4, S:C1A3, FLT1, PON1, AR, LIF, ERBB3, :GA:S1, CD44, TP53, TLR3, GRIA1, GAPDH, AMPA, GRIK1, DES, CHAT, FLT4, CHMP2B, BAG1, CRNA COMPONENT4, GSS, BAK1, KDR, GSTP1, OGG1, IL6
Alzheimer's Disease	Brain		E1; CHIP; UCH; UBB; Tau; LRP; PICALM; CLU; PS1; SORL1; CR1; VLDLR; UBA1; UBA3; CHIP28; AQP1; UCHL1; UCHL3; APP, AAA, CVAP, AD1, APOE, AD2, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3, ALAS2, ABCA1, BIN1, BDNF, BTNL8, C1ORF49, CDH4, CHRN2, CKLFSF2, CLEC4E, CRIL, CSF3R, CST3, CYP2C, DAPK1, ESR1, FCAR, FCGR3B, FFA2, FGA, GAB2, GALP, GAPDHS, GMPB, HP, HTR7, IDE, IFI127, IFI6, IFIT2, IL1RN, IL-1RA, IL8RA, IL8RB, JAG1, KCNJ15, LRP6, MAPT, MARK4, MPHOSPH1, MTHFR, NBN, NCSTN, NIACR2, NMNAT3, NTM, ORM1, P2RY13, PBEF1, PCK1, PICALM, PLA2, PLXNC1, PRNP, PSEN1, PSEN2, PTPRA, RALGPS2, RGS2, SELENBP1, SLC25A37, SORL1, Mitoferrin-1, TF, TFAM, TNF, TNFRSF10C, UBE1C
Amyloidosis			APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB
Amyloid neuropathy			TTR, PALB
Anemia	Blood		CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT
Angelman Syndrome	Nervous system, brain		UBE3A
Attention Deficit Hyperactivity Disorder (ADHD)	Brain		PTCHD1

Autoimmune lymphoproliferative syndrome	Immune system		TNFRSF6, APT1, FAS, CD95, ALPS1A
Autism, Autism spectrum disorders (ASDs), including Asperger's and a general diagnostic category called Pervasive Developmental Disorders (PDDs)	Brain		PTCHD1; Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; GLO1, RTT, PPMX, MRX16, RX79, NLGN3, NLGN4, KIAA1260, AUTSX2, FMR1, FMR2; FXR1; FXR2; MGLUR5, ATP10C, CDH10, GRM6, MGLUR6, CDH9, CNTN4, NLGN2, CNTNAP2, SEMA5A, DHCR7, NLGN4X, NLGN4Y, DPP6, NLGN5, EN2, NRCAM, MDGA2, NRXN1, FMR2, AFF2, FOXP2, OR4M2, OXTR, FXR1, FXR2, PAH, GABRA1, PTEN, GABRA5, PTPRZ1, GABRB3, GABRG1, HIRIP3, SEZ6L2, HOXA1, SHANK3, IL6, SHBZRAP1, LAMB1, SLC6A4, SERT, MAPK3, TAS2R1, MAZ, TSC1, MDGA2, TSC2, MECP2, UBE3A, WNT2, See also 20110023145
autosomal dominant polycystic kidney disease (ADPKD) - (includes diseases such as von Hippel-Lindau disease and tubereous sclerosis complex disease)	kidney	liver	PKD1, PKD2
Autosomal Recessive Polycystic Kidney Disease (ARPKD)	kidney	liver	PKDH1
Ataxia-Telangiectasia (a.k.a Louis Bar syndrome)	Nervous system, immune system	various	ATM
B-Cell Non-Hodgkin Lymphoma			BCL7A, BCL7
Bardet-Biedl syndrome	Eye, musculoskeletal system, kidney, reproductive organs	Liver, ear, gastrointestinal system, brain	ARL6, BBS1, BBS2, BBS4, BBS5, BBS7, BBS9, BBS10, BBS12, CEP290, INPP5E, LZTFL1, MKKS, MKS1, SDCCAG8, TRIM32, TTC8
Bare Lymphocyte Syndrome	blood		TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5
Bartter's Syndrome (types I, II, III, IVA and B, and V)	kidney		SLC12A1 (type I), KCNJ1 (type II), CLCNKB (type III), BSND (type IV A), or both the CLCNKA CLCNKB genes (type IV B), CASR (type V).
Becker muscular dystrophy	Muscle		DMD, BMD, MYF6
Best Disease (Vitelliform Macular Dystrophy type 2)	eye		VMD2
Bleeding Disorders	blood		TBXA2R, P2RX1, P2X1
Blue Cone Monochromacy	eye		OPN1LW, OPN1MW, and LCR
Breast Cancer	Breast tissue		BRCA1, BRCA2, COX-2
Bruton's Disease (aka X-linked Agammaglobulinemia)	Immune system, specifically B cells		BTK
Cancers (e.g., lymphoma, chronic lymphocytic leukemia (CLL), B cell acute lymphocytic leukemia	Various		FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC, TRBC, those described in WO2015048577

(B-ALL), acute lymphoblastic leukemia, acute myeloid leukemia, non-Hodgkin's lymphoma (NHL), diffuse large cell lymphoma (DLCL), multiple myeloma, renal cell carcinoma (RCC), neuroblastoma, colorectal cancer, breast cancer, ovarian cancer, melanoma, sarcoma, prostate cancer, lung cancer, esophageal cancer, hepatocellular carcinoma, pancreatic cancer, astrocytoma, mesothelioma, head and neck cancer, and medulloblastoma			
Cardiovascular Diseases	heart	Vascular system	IL1B, XDH, TP53, PTGS, MB, IL4, ANGPT1, ABCG8, CTSK, PTGIR, KCNJ11, INS, CRP, PDGFRB, CCNA2, PDGFB, KCNJ5, KCNN3, CAPN10, ADRA2B, ABCG5, PRDX2, CPAN5, PARP14, MEX3C, ACE, RNF, IL6, TNF, STN, SERPINE1, ALB, ADIPOQ, APOB, APOE, LEP, MTHFR, APOA1, EDN1, NPPB, NOS3, PPARG, PLAT, PTGS2, CETP, AGTR1, HMGCR, IGF1, SELE, REN, PPARA, PON1, KNG1, CCL2, LPL, VWF, F2, ICAM1, TGFB, NPPA, IL10, EPO, SOD1, VCAM1, IFNG, LPA, MPO, ESR1, MAPK, HP, F3, CST3, COG2, MMP9, SERPINC1, F8, HMOX1, APOC3, IL8, PROL1, CBS, NOS2, TLR4, SELP, ABCA1, AGT, LDLR, GPT, VEGFA, NR3C2, IL18, NOS1, NR3C1, FGB, HGF, IL1A, AKT1, LIPC, HSPD1, MAPK14, SPPI, ITGB3, CAT, UTS2, THBD, F10, CP, TNFRSF11B, EGFR, MMP2, PLG, NPY, RHOD, MAPK8, MYC, FN1, CMA1, PLA2, GNB3, ADRB2, SOD2, F5, VDR, ALOX5, HLA-DRB1, PARP1, CD40LG, PON2, AGER, IRS1, PTGS1, ECE1, F7, IRMN, EPHX2, IGFBP1, MAPK10, FAS, ABCB1, JUN, IGFBP3, CD14, PDE5A, AGTR2, CD40, LCAT, CCR5, MMP1, TIMP1, ADM, DYT10, STAT3, MMP3, ELN, USF1, CFH, HSPA4, MMP12, MME, F2R, SELL, CTSB, ANXA5, ADRB1, CYBA, FGA, GGT1, LIPG, HIF1A, CXCR4, PROC, SCARB1, CD79A, PLTP, ADD1, FGG, SAA1, KCNH2, DPP4, NPR1, VTN, KIAA0101, FOS, TLR2, PPIG, IL1R1, AR, CYP1A1, SERPINA1, MTR, RBP4, APOA4, CDKN2A, FGF2, EDNRB, ITGA2,

		VLA-2, CABIN1, SHBG, HMGB1, HSP90B2P, CYP3A4, GJA1, CAV1, ESR2, LTA, GDF15, BDNF, CYP2D6, NGF, SP1, TGIF1, SRC, EGF, PIK3CG, HLA-A, KCNQ1, CNR1, FBN1, CHKA, BEST1, CTNNB1, IL2, CD36, PRKAB1, TPO, ALDH7A1, CX3CR1, TH, F9, CH1, TF, HFE, IL17A, PTEN, GSTM1, DMD, GATA4, F13A1, TTR, FABP4, PON3, APOC1, INSR, TNFRSF1B, HTR2A, CSF3, CYP2C9, TXN, CYP11B2, PTH, CSF2, KDR, PLA2G2A, THBS1, GCG, RHOA, ALDH2, TCF7L2, NFE2L2, NOTCH1, UGT1A1, IFNA1, PPARG, SIRT1, GNHR1, PAPP, ARR3, NPPC, AHSP, PTK2, IL13, MTOR, ITGB2, GSTT1, IL6ST, CPB2, CYP1A2, HNF4A, SLC64A, PLA2G6, TNFSF11, SLC8A1, F2RL1, AKR1A1, ALDH9A1, BGLAP, MTTP, MTRR, SULT1A3, RAGE, C4B, P2RY12, RNLS, CREB1, POMC, RAC1, LMNA, CD59, SCM5A, CYP1B1, MIF, MMP13, TIMP2, CYP19A1, CUP21A2, PTPN22, MYH14, MBL2, SELPLG, AOC3, CTSL1, PCNA, IGF2, ITGB1, CAST, CXCL12, IGHE, KCNE1, TFRC, COL1A1, COL1A2, IL2RB, PLA2G10, ANGPT2, PROCR, NOX4, HAMP, PTPN11, SLCA1, IL2RA, CCL5, IRF1, CF:AR, CA:CA, EIF4E, GSTP1, JAK2, CYP3A5, HSPG2, CCL3, MYD88, VIP, SOAT1, ADRBK1, NR4A2, MMP8, NPR2, GCH1, EPRS, PPARGC1A, F12, PECAM1, CCL4, CERPINA34, CASR, FABP2, TTF2, PROS1, CTF1, SGCB, YME1L1, CAMP, ZC3H12A, AKR1B1, MMP7, AHR, CSF1, HDAC9, CTGF, KCNMA1, UGT1A, PRKCA, COMT, S100B, EGR1, PRL, IL15, DRD4, CAMK2G, SLC22A2, CCL11, PGF, THPO, GP6, TACR1, NTS, HNF1A, SST, KCDN1, LOC646627, TBXAS1, CUP2J2, TBXA2R, ADH1C, ALOX12, AHSG, BHMT, GJA4, SLC25A4, ACLY, ALOX5AP, NUMA1, CYP27B1, CYSLTR2, SOD3, LTC4S, UCN, GHRL, APOC2, CLEC4A, KBTBD10, TNC, TYMS, SHC1, LRP1, SOCS3, ADH1B, KLK3, HSD11B1, VKORC1, SERPINB2, TNSI, RNF19A, EPOR, ITGAM, PITX2, MAPK7, FCGR3A, LEEPR,
--	--	---

			ENG, GPX1, GOT2, HRH1, NR112, CRH, HTR1A, VDAC1, HPSE, SFTPD, TAP2, RMF123, PTK2Bm NTRK2, IL6R, ACHE, GLP1R, GHR, GSR, NQO1, NR5A1, GJB2, SLC9A1, MAOA, PCSK9, FCGR2A, SERPINF1, EDN3, UCP2, TFAP2A, C4BPA, SERPINF2, TYMP, ALPP, CXCR2, SLC3A3, ABCG2, ADA, JAK3, HSPA1A, FASN, FGF1, F11, ATP7A, CR1, GFPA, ROCK1, MECP2, MYLK, BCHE, LIPE, ADORA1, WRN, CXCR3, CD81, SMAD7, LAMC2, MAP3K5, CHGA, IAPP, RHO, ENPP1, PTHLH, NRG1, VEGFC, ENPEP, CEBPB, NAGLU, F2RL3, CX3CL1, BDKRB1, ADAMTS13, ELANE, ENPP2, CISH, GAST, MYOC, ATP1A2, NF1, GJB1, MEF2A, VCL, BMPR2, TUBB, CDC42, KRT18, HSF1, MYB, PRKAA2, ROCK2, TFP1, PRKG1, BMP2, CTNND1, CTH, CTSS, VAV2, NPY2R, IGFBP2, CD28, GSTA1, PPIA, APOH, S100A8, IL11, ALOX15, FBLN1, NR1H3, SCD, GIP, CHGB, PRKCB, SRD5A1, HSD11B2, CALCRL, GALNT2, ANGPTL4, KCNN4, PIK3C2A, HBEGF, CYP7A1, HLA-DRB5, BNIP3, GCKR, S100A12, PADI4, HSPA14, CXCR1, H19, KRTAP19-3, IDDM2, RAC2, YRY1, CLOCK, NGFR, DBH, CRNA COMPONENT4, CACNA1C, PRKAG2, CHAT, PTGDS, NR1H2, TEK, VEGFB, MEF2C, MAPKAPK2, TNFRSF11A, HSPA9, CYSLTR1, MAT1A, OPR1, IMPA1, CLCN2, DLD, PSMA6, PSMB8, CHI3L1, ALDH1B1, PARP2, STAR, LBP, ABCC6, RGS2, EFNB2, GJB6, APOA2, AMPD1, DYSL, FDFT1, EMD2, CCR6, GJB3, IL1RL1, ENTPD1, BBS4, CELSR2, F11R, RAPGEF3, HYAL1, ZNF259, ATOX1, ATF6, KHK, SAT1, GGH, TIMP4, SLC4A4, PDE2A, PDE3B, FADS1, FADS2, TMSB4X, TXNIP, LIMS1, RHOB, LY96, FOXO1, PNPLA2, TRH, GJC1, SLC17A5, FTO, GJD2, PRSC1, CASP12, GPBAR1, PXX, IL33, TRIB1, PBX4, NUPR1, 15-SEP, CILP2, TERC, GGT2, MTCO1, UOX, AVP
Cataract	eye		CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3,

			CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1
CDKL-5 Deficiencies or Mediated Diseases	Brain, CNS		CDKL5
Charcot-Marie-Tooth (CMT) disease (Types 1, 2, 3, 4,)	Nervous system	Muscles (dystrophy)	PMP22 (CMT1A and E), MPZ (CMT1B), LITAF (CMT1C), EGR2 (CMT1D), NEFL (CMT1F), GJB1 (CMT1X), MFN2 (CMT2A), KIF1B (CMT2A2B), RAB7A (CMT2B), TRPV4 (CMT2C), GARS (CMT2D), NEFL (CMT2E), GAPD1 (CMT2K), HSPB8 (CMT2L), DYNC1H1, CMT2O), LRSAM1 (CMT2P), IGHMBP2 (CMT2S), MORC2 (CMT2Z), GDAP1 (CMT4A), MTMR2 or SBF2/MTMR13 (CMT4B), SH3TC2 (CMT4C), NDRG1 (CMT4D), PRX (CMT4F), FIG4 (CMT4J), NT-3
Chédiak-Higashi Syndrome	Immune system	Skin, hair, eyes, neurons	LYST
Choroideremia			CHM, REP1,
Chorioretinal atrophy	eye		PRDM13, RGR, TEAD1
Chronic Granulomatous Disease	Immune system		CYBA, CYBB, NCF1, NCF2, NCF4
Chronic Mucocutaneous Candidiasis	Immune system		AIRE, CARD9, CLEC7A IL12B, IL12B1, IL1F, IL17RA, IL17RC, RORC, STAT1, STAT3, TRAF31P2
Cirrhosis	liver		KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988
Colon cancer (Familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC))	Gastrointestinal		FAP: APC HNPCC: MSH2, MLH1, PMS2, SH6, PMS1
Combined Immunodeficiency	Immune System		IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228
Cone(-rod) dystrophy	eye		AIPL1, CRX, GUA1A, GUCY2D, PITPM3, PROM1, PRPH2, RIMS1, SEMA4A, ABCA4, ADAM9, ATF6, C21ORF2, C8ORF37, CACNA2D4, CDHR1, CERKL, CNGA3, CNGB3, CNM4, CNAT2, IFT81, KCNV2, PDE6C, PDE6H, POC1B, RAX2, RDH5, RPGRIP1, TTLL5, RetCG1, GUCY2E
Congenital Stationary Night Blindness	eye		CABP4, CACNA1F, CACNA2D4, GNAT1, CPR179, GRK1, GRM6, LRIT3, NYX, PDE6B, RDH5, RHO, RLBP1, RPE65, SAG, SLC24A1, TRPM1,
Congenital Fructose Intolerance	Metabolism		ALDOB

Cori's Disease (Glycogen Storage Disease Type III)	Various- wherever glycogen accumulates, particularly liver, heart, skeletal muscle		AGL
Corneal clouding and dystrophy	eye		APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD
Cornea plana congenital			KERA, CNA2
Cri du chat Syndrome, also known as 5p syndrome and cat cry syndrome			Deletions involving only band 5p15.2 to the entire short arm of chromosome 5, e.g. CTNND2, TERT,
Cystic Fibrosis (CF)	Lungs and respiratory system	Pancreas, liver, digestive system, reproductive system, exocrine, glands,	CTFR, ABCC7, CF, MRP7, SCN1A, those described in WO2015157070
Diabetic nephropathy	kidney		Gremlin, 12/15- lipoxigenase, TIM44,
Dent Disease (Types 1 and 2)	Kidney		Type 1: CLCN5, Type 2: ORCL
Dentatorubro-Pallidolusian Atrophy (DRPLA) (aka Haw River and Naito-Oyanagi Disease)	CNS, brain, muscle		Atrophin-1 and Atn1
Down Syndrome	various		Chromosome 21 trisomy
Drug Addiction	Brain		Prkce; Drd2; Drd4; ABAT; GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1
Duane syndrome (Types 1, 2, and 3, including subgroups A, B and C). Other names for this condition include: Duane's Retraction Syndrome (or DR syndrome), Eye Retraction Syndrome, Retraction Syndrome, Congenital retraction syndrome and Stilling-Turk-Duane Syndrome	eye		CHN1, indels on chromosomes 4 and 8
Duchenne muscular dystrophy (DMD)	muscle	Cardiovascular, respiratory	DMD, BMD, dystrophin gene, intron flanking exon 51 of DMD gene, exon 51 mutations in DMD gene, See also WO2013163628 and US Pat. Pub. 20130145487
Edward's Syndrome (Trisomy 18)			Complete or partial trisomy of chromosome 18
Ehlers-Danlos Syndrome (Types I-VI)	Various depending on type: including musculoskeletal, eye, vasculature, immune, and skin		COL5A1, COL5A2, COL1A1, COL3A1, TNXB, PLOD1, COL1A2, FKBP14 and ADAMTS2

Emery-Dreifuss muscular dystrophy	muscle		LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A
Enhanced S-Cone Syndrome	eye		NR2E3, NRL
Fabry's Disease	Various – including skin, eyes, and gastrointestinal system, kidney, heart, brain, nervous system		GLA
Facioscapulohumeral muscular dystrophy	muscles		FSHMD1A, FSHD1A, FRG1,
Factor H and Factor H-like 1	blood		HF1, CFH, HUS
Factor V Leiden thrombophilia and Factor V deficiency	blood		Factor V (F5)
Factor V and Factor VII deficiency	blood		MCFD2
Factor VII deficiency	blood		F7
Factor X deficiency	blood		F10
Factor XI deficiency	blood		F11
Factor XII deficiency	blood		F12, HAF
Factor XIIIa deficiency	blood		F13A1, F13A
Factor XIIIb deficiency	blood		F13B
Familial Hypercholesterolemia	Cardiovascular system		APOB, LDLR, PCSK9
Familial Mediterranean Fever (FMF) also called recurrent polyserositis or familial paroxysmal polyserositis	Various- organs/tissues with serous or synovial membranes, skin, joints	Heart, kidney, brain/CNS, reproductive organs	MEFV
Fanconi Anemia	Various – blood (anemia), immune system, cognitive, kidneys, eyes, musculoskeletal		FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCC, FANCG, RAD51, BRCA1, BRCA2, BRIP1, BACH1, FANCI, FANCB, FANCD1, FANCD2, FANCD, FAD, FANCE, FACE, FANCF, FANCI, ERCC4, FANCL, FANCM, PALB2, RAD51C, SLX4, UBE2T, FANCB, XRCC9, PHF9, KIAA1596
Fanconi Syndrome Types I (Childhood onset) and II (Adult Onset)	kidneys		FRTS1, GATM
Fragile X syndrome and related disorders	brain		FMR1, FMR2; FXR1; FXR2; mGLUR5
Fragile XE Mental Retardation (aka Martin Bell syndrome)	Brain, nervous system		FMR1
Friedreich Ataxia (FRDA)	Brain, nervous system	heart	FXN/X25
Fuchs endothelial corneal dystrophy	Eye		TCF4; COL8A2
Galactosemia	Carbohydrate metabolism disorder	Various-where galactose accumulates – liver, brain, eyes	GALT, GALK1, and GALE

Gastrointestinal Epithelial Cancer, GI cancer			CISH
Gaucher Disease (Types 1, 2, and 3, as well as other unusual forms that may not fit into these types)	Fat metabolism disorder	Various-liver, spleen, blood, CNS, skeletal system	GBA
Griscelli syndrome			
Glaucoma	eye		MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A, those described in WO2015153780
Glomerulo sclerosis	kidney		CC chemokine ligand 2
Glycogen Storage Diseases Types I-VI -See also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease	Metabolism Diseases		SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM, See also Cori's Disease, Pompe's Disease, McArdle's disease, Hers Disease, and Von Gierke's disease
RBC Glycolytic enzyme deficiency	blood		any mutations in a gene for an enzyme in the glycolysis pathway including mutations in genes for hexokinases I and II, glucokinase, phosphoglucose isomerase, phosphofructokinase, aldolase Bm triosephosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerokinase, phosphoglycerate mutase, enolase I, pyruvate kinase
Hartnup's disease	Malabsorption disease	Various- brain, gastrointestinal, skin,	SLC6A19
Hearing Loss	ear		NOX3, Hes5, BDNF,
Hemochromatosis (HH)	Iron absorption regulation disease	Various- wherever iron accumulates, liver, heart, pancreas, joints, pituitary gland	HFE and H63D
Hemophagocytic lymphohistiocytosis disorders	blood		PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3
Hemorrhagic disorders	blood		PI, ATT, F5
Hers disease (Glycogen storage disease Type VI)	liver	muscle	PYGL
Hereditary angioedema (HAE)			kalikrein B1
Hereditary Hemorrhagic Telangiectasia (Osler-Weber-Rendu Syndrome)	Skin and mucous membranes		ACVRL1, ENG and SMAD4
Hereditary Spherocytosis	blood		NK1, EPB42, SLC4A1, SPTA1, and SPTB
Hereditary Persistence of Fetal Hemoglobin	blood		HBG1, HBG2, BCL11A, promoter region of HBG 1 and/or 2 (in the CCAAT box)

Hemophilia (hemophilia A (Classic) a B (aka Christmas disease) and C)	blood		A: FVIII, F8C, HEMA B: FVIX, HEMB C: F9, F11
Hepatic adenoma	liver		TCF1, HNF1A, MODY3
Hepatic failure, early onset, and neurologic disorder	liver		SCOD1, SCO1
Hepatic lipase deficiency	liver		LIPC
Hepatoblastoma, cancer and carcinomas	liver		CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5
Hermansky-Pudlak syndrome	Skin, eyes, blood, lung, kidneys, intestine		HPS1, HPS3, HPS4, HPS5, HPS6, HPS7, DTNBP1, BLOC1, BLOC1S2, BLOC3
HIV susceptibility or infection	Immune system		IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCR5 (CCR5), those in WO2015148670A1
Holoprosencephaly (HPE) (Alobar, Semilobar, and Lobar)	brain		ACVRL1, ENG, SMAD4
Homocystinuria	Metabolic disease	Various-connective tissue, muscles, CNS, cardiovascular system	CBS, MTHFR, MTR, MTRR, and MMADHC
HPV			HPV16 and HPV18 E6/E7
HSV1, HSV2, and related keratitis	eye		HSV1 genes (immediate early and late HSV-1 genes (UL1, 1.5, 5, 6, 8, 9, 12, 15, 16, 18, 19, 22, 23, 26, 26.5, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 42, 48, 49.5, 50, 52, 54, S6, RL2, RS1, those described in WO2015153789, WO2015153791
Hunter's Syndrome (aka Mucopolysaccharidosis type II)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDS
Huntington's disease (HD) and HD-like disorders	Brain, nervous system		HD, HTT, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17, PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2, and those described in WO2013130824, WO2015089354
Hurler's Syndrome (aka mucopolysaccharidosis type I H, MPS IH)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α -L-iduronidase
Hurler-Scheie syndrome (aka mucopolysaccharidosis type I H-S, MPS I H-S)	Lysosomal storage disease	Various- liver, spleen, eye, joint, heart, brain, skeletal	IDUA, α -L-iduronidase
hyaluronidase deficiency (aka MPS IX)	Soft and connective tissues		HYAL1
Hyper IgM syndrome	Immune system		CD40L

Hyper- tension caused renal damage	kidney		Mineral corticoid receptor
Immunodeficiencies	Immune System		CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI
Inborn errors of metabolism: including urea cycle disorders, organic acidemias, fatty acid oxidation defects, amino acidopathies, carbohydrate disorders, mitochondrial disorders	Metabolism diseases, liver	Various organs and cells	See also: Carbohydrate metabolism disorders (e.g. galactosemia), Amino acid Metabolism disorders (e.g. phenylketonuria), Fatty acid metabolism (e.g. MCAD deficiency), Urea Cycle disorders (e.g. Citrullinemia), Organic acidemias (e.g. Maple Syrup Urine disease), Mitochondrial disorders (e.g. MELAS), peroxisomal disorders (e.g. Zellweger syndrome)
Inflammation	Various		IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
Inflammatory Bowel Diseases (e.g. Ulcerative Colitis and Chron's Disease)	Gastrointestinal	Joints, skin	NOD2, IRGM, LRRK2, ATG5, ATG16L1, IRGM, GATM, ECM1, CDH1, LAMB1, HNF4A, GNA12, IL10, CARD9/15, CCR6, IL2RA, MST1, TNFSF15, REL, STAT3, IL23R, IL12B, FUT2
Interstitial renal fibrosis	kidney		TGF- β type II receptor
Job's Syndrome (aka Hyper IgE Syndrome)	Immune System		STAT3, DOCK8
Juvenile Retinoschisis	eye		RS1, XLR1
Kabuki Syndrome 1			MLL4, KMT2D
Kennedy Disease (aka Spinobulbar Muscular Atrophy)	Muscles, brain, nervous system		SBMA/SMAX1/AR
Klinefelter syndrome	Various- particularly those involved in development of male characteristics		Extra X chromosome in males
Lafora Disease	Brain, CNS		EMP2A and EMP2B
Leber Congenital Amaurosis	eye		CRB1, RP12, CORD2, CRD, CRX, IMPDH1, OTX2, AIPL1, CABP4, CCT2, CEP290, CLUAP1, CRB1, CRX, DTHD1, GDF6, GUCY2D, IFT140, IQCB1, KCNJ13, LCA5, LRAT, NMNAT1, PRPH2, RD3, RDH12, RPE65, RP20, RPGRIP1, SPATA7, TULP1, LCA1, LCA4, GUC2D, CORD6, LCA3,
Lesch-Nyhan Syndrome	Metabolism disease	Various – joints, cognitive, brain, nervous system	HPRT1

Leukocyte deficiencies and disorders	blood		ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4
Leukemia	Blood		TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN
Limb-girdle muscular dystrophy diseases	muscle		LGMD
Lowe syndrome	brain, eyes, kidneys		OCRL
Lupus glomerulo- nephritis	kidney		MAPK1
Machado-Joseph's Disease (also known as Spinocerebellar ataxia Type 3)	Brain, CNS, muscle		ATX3
Macular degeneration	eye		ABC4, CBC1, CHM1, APOE, C1QTNF5, C2, C3, CCL2, CCR2, CD36, CFB, CFH, CFHR1, CFHR3, CNGB3, CP, CRP, CST3, CTSD, CX3CR1, ELOVL4, ERCC6, FBLN5, FBLN6, FSCN2, HMCN1, HTRA1, IL6, IL8, PLEKHA1, PROM1, PRPH2, RPGR, SERPING1, TCOF1, TIMP3, TLR3
Macular Dystrophy	eye		BEST1, C1QTNF5, CTNNA1, EFEMP1, ELOVL4, FSCN2, GUCA1B, HMCN1, IMPG1, OTX2, PRDM13, PROM1, PRPH2, RP1L1, TIMP3, ABCA4, CFH, DRAM2, IMG1, MFSD8, ADMD, STGD2, STGD3, RDS, RP7, PRPH, AVMD, AOFMD, VMD2
Malattia Leventinese	eye		EFEMP1, FBLN3
Maple Syrup Urine Disease	Metabolism disease		BCKDHA, BCKDHB, and DBT
Marfan syndrome	Connective tissue	Musculoskeletal	FBN1
Maroteaux-Lamy Syndrome (aka MPS VI)	Musculoskeletal system, nervous system	Liver, spleen	ARSB

McArdle's Disease (Glycogen Storage Disease Type V)	Glycogen storage disease	muscle	PYGM
Medullary cystic kidney disease	kidney		UMOD, HNFJ, FJHN, MCKD2, ADMCKD2
Metachromatic leukodystrophy	Lysosomal storage disease	Nervous system	ARSA
Methylmalonic acidemia (MMA)	Metabolism disease		MMAA, MMAB, MUT, MMACHC, MMADHC, LMBRD1
Morquio Syndrome (aka MPS IV A and B)	Connective tissue, skin, bone, eyes	heart	GALNS
Mucopolysaccharidosis diseases (Types I H/S, I H, II, III A B and C, I S, IVA and B, IX, VII, and VI)	Lysosomal storage disease – affects various organs/tissues		See also Hurler/Scheie syndrome, Hurler disease, Sanfillipo syndrome, Scheie syndrome, Morquio syndrome, hyaluronidase deficiency, Sly syndrome, and Maroteaux-Lamy syndrome
Muscular Atrophy	muscle		VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1
Muscular dystrophy	muscle		FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1
Myotonic dystrophy (Type 1 and Type 2)	Muscles	Eyes, heart, endocrine	CNBP (Type 2) and DMPK (Type 1)
Neoplasia			PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR gamma; WT1 (Wilms Tumor); FGF Receptor Family members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB (retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR (Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4 variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor; Bax; Bcl2; caspases family (9 members: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc

Neurofibromatosis (NF) (NF1, formerly Recklinghausen's NF, and NF2)	brain, spinal cord, nerves, and skin		NF1, NF2
Niemann-Pick Lipidosis (Types A, B, and C)	Lysosomal Storage Disease	Various- where sphingomyelin accumulates, particularly spleen, liver, blood, CNS	Types A and B: SMPD1; Type C: NPC1 or NPC2
Noonan Syndrome	Various – musculoskeletal, heart, eyes, reproductive organs, blood		PTPN11, SOS1, RAF1 and KRAS
Norrie Disease or X-linked Familial Exudative Vitreoretinopathy	eye		NDP
North Carolina Macular Dystrophy	eye		MCDR1
Osteogenesis imperfecta (OI) (Types I, II, III, IV, V, VI, VII)	bones, musculoskeletal		COL1A1, COL1A2, CRTAP, P3H
Osteopetrosis	bones		LRP5, BMND1, LRP7, LR3, OPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1
Patau's Syndrome (Trisomy 13)	Brain, heart, skeletal system		Additional copy of chromosome 13
Parkinson's disease (PD)	Brain, nervous system		SNCA (PARK1), UCHL1 (PARK 5), and LRRK2 (PARK8), (PARK3), PARK2, PARK4, PARK7 (PARK7), PINK1 (PARK6); x-Synuclein, DJ-1, Parkin, NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, NCAP, PRKN, PDJ, DBH, NDUFV2
Pattern Dystrophy of the RPE	eye		RDS/peripherin
Phenylketonuria (PKU)	Metabolism disorder	Various due to build-up of phenylalanine, phenyl ketones in tissues and CNS	PAH, PKU1, QDPR, DHPR, PTS
Polycystic kidney and hepatic disease	Kidney, liver		FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63
Pompe's Disease	Glycogen storage disease	Various – heart, liver, spleen	GAA
Porphyria (actually refers to a group of different diseases all having a specific heme production process abnormality)	Various- wherever heme precursors accumulate		ALAD, ALAS2, CPOX, FECH, HMBS, PPOX, UROD, or UROS
posterior polymorphous corneal dystrophy	eyes		TCF4; COL8A2
Primary Hyperoxaluria (e.g. type 1)	Various – eyes, heart, kidneys, skeletal system		LDHA (lactate dehydrogenase A) and hydroxyacid oxidase 1 (HAO1)
Primary Open Angle Glaucoma (POAG)	eyes		MYOC

Primary sclerosing cholangitis	Liver, gallbladder		TCF4; COL8A2
Progeria (also called Hutchinson-Gilford progeria syndrome)	All		LMNA
Prader-Willi Syndrome	Musculoskeletal system, brain, reproductive and endocrine system		Deletion of region of short arm of chromosome 15, including UBE3A
Prostate Cancer	prostate		HOXB13, MSMB, GPRC6A, TP53
Pyruvate Dehydrogenase Deficiency	Brain, nervous system		PDHA1
Kidney/Renal carcinoma	kidney		RLIP76, VEGF
Rett Syndrome	Brain		MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1
Retinitis pigmentosa (RP)	eye		ADIPOR1, ABCA4, AGBL5, ARHGEF18, ARL2BP, ARL3, ARL6, BEST1, BBS1, BBS2, C2ORF71, C8ORF37, CA4, CERKL, CLRN1, CNGA1, CMGB1, CRB1, CRX, CYP4V2, DHDDS, DHX38, EMC1, EYS, FAM161A, FSCN2, GPR125, GUCA1B, HK1, HPRPF3, HGSNAT, IDH3B, IMPDH1, IMPG2, IFT140, IFT172, KLHL7, KIAA1549, KIZ, LRAT, MAK, MERTK, MVK, NEK2, NUROD1, NR2E3, NRL, OFD1, PDE6A, PDE6B, PDE6G, POMGNT1, PRCD, PROM1, PRPF3, PRPF4, PRPF6, PRPF8, PRPF31, PRPH2, RPB3, RDH12, REEP6, RP39, RGR, RHO, RLBP1, ROM1, RP1, RP1L1, RPY, RP2, RP9, RPE65, RPGR, SAMD11, SAG, SEMA4A, SLC7A14, SNRNP200, SPP2, SPATA7, TRNT1, TOPORS, TTC8, TULP1, USH2A, ZFN408, ZNF513, See also 20120204282
Scheie syndrome (also known as mucopolysaccharidosis type I S(MPS I-S))	Various- liver, spleen, eye, joint, heart, brain, skeletal		IDUA, α -L-iduronidase
Schizophrenia	Brain		Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin); Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2 Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a; GSK3b; 5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA; DTNBP1; Dao (Dao1); TCF4; COL8A2
Secretase Related Disorders	Various		APH-1 (alpha and beta); PSEN1; NCSTN; PEN-2; Nos1, Parp1, Nat1, Nat2, CTSB, APP, APH1B, PSEN2,

			PSENEN, BACE1, ITM2B, CTSD, NOTCH1, TNF, INS, DYT10, ADAM17, APOE, ACE, STN, TP53, IL6, NGFR, IL1B, ACHE, CTNNB1, IGF1, IFNG, NRG1, CASP3, MAPK1, CDH1, APBB1, HMGCR, CREB1, PTGS2, HES1, CAT, TGFB1, ENO2, ERBB4, TRAPPC10, MAOB, NGF, MMP12, JAG1, CD40LG, PPARG, FGF2, LRP1, NOTCH4, MAPK8, PREP, NOTCH3, PRNP, CTSG, EGF, REN, CD44, SELP, GHR, ADCYAP1, INSR, GFAP, MMP3, MAPK10, SP1, MYC, CTSE, PPARA, JUN, TIMP1, IL5, IL1A, MMP9, HTR4, HSPG2, KRAS, CYCS, SMG1, IL1R1, PROK1, MAPK3, NTRK1, IL13, MME, TKT, CXCR2, CHRM1, ATXN1, PAWR, NOTCH2, M6PR, CYP46A1, CSNK1D, MAPK14, PRG2, PRKCA, L1 CAM, CD40, NR1H2, JAG2, CTNND1, CMA1, SORT1, DLK1, THEM4, JUP, CD46, CCL11, CAV3, RNASE3, HSPA8, CASP9, CYP3A4, CCR3, TFAP2A, SCP2, CDK4, JOF1A, TCF7L2, B3GALT, MDM2, RELA, CASP7, IDE, FANP4, CASK, ADCYAP1R1, ATF4, PDGFA, C21ORF33, SCG5, RMF123, NKFB1, ERBB2, CAV1, MMP7, TGFA, RXRA, STX1A, PSMC4, P2RY2, TNFRSF21, DLG1, NUMBL, SPN, PLSCR1, UBQLN2, UBQLN1, PCSK7, SPON1, SILV, QPCT, HESS, GCC1
Selective IgA Deficiency	Immune system		Type 1: MSH5; Type 2: TNFRSF13B
Severe Combined Immunodeficiency (SCID) and SCID-X1, and ADA-SCID	Immune system		JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4, those identified in US Pat. App. Pub. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937;
Sickle cell disease	blood		HBB, BCL11A, BCL11Ae, cis-regulatory elements of the B-globin locus, HBG 1/2 promoter, HBG distal CCAAT box region between -92 and -130 of the HBG Transcription Start Site, those described in WO2015148863, WO 2013/126794, US Pat. Pub. 20110182867
Sly Syndrome (aka MPS VII)			GUSB
Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, 8, 12 and 17)			ATXN1, ATXN2, ATX3
Sorsby Fundus Dystrophy	eye		TIMP3
Stargardt disease	eye		ABCR, ELOVL4, ABCA4, PROM1

Tay-Sachs Disease	Lysosomal Storage disease	Various – CNS, brain, eye	HEX-A
Thalassemia (Alpha, Beta, Delta)	blood		HBA1, HBA2 (Alpha), HBB (Beta), HBB and HBD (delta), LCRB, BCL11A, BCL11Ae, cis-regulatory elements of the B-globin locus, HBG 1/2 promoter, those described in WO2015148860, US Pat. Pub. 20110182867, 2015/148860
Thymic Aplasia (DiGeorge Syndrome; 22q11.2 deletion syndrome)	Immune system, thymus		deletion of 30 to 40 genes in the middle of chromosome 22 at a location known as 22q11.2, including TBX1, DGCR8
Transthyretin amyloidosis (ATTR)	liver		TTR (transthyretin)
trimethylaminuria	Metabolism disease		FMO3
Trinucleotide Repeat Disorders (generally)	Various		HTT; SBMA/SMAX1/AR; FXN/X25 ATX3; ATXN1; ATXN2; DMPK; Atrophin-1 and Atn1 (DRPLA Dx); CBP (Creb-BP - global instability); VLDLR; Atxn7; Atxn10; FEN1, TNRC6A, PABPN1, JPH3, MED15, ATXN1, ATXN3, TBP, CACNA1A, ATXN80S, PPP2R2B, ATXN7, TNRC6B, TNRC6C, CELF3, MAB21L1, MSH2, TMEM185A, SIX5, CNPY3, RAXE, GNB2, RPL14, ATXN8, ISR, TTR, EP400, GIGYF2, OGG1, STC1, CNDP1, C10ORF2, MAML3, DKC1, PAXIP1, CASK, MAPT, SP1, POLG, AFF2, THBS1, TP53, ESR1, CGGBP1, ABT1, KLK3, PRNP, JUN, KCNN3, BAX, FRAXA, KBTBD10, MBNL1, RAD51, NCOA3, ERDA1, TSC1, COMP, GGLC, RRAD, MSH3, DRD2, CD44, CTCF, CCND1, CLSPN, MEF2A, PTPRU, GAPDH, TRIM22, WT1, AHR, GPX1, TPMT, NDP, ARX, TYR, EGR1, UNG, NUMBL, FABP2, EN2, CRYGC, SRP14, CRYGB, PDCD1, HOXA1, ATXN2L, PMS2, GLA, CBL, FTH1, IL12RB2, OTX2, HOXA5, POLG2, DLX2, AHRR, MANF, RMEM158, See also 20110016540
Turner's Syndrome (XO)	Various – reproductive organs, and sex characteristics, vasculature		Monosomy X
Tuberous Sclerosis	CNS, heart, kidneys		TSC1, TSC2
Usher syndrome (Types I, II, and III)	Ears, eyes		ABHD12, CDH23, CIB2, CLRN1, DFNB31, GPR98, HARS, MYO7A,

			PCDH15, USH1C, USH1G, USH2A, USH11A, those described in WO2015134812A1
Velocardiofacial syndrome (aka 22q11.2 deletion syndrome, DiGeorge syndrome, conotruncal anomaly face syndrome (CTAF), autosomal dominant Opitz G/BB syndrome or Cayler cardiofacial syndrome)	Various – skeletal, heart, kidney, immune system, brain		Many genes are deleted, COM, TBX1, and other are associated with symptoms
Von Gierke's Disease (Glycogen Storage Disease type I)	Glycogen Storage disease	Various – liver, kidney	G6PC and SLC37A4
Von Hippel-Lindau Syndrome	Various – cell growth regulation disorder	CNS, Kidney, Eye, visceral organs	VHL
Von Willebrand Disease (Types I, II and III)	blood		VWF
Wilson Disease	Various - Copper Storage Disease	Liver, brains, eyes, other tissues where copper builds up	ATP7B
Wiskott-Aldrich Syndrome	Immune System		WAS
Xeroderma Pigmentosum	Skin	Nervous system	POLH
XXX Syndrome	Endocrine, brain		X chromosome trisomy

[0617] In one embodiment, the compositions, systems, or components thereof can be used to treat or prevent a disease in a subject by modifying one or more genes associated with one or more cellular functions, such as any one or more of those in **Table 5B**. In one embodiment, the disease is a genetic disease or disorder. In some of embodiments, the composition, system, or component thereof can modify one or more genes or polynucleotides associated with one or more genetic diseases such as any set forth in **Table 5B**.

[0618] **Table 5B.**

Exemplary Genes controlling Cellular Functions	
CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2; PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1; AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2; PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2; ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3; PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7; YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A; CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1; PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2; TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK; HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1; PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A; PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN; EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ; PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1; PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1; CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor Signaling	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1; MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I; PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2; MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1; MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13; RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1; PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3; MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP; CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1; ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1; STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12; IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2; ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2; CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11; PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1; GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B; AKT3; PRKCA

Ephrin Receptor Signaling Actin Cytoskeleton Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1; PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2; MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2; DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14; CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2; PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1; MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10; MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2; EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4; AKT3; SGK
	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1; PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6; ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8; F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7; PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1; MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3; ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL; BRAF; VAV3; SGK
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2; MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2; PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST; GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1; GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2; HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A; HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1; PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX; ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1; BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB; CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8; BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF; RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2; CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2; BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK; CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11; AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A; MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1; MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9; EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1; NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN; GSK3B; ATF4; AKT3; VAV3; RPS6KB1

Leukocyte Extravasation Signaling	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA; RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11; MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12; PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB; MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK; MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2; CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK; CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A; TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2; CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8; CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA; SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP; RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1; TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2; CRKL; BRAF; GSK3B; AKT3
Acute Phase Response Signaling	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11; AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14; PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1; TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1; IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN; AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11; MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA; CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1; MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR; RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2; AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1; NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2; GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 Signaling Aryl Hydrocarbon Receptor Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A; BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2; PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1; PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9; CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A; HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1; SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN; SNAI2; GSK3B; BAX; AKT3 HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1; NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1; SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1; MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1; SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF; CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1; CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1; HSP90AA1

Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1; NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A; PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1; ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD; GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL; NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1; CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1; NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1; HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1; GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA; FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1; GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2; PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1; CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN; RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2; ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8; IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A; NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1; TGFB1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling Wnt & Beta catenin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1 CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFB1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2

Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8; IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A; YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3; FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative Stress Response	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1; NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL; NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP; MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1; GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic Stellate Cell Activation	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF; SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9; IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8; PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX; IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB; NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3; NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2; PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG; RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA; MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8; PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD; MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3; VAV3; PRKCA

G-Protein Coupled Receptor Signaling	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB; PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB; PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1; IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK; PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3; PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6; MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3; MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1; MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC; PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2; PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC; JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF; AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3; BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN; RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN; VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11; KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6; PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1; PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S Checkpoint Regulation	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC; ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11; HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1; E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1; GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS; NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN; MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10; JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD; FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8; DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB; CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3; BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11; AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8; MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1; AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4; AKT3; PRKCA; HGF

GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A; STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3; ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3; STAT1
Amyotrophic Lateral Sclerosis Signaling	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2; PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1; PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1; APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B; PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A; PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3; STAT1
Nicotinate and Nicotinamide Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1; PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1; PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2; MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ; CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13; RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1; MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS; STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2; JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term Depression	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS; PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3; KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA; YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2; SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1; HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP; MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4; CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7; USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8; USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFkB2; MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF; IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFkB1; JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1; NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD; RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1; LRP5; CEBPB; FOXO1; PRKCA

TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1; FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2; SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2; MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1; IKKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBK; RELB; MAP3K7; CHUK; NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS; CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2; MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1; SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A; RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8; APOB; MAPK10; PPARG; MTPP; MAPK9; PPARGC1A; TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1; PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS; PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1; ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1; CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11; HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4; HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3; MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1; STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT; HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM; VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1; MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2; TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA; NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1; SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2; CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1; PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1; PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1; FRAP1; AKT3; RPS6KB1

Cell Cycle: G2/M DNA Damage Checkpoint Regulation	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC; CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A; PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the Cardiovascular System	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3; CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1; VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4; PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C; NT5E; POLD1; NME1
cAMP-mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3; SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction Notch Signaling	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3 HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum Stress Pathway Pyrimidine Metabolism	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3 NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B; NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7; PARK2; CASP3
Cardiac & Beta Adrenergic Signaling	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC; PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid Metabolism	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair Pathway	ERCC5; ERCC4; XPA; XPC; ERCC1
Starch and Sucrose Metabolism	UCHL1; HK2; GCK; GPI; HK1
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor Signaling	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline Metabolism	ALDH1A1; NOS3; NOS2A

Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose Metabolism	HK2; GCK; HK1
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR
Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid Metabolism	PRDX6; CHKA
Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics by Cytochrome p450	GSTP1; CYP1B1
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate Metabolism	PRMT5
Androgen and Estrogen Metabolism	PRMT5
Ascorbate and Aldarate Metabolism	ALDH1A1
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor Signaling	GNB2L1
NRF2-mediated Oxidative Stress Response	PRDX1
Pentose Phosphate Pathway	GPI
Pentose and Glucuronate Interconversions	UCHL1
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5; TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and Isoleucine Degradation	ALDH1A1

Glycine, Serine and Threonine Metabolism	CHKA
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2; Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca; Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2; Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b; Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin; Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8; Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[0619] In an aspect, the invention provides a method of individualized or personalized treatment of a genetic disease in a subject in need of such treatment comprising: (a) introducing one or more mutations *ex vivo* in a tissue, organ or a cell line, or *in vivo* in a transgenic non-human mammal, comprising delivering to cell(s) of the tissue, organ, cell or mammal a composition comprising the particle delivery system or the delivery system or the virus particle of any one of the above embodiment or the cell of any one of the above embodiment, wherein the specific mutations or precise sequence substitutions are or have been correlated to the genetic disease; (b) testing treatment(s) for the genetic disease on the cells to which the vector has been delivered that have the specific mutations or precise sequence substitutions correlated to the genetic disease; and (c) treating the subject based on results from the testing of treatment(s) of step (b).

Infectious Diseases

[0620] In one embodiment, the composition, system(s) or component(s) thereof can be used to diagnose, prognose, treat, and/or prevent an infectious disease caused by a microorganism, such as bacteria, virus, fungi, parasites, or combinations thereof.

[0621] In one embodiment, the system(s) or component(s) thereof can be capable of targeting specific microorganism within a mixed population. Exemplary methods of such techniques are described in e.g., Gomaa AA, Klumpe HE, Luo ML, Selle K, Barrangou R, Beisel CL. 2014. Programmable removal of bacterial strains by use of genome-targeting composition, systems, mBio 5:e00928-13; Citorik RJ, Mimee M, Lu TK. 2014. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. Nat Biotechnol 32:1141–1145, the teachings of which can be adapted for use with the compositions, systems, and components thereof disclosed herein.

[0622] In one embodiment, the composition, system(s) and/or components thereof can be capable of targeting pathogenic and/or drug-resistant microorganisms, such as bacteria, virus, parasites, and fungi. In one embodiment, the composition, system(s) and/or components thereof can be capable of targeting and modifying one or more polynucleotides in a pathogenic microorganism such that the microorganism is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host cell.

[0623] In one embodiment, the pathogenic bacteria that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, those of the genus *Actinomyces* (e.g. *A. israelii*), *Bacillus* (e.g. *B. anthracis*, *B. cereus*), *Bacteroides* (e.g. *B. fragilis*), *Bartonella* (*B. henselae*, *B. quintana*), *Bordetella* (*B. pertussis*), *Borrelia* (e.g. *B. burgdorferi*, *B. garinii*, *B. afzelii*, and *B. recurrentis*), *Brucella* (e.g. *B. abortus*, *B. canis*, *B. melitensis*, and *B. suis*), *Campylobacter* (e.g. *C. jejuni*), *Chlamydia* (e.g. *C. pneumoniae* and *C. trachomatis*), *Chlamydophila* (e.g. *C. psittaci*), *Clostridium* (e.g. *C. botulinum*, *C. difficile*, *C. perfringens*, *C. tetani*), *Corynebacterium* (e.g. *C. diphtheriae*), *Enterococcus* (e.g. *E. Faecalis*, *E. faecium*), *Ehrlichia* (*E. canis* and *E. chaffensis*), *Escherichia* (e.g. *E. coli*), *Francisella* (e.g. *F. tularensis*), *Haemophilus* (e.g. *H. influenzae*), *Helicobacter* (*H. pylori*), *Klebsiella* (e.g. *K. pneumoniae*), *Legionella* (e.g. *L. pneumophila*), *Leptospira* (e.g. *L. interrogans*, *L. santarosai*, *L. weilii*, *L. noguchii*), *Listeria* (e.g. *L. monocytogenes*), *Mycobacterium* (e.g. *M. leprae*, *M. tuberculosis*, *M. ulcerans*), *Mycoplasma* (*M. pneumoniae*), *Neisseria* (*N. gonorrhoeae* and *N. meningitidis*), *Nocardia* (e.g. *N. asteroides*), *Pseudomonas* (*P. aeruginosa*), *Rickettsia* (*R. rickettsia*), *Salmonella* (*S. typhi* and *S. typhimurium*), *Shigella* (*S. sonnei* and *S. dysenteriae*), *Staphylococcus* (*S. aureus*, *S. epidermidis*, and *S. saprophyticus*), *Streptococcus* (*S. agalactiae*, *S. pneumoniae*, *S. pyogenes*), *Treponema* (*T. pallidum*), *Ureaplasma* (e.g. *U. urealyticum*), *Vibrio* (e.g. *V. cholerae*), *Yersinia* (e.g. *Y. pestis*, *Y. enterocolitica*, and *Y. pseudotuberculosis*).

[0624] In one embodiment, the pathogenic virus that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, a double-stranded DNA virus, a partly double-stranded DNA virus, a single-stranded DNA virus, a positive single-stranded RNA virus, a negative single-stranded RNA virus, or a double stranded RNA virus. In one embodiment, the pathogenic virus can be from the family *Adenoviridae* (e.g. *Adenovirus*), *Herpesviridae* (e.g. *Herpes simplex*, type 1, *Herpes simplex*, type 2, *Varicella-zoster virus*, *Epstein–Barr virus*, *Human cytomegalovirus*, *Human*

herpesvirus, type 8), Papillomaviridae (e.g. Human papillomavirus), Polyomaviridae (e.g. BK virus, JC virus), Poxviridae (e.g. smallpox), Hepadnaviridae (e.g. Hepatitis B), Parvoviridae (e.g. Parvovirus B19), Astroviridae (e.g. Human astrovirus), Caliciviridae (e.g. Norwalk virus), Picornaviridae (e.g. coxsackievirus, hepatitis A virus, poliovirus, rhinovirus), Coronaviridae (e.g. Severe acute respiratory syndrome-related coronavirus, strains: Severe acute respiratory syndrome virus, Severe acute respiratory syndrome coronavirus 2 (COVID-19 and variants)), Flaviviridae (e.g. Hepatitis C virus, yellow fever virus, dengue virus, West Nile virus, TBE virus), Togaviridae (e.g. Rubella virus), Hepeviridae (e.g. Hepatitis E virus), Retroviridae (Human immunodeficiency virus (HIV)), Orthomyxoviridae (e.g. Influenza virus), Arenaviridae (e.g. Lassa virus), Bunyaviridae (e.g. Crimean-Congo hemorrhagic fever virus, Hantaan virus), Filoviridae (e.g. Ebola virus and Marburg virus), Paramyxoviridae (e.g. Measles virus, Mumps virus, Parainfluenza virus, Respiratory syncytial virus), Rhabdoviridae (Rabies virus), Hepatitis D virus, Reoviridae (e.g. Rotavirus, Orbivirus, Coltivirus, Banna virus).

[0625] In one embodiment, the pathogenic fungi that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, those of the genus *Candida* (e.g. *C. albicans*), *Aspergillus* (e.g. *A. fumigatus*, *A. flavus*, *A. clavatus*), *Cryptococcus* (e.g. *C. neoformans*, *C. gattii*), *Histoplasma* (*H. capsulatum*), *Pneumocystis* (e.g. *P. jirovecii*), *Stachybotrys* (e.g. *S. chartarum*).

[0626] In one embodiment, the pathogenic parasites that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, protozoa, helminths, and ectoparasites. In one embodiment, the pathogenic protozoa that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, those from the groups Sarcodina (e.g. ameba such as *Entamoeba*), Mastigophora (e.g. flagellates such as *Giardia* and *Leishmania*), Cilophora (e.g. ciliates such as *Balantidium*), and sporozoa (e.g. *Plasmodium* and *Cryptosporidium*). In one embodiment, the pathogenic helminths that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, flatworms (platyhelminths), thorny-headed worms (acanthocephalins), and roundworms (nematodes). In one embodiment, the pathogenic ectoparasites that can be targeted and/or modified by the composition, system(s) and/or component(s) thereof disclosed herein include, but are not limited to, ticks, fleas, lice, and mites.

[0627] In one embodiment, the pathogenic parasite that can be targeted and/or modified by the composition, system,(s) and/or component(s) thereof disclosed herein include, but are not limited to, *Acanthamoeba* spp., *Balamuthia mandrillaris*, Babesiosis spp. (e.g. *Babesia B. divergens*, *B. bigemina*, *B. equi*, *B. microfti*, *B. duncani*), Balantidiasis spp. (e.g. *Balantidium coli*), Blastocystis spp., *Cryptosporidium* spp., Cyclosporiasis spp. (e.g. *Cyclospora cayetanensis*), Dientamoebiasis spp. (e.g. *Dientamoeba fragilis*), Amoebiasis spp. (e.g. *Entamoeba histolytica*), Giardiasis spp. (e.g. *Giardia lamblia*), Isosporiasis spp. (e.g. *Isospora belli*), Leishmania spp., Naegleria spp. (e.g. *Naegleria fowleri*), Plasmodium spp. (e.g. *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale curtisi*, *Plasmodium ovale wallikeri*, *Plasmodium malariae*, *Plasmodium knowlesi*), Rhinosporidiosis spp. (e.g. *Rhinosporidium Seeberi*), Sarcocystosis spp. (e.g. *Sarcocystis bovi hominis*, *Sarcocystis sui hominis*), Toxoplasma spp. (e.g. *Toxoplasma gondii*), Trichomonas spp. (e.g. *Trichomonas vaginalis*), Trypanosoma spp. (e.g. *Trypanosoma brucei*), Trypanosoma spp. (e.g. *Trypanosoma cruzi*), Tapeworm (e.g. *Cestoda*, *Taenia multiceps*, *Taenia saginata*, *Taenia solium*), Diphyllbothrium latum spp., Echinococcus spp. (e.g. *Echinococcus granulosus*, *Echinococcus multilocularis*, *E. vogeli*, *E. oligarthrus*), Hymenolepis spp. (e.g. *Hymenolepis nana*, *Hymenolepis diminuta*), Bertiella spp. (e.g. *Bertiella mucronata*, *Bertiella studeri*), Spirometra (e.g. *Spirometra erinaceieuropaei*), Clonorchis spp. (e.g. *Clonorchis sinensis*; *Clonorchis viverrini*), Dicrocoelium spp. (e.g. *Dicrocoelium dendriticum*), Fasciola spp. (e.g. *Fasciola hepatica*, *Fasciola gigantica*), Fasciolopsis spp. (e.g. *Fasciolopsis buski*), Metagonimus spp. (e.g. *Metagonimus yokogawai*), Metorchis spp. (e.g. *Metorchis conjunctus*), Opisthorchis spp. (e.g. *Opisthorchis viverrini*, *Opisthorchis felinus*), Clonorchis spp. (e.g. *Clonorchis sinensis*), Paragonimus spp. (e.g. *Paragonimus westermani*; *Paragonimus africanus*; *Paragonimus caliensis*; *Paragonimus kellicotti*; *Paragonimus skrjabini*; *Paragonimus uterobilateralis*), Schistosoma sp., Schistosoma spp. (e.g. *Schistosoma mansoni*, *Schistosoma haematobium*, *Schistosoma japonicum*, *Schistosoma mekongi*, and *Schistosoma intercalatum*), Echinostoma spp. (e.g. *E. echinatum*), Trichobilharzia spp. (e.g. *Trichobilharzia regenti*), Ancylostoma spp. (e.g. *Ancylostoma duodenale*), Necator spp. (e.g. *Necator americanus*), Angiostrongylus spp., Anisakis spp., Ascaris spp. (e.g. *Ascaris lumbricoides*), Baylisascaris spp. (e.g. *Baylisascaris procyonis*), Brugia spp. (e.g. *Brugia malayi*, *Brugia timori*), Dioctophyme spp. (e.g. *Dioctophyme renale*), Dracunculus spp. (e.g. *Dracunculus medinensis*), Enterobius spp. (e.g. *Enterobius vermicularis*, *Enterobius gregorii*), Gnathostoma

spp. (e.g. *Gnathostoma spinigerum*, *Gnathostoma hispidum*), Halicephalobus spp. (e.g. *Halicephalobus gingivalis*), Loa loa spp. (e.g. *Loa loa filaria*), Mansonella spp. (e.g. *Mansonella streptocerca*), Onchocerca spp. (e.g. *Onchocerca volvulus*), Strongyloides spp. (e.g. *Strongyloides stercoralis*), Thelazia spp. (e.g. *Thelazia californiensis*, *Thelazia callipaeda*), Toxocara spp. (e.g. *Toxocara canis*, *Toxocara cati*, *Toxascaris leonine*), Trichinella spp. (e.g. *Trichinella spiralis*, *Trichinella britovi*, *Trichinella nelsoni*, *Trichinella nativa*), Trichuris spp. (e.g. *Trichuris trichiura*, *Trichuris vulpis*), Wuchereria spp. (e.g. *Wuchereria bancrofti*), Dermatobia spp. (e.g. *Dermatobia hominis*), Tunga spp. (e.g. *Tunga penetrans*), Cochliomyia spp. (e.g. *Cochliomyia hominivorax*), Linguatula spp. (e.g. *Linguatula serrata*), Archiacanthocephala sp., Moniliformis sp. (e.g. *Moniliformis moniliformis*), Pediculus spp. (e.g. *Pediculus humanus capitis*, *Pediculus humanus humanus*), Pthirus spp. (e.g. *Pthirus pubis*), Arachnida spp. (e.g. *Trombiculidae*, *Ixodidae*, *Argasidae*), Siphonaptera spp (e.g. *Siphonaptera: Pulicinae*), Cimicidae spp. (e.g. *Cimex lectularius* and *Cimex hemipterus*), Diptera spp., Demodex spp. (e.g. *Demodex folliculorum/brevis/canis*), Sarcoptes spp. (e.g. *Sarcoptes scabiei*), Dermanyssus spp. (e.g. *Dermanyssus gallinae*), Ornithonyssus spp. (e.g. *Ornithonyssus sylviarum*, *Ornithonyssus bursa*, *Ornithonyssus bacoti*), Laelaps spp. (e.g. *Laelaps echidnina*), Liponyssoides spp. (e.g. *Liponyssoides sanguineus*).

[0628] In one embodiment the gene targets can be any of those as set forth in Table 1 of Strich and Chertow, 2019. J. Clin. Microbio. 57:4 e01307-18, which is incorporated herein in its entirety and can be adapted for use with the engineered Type V Cas proteins disclosed herein.

[0629] In one embodiment, the method can include delivering a composition, system, and/or component thereof to a pathogenic organism disclosed herein, allowing the composition, system, and/or component thereof to specifically bind and modify one or more targets in the pathogenic organism, whereby the modification kills, inhibits, reduces the pathogenicity of the pathogenic organism, or otherwise renders the pathogenic organism non-pathogenic. In one embodiment, delivery of the composition, system, occurs *in vivo* (i.e., in the subject being treated). In one embodiment occurs by an intermediary, such as microorganism or phage that is non-pathogenic to the subject but is capable of transferring polynucleotides and/or infecting the pathogenic microorganism. In one embodiment, the intermediary microorganism can be an engineered bacteria, virus, or phage that contains the composition,

system(s) and/or component(s) thereof and/or vectors and/or vector systems. The method can include administering an intermediary microorganism containing the composition, system(s) and/or component(s) thereof and/or vectors and/or vector systems to the subject to be treated. The intermediary microorganism can then produce the compositions and/or component thereof or transfer a composition, system, polynucleotide to the pathogenic organism. In embodiments, where the compositions and/or component thereof, vector, or vector system is transferred to the pathogenic microorganism, the composition, system, or component thereof is then produced in the pathogenic microorganism and modifies the pathogenic microorganism such that it is less virulent, killed, inhibited, or is otherwise rendered incapable of causing disease and/or infecting and/or replicating in a host or cell thereof.

[0630] In one embodiment, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g., a virus), the composition, system, can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA cannot be replicated by the host cell's machinery into a functional virus. In one embodiment, where the pathogenic microorganism inserts its genetic material into the host cell's genome (e.g., a virus), the composition, system, can be designed such that it modifies the host cell's genome such that the viral DNA or cDNA is deleted from the host cell's genome.

[0631] It will be appreciated that inhibiting or killing the pathogenic microorganism, the disease and/or condition that its infection causes in the subject can be treated or prevented. Thus, also provided herein are methods of treating and/or preventing one or more diseases or symptoms thereof caused by any one or more pathogenic microorganisms, such as any of those disclosed herein.

Mitochondrial Diseases

[0632] Some of the most challenging mitochondrial disorders arise from mutations in mitochondrial DNA (mtDNA), a high copy number genome that is maternally inherited. In one embodiment, mtDNA mutations can be modified using a composition, system, disclosed herein. In one embodiment, the mitochondrial disease that can be diagnosed, prognosed, treated, and/or prevented can be MELAS (mitochondrial myopathy encephalopathy, and lactic acidosis and stroke-like episodes), CPEO/PEO (chronic progressive external ophthalmoplegia syndrome/progressive external ophthalmoplegia), KSS (Kearns-Sayre syndrome), MIDD (maternally inherited diabetes and deafness), MERRF (myoclonic epilepsy associated with ragged red fibers), NIDDM (noninsulin-dependent diabetes mellitus), LHON (Leber hereditary

optic neuropathy), LS (Leigh Syndrome) an aminoglycoside induced hearing disorder, NARP (neuropathy, ataxia, and pigmentary retinopathy), Extrapyrarnidal disorder with akinesia-rigidity, psychosis and SNHL, Nonsyndromic hearing loss a cardiomyopathy, an encephalomyopathy, Pearson's syndrome, or a combination thereof.

[0633] In one embodiment, the mtDNA of a subject can be modified *in vivo* or *ex vivo*. In one embodiment, where the mtDNA is modified *ex vivo*, after modification the cells containing the modified mitochondria can be administered back to the subject. In one embodiment, the composition, system, or component thereof can be capable of correcting an mtDNA mutation, or a combination thereof.

[0634] In one embodiment, at least one of the one or more mtDNA mutations is selected from the group consisting of: A3243G, C3256T, T3271C, G1019A, A1304T, A15533G, C1494T, C4467A, T1658C, G12315A, A3421G, A8344G, T8356C, G8363A, A13042T, T3200C, G3242A, A3252G, T3264C, G3316A, T3394C, T14577C, A4833G, G3460A, G9804A, G11778A, G14459A, A14484G, G15257A, T8993C, T8993G, G10197A, G13513A, T1095C, C1494T, A1555G, G1541A, C1634T, A3260G, A4269G, T7587C, A8296G, A8348G, G8363A, T9957C, T9997C, G12192A, C12297T, A14484G, G15059A, duplication of CCCCCTCCCC-tandem repeats (SEQ ID NO: 56) at positions 305-314 and/or 956-965, deletion at positions from 8,469-13,447, 4,308-14,874, and/or 4,398-14,822, 961ins/delC, the mitochondrial common deletion (e.g. mtDNA 4,977 bp deletion), and combinations thereof.

[0635] In one embodiment, the mitochondrial mutation can be any mutation as set forth in or as identified by use of one or more bioinformatic tools available at Mitomap available at mitomap.org. Such tools include, but are not limited to, "Variant Search, aka Market Finder", Find Sequences for Any Haplogroup, aka "Sequence Finder", "Variant Info", "POLG Pathogenicity Prediction Server", "MITOMASTER", "Allele Search", "Sequence and Variant Downloads", "Data Downloads". MitoMap contains reports of mutations in mtDNA that can be associated with disease and maintains a database of reported mitochondrial DNA Base Substitution Diseases: rRNA/tRNA mutations.

[0636] In one embodiment, the method includes delivering a composition, system, and/or a component thereof to a cell, and more specifically one or more mitochondria in a cell, allowing the composition, system, and/or component thereof to modify one or more target polynucleotides in the cell, and more specifically one or more mitochondria in the cell. The target polynucleotides can correspond to a mutation in the mtDNA, such as any one or more of

those disclosed herein. In one embodiment, the modification can alter a function of the mitochondria such that the mitochondria functions normally or at least is/are less dysfunctional as compared to an unmodified mitochondria. Modification can occur *in vivo* or *ex vivo*. Where modification is performed *ex vivo*, cells containing modified mitochondria can be administered to a subject in need thereof in an autologous or allogenic manner.

Microbiome Modification

[0637] Microbiomes play important roles in health and disease. For example, the gut microbiome can play a role in health by controlling digestion, preventing growth of pathogenic microorganisms and have been suggested to influence mood and emotion. Imbalanced microbiomes can promote disease and are suggested to contribute to weight gain, unregulated blood sugar, high cholesterol, cancer, and other disorders. A healthy microbiome has a series of joint characteristics that can be distinguished from non-healthy individuals, thus detection and identification of the disease-associated microbiome can be used to diagnose and detect disease in an individual. The compositions, systems, and components thereof can be used to screen the microbiome cell population and be used to identify a disease associated microbiome. Cell screening methods utilizing compositions, systems, and components thereof are described elsewhere herein and can be applied to screening a microbiome, such as a gut, skin, vagina, and/or oral microbiome, of a subject.

[0638] In one embodiment, the microbe population of a microbiome in a subject can be modified using a composition, system, and/or component thereof disclosed herein. In one embodiment, the composition, system, and/or component thereof can be used to identify and select one or more cell types in the microbiome and remove them from the microbiome population. Exemplary methods of selecting cells using a composition, system, and/or component thereof are described elsewhere herein. In this way the make-up or microorganism profile of the microbiome can be altered. In one embodiment, the alteration causes a change from a diseased microbiome composition to a healthy microbiome composition. In this way the ratio of one type or species of microorganism to another can be modified, such as going from a diseased ratio to a healthy ratio. In one embodiment, the cells selected are pathogenic microorganisms.

[0639] In one embodiment, the compositions and systems disclosed herein can be used to modify a polynucleotide in a microorganism of a microbiome in a subject. In one embodiment, the microorganism is a pathogenic microorganism. In one embodiment, the microorganism is

a commensal and non-pathogenic microorganism. Methods of modifying polynucleotides in a cell in the subject are described elsewhere herein and can be applied to these embodiments.

Gene Drives

[0640] The present invention also contemplates use of the systems, compositions, system components, vectors, engineered cells, disclosed herein to generate a gene drive via delivery of one or more system polypeptides and/or polynucleotides capable of producing a gene drive. In some embodiments, the gene drive can be a Cas-mediated RNA-guided gene drive e.g., Cas-to provide RNA-guided gene drives, for example in systems analogous to gene drives described in PCT Patent Publication WO 2015/105928. Systems of this kind may for example provide methods for altering eukaryotic germline cells, by introducing into the germline cell a nucleic acid sequence encoding an RNA-guided DNA nuclease and one or more guide RNAs. The guide RNAs may be designed to be complementary to one or more target locations on genomic DNA of the germline cell. The nucleic acid sequence encoding the RNA guided DNA nuclease and the nucleic acid sequence encoding the guide RNAs may be provided on constructs between flanking sequences, with promoters arranged such that the germline cell may express the RNA guided DNA nuclease and the guide RNAs, together with any desired cargo-encoding sequences that are also situated between the flanking sequences. The flanking sequences will typically include a sequence which is identical to a corresponding sequence on a selected target chromosome, so that the flanking sequences work with the components encoded by the construct to facilitate insertion of the foreign nucleic acid construct sequences into genomic DNA at a target cut site by mechanisms such as homologous recombination, to render the germline cell homozygous for the foreign nucleic acid sequence. In this way, gene-drive systems are capable of introgressing desired cargo genes throughout a breeding population (Gantz et al., 2015, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*, PNAS 2015, published ahead of print November 23, 2015, doi:10.1073/pnas.1521077112; Esvelt et al., 2014, Concerning RNA-guided gene drives for the alteration of wild populations eLife 2014;3:e03401). In select embodiments, target sequences may be selected which have few potential off-target sites in a genome. Targeting multiple sites within a target locus, using multiple guide RNAs, may increase the cutting frequency and hinder the evolution of drive resistant alleles. Truncated guide RNAs may reduce off-target cutting. Paired nickases may be used instead of a single nuclease, to further increase specificity. Gene drive constructs (such as gene drive engineered

delivery system constructs) may include cargo sequences encoding transcriptional regulators, for example to activate homologous recombination genes and/or repress non-homologous end-joining. Target sites may be chosen within an essential gene, so that non-homologous end-joining events may cause lethality rather than creating a drive-resistant allele. The gene drive constructs can be engineered to function in a range of hosts at a range of temperatures (Cho et al. 2013, Rapid and Tunable Control of Protein Stability in *Caenorhabditis elegans* Using a Small Molecule, PLoS ONE 8(8): e72393. doi:10.1371/journal.pone.0072393).

Transplantation and Xenotransplantation

[0641] The present invention also contemplates use of the composition disclosed herein, e.g., Type V Cas polypeptide protein systems, to provide RNA-guided DNA nucleases adapted to be used to provide modified tissues for transplantation (transplantation between two subjects of the same species) and/or xenotransplantation (transplantation between species). For example, RNA-guided DNA nucleases may be used to knockout, knockdown or disrupt selected genes in an animal, such as a transgenic pig (such as the human heme oxygenase-1 transgenic pig line), for example by disrupting expression of genes that encode epitopes recognized by the human immune system, i.e., xenoantigen genes. Candidate porcine genes for disruption may for example include $\alpha(1,3)$ -galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase genes (*See* PCT Patent Publication WO 2014/066505). In addition, genes encoding endogenous retroviruses may be disrupted, for example the genes encoding all porcine endogenous retroviruses (*See* Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), *Science* 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). In addition, RNA-guided DNA nucleases may be used to target a site for integration of additional genes in xenotransplant donor animals, such as a human CD55 gene to improve protection against hyperacute rejection.

[0642] In other examples, such as in the context of transplantation, the systems and compositions disclosed herein can be used to modify the organ or tissue to be transplanted to reduce, for example, immunogenicity, acute rejection, GvD, and other disorders and complications associated with transplantation. Such methods can allow for utilization of tissues that would otherwise not be a good or ideal match to a recipient for transplantation insofar as they can be modified prior to transplantation to improve the match characteristics and/or minimize tissue antigens and/or the like. For example, where it is interspecies transplantation (such as human to human) the systems, compositions, vectors, cells, and/or the like disclosed

herein, can be used to deliver cargo polynucleotides and/or otherwise be involved to modify the tissue to be transplanted. In some embodiments, the modification can include modifying one or more HLA antigens or other tissue type determinants, such that the immunogenic profile is more similar or identical to the recipient's immunogenic profile than to the donor's so as to reduce the occurrence of rejection by the recipient. Relevant tissue type determinants are known in the art (such as those used to determine organ matching) and techniques to determine the immunogenic profile (which is made up of the expression signature of the tissue type determinants) are generally known in the art.

[0643] In some embodiments, the donor (such as before harvest) or recipient (after transplantation) can receive one or more of the systems, composition, vectors, cells, and/or the like of the present invention disclosed herein that are capable of modifying the immunogenic profile of the transplanted cells, tissue, and/or organ. In some embodiments, the transplanted cells, tissue, and/or organ can be harvested from the donor and the systems, composition, vectors, cells, and/or the like of the present invention disclosed herein capable of modifying the harvested cells, tissue, and/or organ to be, for example, less immunogenic or be modified to have some specific characteristic when transplanted in the recipient can be delivered to the harvested cells, tissue, and/or organ ex vivo. After delivery the cells, tissue, and/or organs can be transplanted into the donor.

[0644] Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The present effector protein systems may be harnessed to correct these defects of genomic instability.

[0645] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar

Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

Considerations for Therapeutic Applications

[0646] A consideration in genome editing therapy is the choice of sequence-specific nuclease, such as a variant of a Type V Cas polypeptide. Each nuclease variant may possess its own unique set of strengths and weaknesses, many of which must be balanced in the context of treatment to maximize therapeutic benefit. For a specific editing therapy to be efficacious, a sufficiently high level of modification must be achieved in target cell populations to reverse disease symptoms. This therapeutic modification 'threshold' is determined by the fitness of edited cells following treatment and the amount of gene product necessary to reverse symptoms. With regard to fitness, editing creates three potential outcomes for treated cells relative to their unedited counterparts: increased, neutral, or decreased fitness. In the case of increased fitness, corrected cells may be able to expand relative to their diseased counterparts to mediate therapy. In this case, where edited cells possess a selective advantage, even low numbers of edited cells can be amplified through expansion, providing a therapeutic benefit to the patient. Where the edited cells possess no change in fitness, an increase in the therapeutic modification threshold can be warranted. As such, significantly greater levels of editing may be needed to treat diseases where editing creates a neutral fitness advantage, relative to diseases where editing creates increased fitness for target cells. If editing imposes a fitness disadvantage, as would be the case for restoring function to a tumor suppressor gene in cancer cells, modified cells would be outcompeted by their diseased counterparts, causing the benefit of treatment to be low relative to editing rates. This may be overcome with supplemental therapies to increase the potency and/or fitness of the edited cells relative to the diseased counterparts.

[0647] In addition to cell fitness, the amount of gene product necessary to treat disease can also influence the minimal level of therapeutic genome editing that can treat or prevent a disease or a symptom thereof. In cases where a small change in the gene product levels can result in significant changes in clinical outcome, the minimal level of therapeutic genome editing is less relative to cases where a larger change in the gene product levels is needed to

gain a clinically relevant response. In one embodiment, the minimal level of therapeutic genome editing can range from 0.1 to 1 %, 1-5%, 5-10%, 10-15%, 15-20%, 20-25%, 25-30%, 30-35%, 35-40%, 40-45%, 45-50%, or 50-55%. Thus, where a small change in gene product levels can influence clinical outcomes and diseases where there is a fitness advantage for edited cells, are ideal targets for genome editing therapy, as the therapeutic modification threshold is low enough to permit a high chance of success.

[0648] The activity of NHEJ and HDR DSB repair can vary by cell type and cell state. NHEJ is not highly regulated by the cell cycle and is efficient across cell types, allowing for high levels of gene disruption in accessible target cell populations. In contrast, HDR acts primarily during S/G2 phase, and is therefore restricted to cells that are actively dividing, limiting treatments that require precise genome modifications to mitotic cells [Ciccia, A. & Elledge, S.J. *Molecular cell* 40, 179-204 (2010); Chapman, J.R., et al. *Molecular cell* 47, 497-510 (2012)].

[0649] The efficiency of correction via HDR may be controlled by the epigenetic state or sequence of the targeted locus, or the specific repair template configuration (single vs. double stranded, long vs. short homology arms) used [Hacein-Bey-Abina, S., et al. *The New England journal of medicine* 346, 1185-1193 (2002); Gaspar, H.B., et al. *Lancet* 364, 2181-2187 (2004); Beumer, K.J., et al. *G3* (2013)]. The relative activity of NHEJ and HDR machineries in target cells may also affect gene correction efficiency, as these pathways may compete to resolve DSBs [Beumer, K.J., et al. *Proceedings of the National Academy of Sciences of the United States of America* 105, 19821-19826 (2008)]. HDR also imposes a delivery challenge not Seen with NHEJ strategies, as it uses the concurrent delivery of nucleases and repair templates. Thus, these differences can be kept in mind when designing, optimizing, and/or selecting a Type V Cas polypeptide based therapeutic as described in greater detail elsewhere herein.

[0650] Type V Cas polypeptide-based polynucleotide modification application can include combinations of proteins, small RNA molecules, and/or repair templates, and can make, In one embodiment, delivery of these multiple parts substantially more challenging than, for example, traditional small molecule therapeutics. Two main strategies for delivery of compositions, systems, and components thereof have been developed: ex vivo and in vivo. In one embodiment of ex vivo treatments, diseased cells are removed from a subject, edited and then transplanted back into the patient. In other embodiments, cells from a healthy allogeneic donor are collected, modified using a composition or component thereof, to impart various

functionalities and/or reduce immunogenicity, and administered to an allogeneic recipient in need of treatment. Ex vivo editing has the advantage of allowing the target cell population to be well defined and the specific dosage of therapeutic molecules delivered to cells to be specified. The latter consideration may be particularly important when off-target modifications are a concern, as titrating the amount of nuclease may decrease such mutations (Hsu et al., 2013). Another advantage of ex vivo approaches is the typically high editing rates that can be achieved, due to the development of efficient delivery systems for proteins and nucleic acids into cells in culture for research and gene therapy applications.

[0651] In vivo polynucleotide modification via compositions, systems, and/or components thereof involves direct delivery of the compositions, systems, and/or components thereof to cell types in their native tissues. In vivo polynucleotide modification via compositions, systems, and/or components thereof allows diseases in which the affected cell population is not amenable to ex vivo manipulation to be treated. Furthermore, delivering compositions, systems, and/or components thereof to cells in situ allows for the treatment of multiple tissue and cell types.

[0652] In one embodiment, such as those where viral vector systems are used to generate viral particles to deliver the composition and/or component thereof to a cell, the total cargo size of the composition and/or component thereof should be considered as vector systems can have limits on the size of a polynucleotide that can be expressed therefrom and/or packaged into cargo inside of a viral particle. In one embodiment, the tropism of a vector system, such as a viral vector system, should be considered as it can impact the cell type to which the composition or component thereof can be efficiently and/or effectively delivered.

[0653] When delivering a system or component thereof via a viral-based system, it can be important to consider the amount of viral particles that will be needed to achieve a therapeutic effect so as to account for the potential immune response that can be elicited by the viral particles when delivered to a subject or cell(s). When delivering a system or component thereof via a viral based system, it can be important to consider mechanisms of controlling the distribution and/or dosage of the system in vivo. Generally, to reduce the potential for off-target effects, it is optimal but not necessarily required, that the amount of the system be as close to the minimum or least effective dose. In practice this can be challenging to do.

[0654] In one embodiment, it can be important to consider the immunogenicity of the system or component thereof. In embodiments, where the immunogenicity of the system or

component thereof is of concern, the immunogenicity system or component thereof can be reduced. By way of example only, the immunogenicity of the system or component thereof can be reduced using the approach set out in Tangri et al. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the Type V Cas polypeptide in the host species (human or other species).

Therapeutic Agent Development

[0655] The compositions, systems, and components thereof disclosed herein can be used to develop Type V Cas polypeptide-based biologically active agents, such as small molecule therapeutics. Thus, disclosed herein are methods for developing a biologically active agent that modulates a cell function and/or signaling event associated with a disease and/or disease gene. In one embodiment, the method comprises (a) contacting a test compound with a diseased cell and/or a cell containing a disease gene cell; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event or other cell functionality associated with said disease or disease gene, thereby developing said biologically active agent that modulates said cell signaling event or other functionality associated with said disease gene. In one embodiment, the diseased cell is a model cell described elsewhere herein. In one embodiment, the diseased cell is a diseased cell isolated from a subject in need of treatment. In one embodiment, the test compound is a small molecule agent. In one embodiment, test compound is a small molecule agent. In one embodiment, the test compound is a biologic molecule agent.

[0656] In one embodiment, the method involves developing a therapeutic based on the composition, system, disclosed herein. In particular embodiments, the therapeutic comprises a Type V Cas polypeptide and/or a guide molecule with a reprogrammable spacer capable of hybridizing to a target sequence of interest. In particular embodiments, the therapeutic is a vector or vector system that can contain a) a first regulatory element operably linked to a nucleotide sequence encoding the Type V Cas polypeptide; and b) a second regulatory element operably linked to one or more nucleotide sequences encoding one or more nucleic acid molecules comprising a guide molecule comprising a reprogrammable spacer sequence, a conserved RNA sequence; wherein components (a) and (b) are located on same or different vectors. In particular embodiments, the biologically active agent is a composition comprising a delivery system operably configured to deliver composition, system, or components thereof, and/or or one or more polynucleotide sequences, vectors, or vector systems containing or

encoding said components into a cell and capable of forming a complex with the components of the composition and system herein, and wherein said complex is operable in the cell. In one embodiment, the complex can include the Type V Cas polypeptide as disclosed herein, guide molecule scaffold comprising the guide sequence (reprogrammable spacer sequence), and a conserved nucleotide sequence. In any such compositions, the delivery system can be a yeast system, a lipofection system, a microinjection system, a biolistic system, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates or artificial virions, or any other system as disclosed herein. In particular embodiments, the delivery is via a particle, a nanoparticle, a lipid or a cell penetrating peptide (CPP).

[0657] Also disclosed herein are methods for developing or designing a composition, system, optionally a composition, system, based therapy or therapeutic, comprising (a) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest guide molecule target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, and optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more guide molecule recognizing one or more of said (sub)selected target sites.

[0658] In one embodiment, the method for developing or designing a guide molecule for use in a composition, system, optionally a composition, system, based therapy or therapeutic, can include (a) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest guide molecule target sites, wherein a guide molecule directed against said target

sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more guide molecule recognizing one or more of said (sub)selected target sites.

[0659] In one embodiment, the method for developing or designing a composition, system, optionally a composition, system, based therapy or therapeutic in a population, can include (a) selecting for a (therapeutic) locus of interest reprogrammable spacer target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest guide molecule reprogrammable spacer target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest guide molecule reprogrammable spacer target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected target sites for an individual subject, optionally designing one or more guide molecule recognizing one or more of said (sub)selected target sites.

[0660] In one embodiment the method for developing or designing a guide molecule for use in a composition, system, optionally a composition, system, based therapy or therapeutic in a population, can include (a) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, and from said selected target sites subselecting target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, or (b) selecting for a (therapeutic) locus of interest guide molecule target sites, wherein said target sites have minimal sequence variation across a population, or selecting for a (therapeutic) locus of interest guide molecule target sites, wherein a guide molecule directed against said target sites recognizes a minimal number of off-target sites across said population, and optionally estimating the number of (sub)selected target sites needed to treat or otherwise modulate or manipulate a population, optionally validating one or more of the (sub)selected

target sites for an individual subject, optionally designing one or more guide molecule reprogrammable spacer recognizing one or more of said (sub)selected target sites.

[0661] In one embodiment, the method for developing or designing a composition, system, such as a composition, system, based therapy or therapeutic, optionally in a population; or for developing or designing a guide molecule reprogrammable spacer for use in a composition, system, optionally a composition, system, based therapy or therapeutic, optionally in a population, can include selecting a set of target sequences for one or more loci in a target population, wherein the target sequences do not contain variants occurring above a threshold allele frequency in the target population (i.e. platinum target sequences); removing from said selected (platinum) target sequences any target sequences having high frequency off-target candidates (relative to other (platinum) targets in the set) to define a final target sequence set; preparing one or more, such as a set of compositions, systems, based on the final target sequence set, optionally wherein a number of composition prepared is based (at least in part) on the size of a target population.

[0662] In one embodiment, off-target candidates/off-targets, TAMrestrictiveness, target cleavage efficiency, or effector protein specificity is identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as disclosed herein elsewhere. In one embodiment, off-target candidates/off-targets are identified or determined using a sequencing-based double-strand break (DSB) detection assay, such as disclosed herein elsewhere. In one embodiment, off-targets, or off target candidates have at least 1, preferably 1-3, mismatches or (distal) TAMmismatches, such as 1 or more, such as 1, 2, 3, or more (distal) TAMmismatches. In one embodiment, sequencing-based DSB detection assay comprises labeling a site of a DSB with an adapter comprising a primer binding site, labeling a site of a DSB with a barcode or unique molecular identifier, or combination thereof, as disclosed herein elsewhere.

[0663] It will be understood that the reprogrammable spacer sequence of the guide molecule is 100% complementary to the target site, i.e., does not comprise any mismatch with the target site. It will be further understood that “recognition” of an (off-)target site by a reprogrammable spacer presupposes composition, system, functionality, i.e., an (off-)target site is only recognized by a reprogrammable spacer RNA if binding of the reprogrammable spacer RNA to the (off-)target site leads to composition, system, activity (such as induction of single or double strand DNA cleavage, transcriptional modulation, etc.).

[0664] In one embodiment, the target sites having minimal sequence variation across a population are characterized by absence of sequence variation in at least 99%, preferably at least 99.9%, more preferably at least 99.99% of the population. In one embodiment, optimizing target location comprises selecting target sequences or loci having an absence of sequence variation in at least 99%, preferably at least 99.9%, more preferably at least 99.99% of a population. These targets are referred to herein elsewhere also as “platinum targets”. In one embodiment, said population comprises at least 1000 individuals, such as at least 5000 individuals, such as at least 10000 individuals, such as at least 50000 individuals.

[0665] In one embodiment, the off-target sites are characterized by at least one mismatch between the off-target site and the guide molecule. In one embodiment, the off-target sites are characterized by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the guide molecule. In one embodiment, the off-target sites are characterized by at least one mismatch between the off-target site and the guide molecule and by at most five, preferably at most four, more preferably at most three mismatches between the off-target site and the guide molecule.

[0666] In one embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes in said population. In one embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the off-target site locus in said population. In one embodiment, said minimal number of off-target sites across said population is determined for high-frequency haplotypes of the target site locus in said population. In one embodiment, the high-frequency haplotypes are characterized by occurrence in at least 0.1% of the population.

[0667] In one embodiment, the number of (sub)selected target sites needed to treat a population is estimated based on based low frequency sequence variation, such as low frequency sequence variation captured in large scale sequencing datasets. In one embodiment, the number of (sub)selected target sites needed to treat a population of a given size is estimated.

[0668] In one embodiment, the method further comprises obtaining genome sequencing data of a subject to be treated; and treating the subject with a composition, system, selected from the set of compositions, systems, wherein the composition, system, selected is based (at least in part) on the genome sequencing data of the individual. In one embodiment, the ((sub)selected) target is validated by genome sequencing, preferably whole genome sequencing.

[0669] In one embodiment, target sequences or loci as disclosed herein are (further) selected based on optimization of one or more parameters, such as TAMtype (natural or modified), TAMnucleotide content, TAMlength, target sequence length, TAMrestrictiveness, target cleavage efficiency, and target sequence position within a gene, a locus or other genomic region. Methods of optimization are discussed in greater detail elsewhere herein.

[0670] In one embodiment, target sequences or loci as disclosed herein are (further) selected based on optimization of one or more of target loci location, target length, target specificity, and TAMcharacteristics. As used herein, unless otherwise indicated, TAMcharacteristics may comprise for instance TAMsequence, TAMlength, and/or TAM GC contents. In one embodiment, optimizing TAMcharacteristics comprises optimizing nucleotide content of aTAM. In one embodiment, optimizing nucleotide content of TAMis selecting a TAMwith a motif that maximizes abundance in the one or more target loci, minimizes mutation frequency, or both. Minimizing mutation frequency can for instance be achieved by selecting TAMsequences devoid of or having low or minimal CpG.

[0671] In one embodiment, the effector protein for each composition and system, in the set of compositions, systems, is selected based on optimization of one or more parameters selected from the group consisting of; effector protein size, ability of effector protein to access regions of high chromatin accessibility, degree of uniform enzyme activity across genomic targets, epigenetic tolerance, mismatch/budge tolerance, effector protein specificity, effector protein stability or half-life, effector protein immunogenicity or toxicity. Methods of optimization are discussed in greater detail elsewhere herein.

Exemplary Applications in Plants and Fungi

[0672] The compositions, systems, and methods disclosed herein can be used to perform gene or genome interrogation or editing or manipulation in plants and fungi. For example, the applications include investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes, e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant or fungus genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The compositions, systems, and methods can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques.

[0673] The compositions, systems, and methods herein may be used to confer desired traits (e.g., enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds) on essentially any plants and fungi, and their cells and tissues. The compositions, systems, and methods may be used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of any foreign gene.

[0674] In one embodiment, compositions, systems, and methods may be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; *See, e.g.*, Nekrasov, “Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system,” *Plant Methods* 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, “Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system,” *Plant Physiology* September 2014 pp 114.247577; Shan, “Targeted genome modification of crop plants using a CRISPR-Cas system,” *Nature Biotechnology* 31, 686-688 (2013); Feng, “Efficient genome editing in plants using a CRISPR/Cas system,” *Cell Research* (2013) 23:1229–1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, “RNA-guided genome editing in plants using a CRISPR-Cas system,” *Mol Plant*. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, “Gene targeting using the *Agrobacterium tumefaciens*-mediated CRISPR-Cas system in rice,” *Rice* 2014, 7:5 (2014), Zhou et al., “Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate: CoA ligase specificity and Redundancy,” *New Phytologist* (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliendo et al, “Targeted DNA degradation using a CRISPR device stably carried in the host genome, *NATURE COMMUNICATIONS* 6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - *Agrobacterium*-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, Morrell et al “Crop genomics: advances and applications,” *Nat Rev Genet*. 2011 Dec 29;13(2):85-96, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. Aspects of utilizing the compositions, systems, and methods may be analogous to the use of the composition in plants, and mention is made of the University of Arizona website “CRISPR-PLANT” (genome.arizona.edu/crispr/) (supported by Penn State and AGI) for directions to nucleic acid

modification in plant systems. This approach can be adapted to and/or applied to the Type V Cas compositions, systems, disclosed herein.

[0675] The compositions, systems, and methods may also be used on protoplasts. A “protoplast” refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate, and regenerate grow into a whole plant under proper growing conditions.

[0676] The compositions, systems, and methods may be used for screening genes (e.g., endogenous, mutations) of interest. In some examples, genes of interest include those encoding enzymes involved in the production of a component of added nutritional value or generally genes affecting agronomic traits of interest, across species, phyla, and plant kingdom. By selectively targeting e.g., genes encoding enzymes of metabolic pathways, the genes responsible for certain nutritional aspects of a plant can be identified. Similarly, by selectively targeting genes which may affect a desirable agronomic trait, the relevant genes can be identified. Accordingly, the present invention encompasses screening methods for genes encoding enzymes involved in the production of compounds with a particular nutritional value and/or agronomic traits.

[0677] It is also understood that reference herein to animal cells may also apply, *mutatis mutandis*, to plant or fungal cells unless otherwise apparent; and the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein.

[0678] In some cases, nucleic acids introduced to plants and fungi may be codon optimized for expression in the plants and fungi. Methods of codon optimization include those described in Kwon KC, et al., Codon Optimization to Enhance Expression Yields Insights into Chloroplast Translation, Plant Physiol. 2016 Sep;172(1):62-77.

[0679] The components (e.g., Type V Cas polypeptide) in the compositions and systems may further comprise one or more functional domains disclosed herein. In some examples, the functional domains may be an exonuclease. Such exonuclease may increase the efficiency of the Type V Cas polypeptide’ function, e.g., mutagenesis efficiency. An example of the functional domain is Trex2, as described in Weiss T et al., [biorxiv.org/content/10.1101/2020.04.11.037572v1](https://doi.org/10.1101/2020.04.11.037572), doi.org/10.1101/2020.04.11.037572.

Examples of plants

[0680] The compositions, systems, and methods herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics. In general, the term “plant” relates to any various photosynthetic, eukaryotic, unicellular, or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants.

[0681] The compositions, systems, and methods may be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders *Magnoliales*, *Illiciales*, *Laurales*, *Piperales*, *Aristolochiales*, *Nymphaeales*, *Ranunculales*, *Papaverales*, *Sarraceniales*, *Trochodendrales*, *Hamamelidales*, *Eucomiales*, *Leitneriales*, *Myricales*, *Fagales*, *Casuarinales*, *Caryophyllales*, *Batales*, *Polygonales*, *Plumbaginales*, *Dilleniales*, *Theales*, *Malvales*, *Urticales*, *Lecythidales*, *Violales*, *Salicales*, *Capparales*, *Ericales*, *Diapensales*, *Ebenales*, *Primulales*, *Rosales*, *Fabales*, *Podostemales*, *Haloragales*, *Myrtales*, *Cornales*, *Proteales*, *Santalales*, *Rafflesiales*, *Celastrales*, *Euphorbiales*, *Rhamnales*, *Sapindales*, *Juglandales*, *Geraniales*, *Polygalales*, *Umbellales*, *Gentianales*, *Polemoniales*, *Lamiales*, *Plantaginales*, *Scrophulariales*, *Campanulales*, *Rubiales*, *Dipsacales*, and *Asterales*; monocotyledonous plants such as those belonging to the orders *Alismatales*, *Hydrocharitales*, *Najadales*, *Triuridales*, *Commelinales*, *Eriocaulales*, *Restionales*, *Poales*, *Juncals*, *Cyperales*, *Typhales*, *Bromeliales*, *Zingiberales*, *Arecales*, *Cyclanthales*, *Pandanals*, *Arales*, *Lilliales*, and *Orchidales*, or with plants belonging to *Gymnospermae*, e.g., those belonging to the orders *Pinales*, *Ginkgoales*, *Cycadales*, *Araucariales*, *Cupressales* and *Gnetales*.

[0682] The compositions, systems, and methods herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa*, *Alseodaphne*, *Anacardium*, *Arachis*, *Beilschmiedia*, *Brassica*, *Carthamus*, *Cocculus*, *Croton*, *Cucumis*, *Citrus*, *Citrullus*, *Capsicum*, *Catharanthus*, *Cocos*, *Coffea*, *Cucurbita*, *Daucus*, *Duguetia*, *Eschscholzia*, *Ficus*, *Fragaria*, *Glaucium*, *Glycine*, *Gossypium*, *Helianthus*, *Hevea*, *Hyoscyamus*, *Lactuca*, *Landolphia*, *Limum*, *Litsea*, *Lycopersicon*, *Lupinus*, *Manihot*, *Majorana*, *Malus*, *Medicago*, *Nicotiana*, *Olea*, *Parthenium*, *Papaver*, *Persea*, *Phaseolus*, *Pistacia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Senecio*, *Sinomenium*, *Stephania*, *Sinapis*, *Solanum*, *Theobroma*, *Trifolium*, *Trigonella*, *Vicia*, *Vinca*, *Vitis*, and

Vigna; and the genera *Allium*, *Andropogon*, *Aragrostis*, *Asparagus*, *Avena*, *Cynodon*, *Elaeis*, *Festuca*, *Festulolium*, *Heterocallis*, *Hordeum*, *Lemna*, *Lolium*, *Musa*, *Oryza*, *Panicum*, *Pannisetum*, *Phleum*, *Poa*, *Secale*, *Sorghum*, *Triticum*, *Zea*, *Abies*, *Cunninghamia*, *Ephedra*, *Picea*, *Pinus*, and *Pseudotsuga*.

[0683] In one embodiment, target plants and plant cells for engineering include those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape Seed) and plants used for experimental purposes (e.g., *Arabidopsis*). Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticales, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini.

[0684] The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants. The compositions, systems, and methods can be used over a broad range of "algae" or "algae cells." Examples of algae include eukaryotic phyla, including the *Rhodophyta* (red algae), *Chlorophyta* (green algae), *Phaeophyta* (brown algae), *Bacillariophyta* (diatoms), *Eustigmatophyta* and dinoflagellates as well as the prokaryotic phylum *Cyanobacteria* (blue-

green algae). Examples of algae species include those of *Amphora*, *Anabaena*, *Anikstrodesmis*, *Botryococcus*, *Chaetoceros*, *Chlamydomonas*, *Chlorella*, *Chlorococcum*, *Cyclotella*, *Cylindrotheca*, *Dunaliella*, *Emiliana*, *Euglena*, *Hematococcus*, *Isochrysis*, *Monochrysis*, *Monoraphidium*, *Nannochloris*, *Nannochloropsis*, *Navicula*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Nodularia*, *Nostoc*, *Oochromonas*, *Oocystis*, *Oscillatoria*, *Pavlova*, *Phaeodactylum*, *Playtmonas*, *Pleurochrysis*, *Porhyra*, *Pseudoanabaena*, *Pyramimonas*, *Stichococcus*, *Synechococcus*, *Synechocystis*, *Tetraselmis*, *Thalassiosira*, and *Trichodesmium*.

Plant promoters

[0685] In order to ensure appropriate expression in a plant cell, the components of the components and systems herein may be placed under control of a plant promoter. A plant promoter is a promoter operable in plant cells. A plant promoter is capable of initiating transcription in plant cells, whether or not its origin is a plant cell. The use of different types of promoters is envisaged.

[0686] In some examples, the plant promoter is a constitutive plant promoter, which is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. In some examples, the plant promoter is a regulated promoter, which directs gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In some examples, the plant promoter is a tissue-preferred promoters, which can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the Seed.

[0687] Exemplary plant promoters include those obtained from plants, plant viruses, and bacteria such as *Agrobacterium* or *Rhizobium* which comprise genes expressed in plant cells. Additional examples of promoters include those described in Kawamata et al., (1997) *Plant Cell Physiol* 38:792-803; Yamamoto et al., (1997) *Plant J* 12:255-65; Hire et al, (1992) *Plant Mol Biol* 20:207-18, Kuster et al, (1995) *Plant Mol Biol* 29:759-72, and Capana et al., (1994) *Plant Mol Biol* 25:681 -91.

[0688] In some examples, a plant promoter may be an inducible promoter, which is inducible and allows for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome), such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. In a particular example, of the components of a light inducible system include a Type V Cas polypeptide, a light-responsive cytochrome heterodimer (e.g., from *Arabidopsis thaliana*), and a transcriptional activation/repression domain.

[0689] In some examples, the promoter may be a chemical-regulated promoter (where the application of an exogenous chemical induces gene expression) or a chemical-repressible promoter (where application of the chemical represses gene expression). Examples of chemical-inducible promoters include maize In2-2 promoter (activated by benzene sulfonamide herbicide safeners), the maize GST promoter (activated by hydrophobic electrophilic compounds used as pre-emergent herbicides), the tobacco PR-1 a promoter (activated by salicylic acid), promoters regulated by antibiotics (such as tetracycline-inducible and tetracycline-repressible promoters).

[0690] Vectors and other regulatory polynucleotides, tags, reporters, and/or the like are described elsewhere herein and, in some embodiments, can be utilized in and/or to produce a modified plant disclosed herein.

Stable integration in the genome of plants

[0691] In one embodiment, polynucleotides encoding the components of the compositions and systems may be introduced for stable integration into the genome of a plant cell. In some cases, vectors or expression systems may be used for such integration. The design of the vector or the expression system can be adjusted depending on for when, where and under what conditions the guide molecule and/or the Type V Cas polypeptide gene are expressed. In some cases, the polynucleotides may be integrated into an organelle of a plant, such as a plastid, mitochondrion, or a chloroplast. The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

[0692] In one embodiment, the method of integration generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom. In some examples, the expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the RNA and/or Type V Cas polypeptide in a plant cell; a 5' untranslated region to enhance expression; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the guide molecule and/or the Type V Cas polypeptide gene sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

Transient expression in plants

[0693] In one embodiment, the components of the compositions and systems may be transiently expressed in the plant cell. In some examples, the compositions and systems may modify a target polynucleotide only when both the guide molecule and the Type V Cas polypeptide are present in a cell, such that genomic modification can further be controlled. As the expression of the Type V Cas polypeptide is transient, plants regenerated from such plant cells typically contain no foreign DNA. In certain examples, the Type V Cas polypeptide is stably expressed, and the guide molecule sequence is transiently expressed.

[0694] DNA and/or RNA (e.g., mRNA) may be introduced to plant cells for transient expression. In such cases, the introduced nucleic acid may be provided in sufficient quantity to modify the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions.

[0695] The transient expression may be achieved using suitable vectors. Exemplary vectors that may be used for transient expression include a pEAQ vector (may be tailored for Agrobacterium-mediated transient expression) and Cabbage Leaf Curl virus (CaLCuV), and vectors described in Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93; and Yin K et al., Scientific Reports volume 5, Article number: 14926 (2015).

[0696] Combinations of the different methods described above are also envisaged.

Translocation to and/or expression in specific plant organelles

[0697] The compositions and systems herein may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[0698] In one embodiment, it is envisaged that the compositions and systems are used to specifically modify chloroplast genes or to ensure expression in the chloroplast. The compositions and systems (e.g., Type V Cas polypeptide, guide molecules, or their encoding polynucleotides) may be transformed, compartmentalized, and/or targeted to the chloroplast. In an example, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[0699] Examples of methods of chloroplast transformation include Particle bombardment, PEG treatment, and microinjection, and the translocation of transformation cassettes from the nuclear genome to the plastid. In some examples, targeting of chloroplasts may be achieved by incorporating in chloroplast localization sequence, and/or the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the components of the compositions and systems. Additional examples of transforming, targeting and localization of chloroplasts include those described in WO2010061186, Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180, and US 20040142476, which are incorporated by reference herein in their entireties.

Modification of polyploid plants

[0700] The compositions, systems, and methods may be used to modify polyploid plants. Polyploid plants carry duplicate copies of their genomes (e.g., as many as six, such as in wheat). In some cases, the compositions, systems, and methods can be multiplexed to affect all copies of a gene, or to target dozens of genes at once. For instance, the compositions, systems, and methods may be used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defenses against a disease. The modification may be simultaneous suppression the expression of the TaMLO-A1, TaMLO-B1 and TaMLO-D1 nucleic acid sequence in a wheat plant cell and regenerating a wheat plant therefrom, in order to ensure that the wheat plant is resistant to powdery mildew (e.g., as described in WO2015109752).

Plant cultures and regeneration

[0701] In one embodiment, the modified plants or plant cells may be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Examples of regeneration techniques include those relying on manipulation of certain phytohormones in a tissue culture growth medium, relying on a biocide and/or herbicide

marker which has been introduced together with the desired nucleotide sequences, obtaining from cultured protoplasts, plant callus, explants, organs, pollens, embryos, or parts thereof.

Exemplary applications in plants

[0702] The compositions, systems, and methods may be used to generate genetic variation(s) in a plant (e.g., crop) of interest. One or more, e.g., a library of, guide molecules targeting one or more locations in a genome may be provided and introduced into plant cells together with the Type V Cas polypeptide. For example, a collection of genome-scale point mutations and gene knock-outs can be generated. In some examples, the compositions, systems, and methods may be used to generate a plant part or plant from the cells so obtained and screening the cells for a trait of interest. The target genes may include both coding and non-coding regions. In some cases, the trait is stress tolerant and the method is a method for the generation of stress-tolerant crop varieties.

[0703] In one embodiment, the compositions, systems, and methods are used to modify endogenous genes or to modify their expression. The expression of the components may induce targeted modification of the genome, either by direct activity of the Type V Cas polypeptide and optionally introduction of recombination template DNA, or by modification of genes targeted. The different strategies disclosed herein above allow Type V Cas polypeptide-mediated targeted genome editing without requiring the introduction of the components into the plant genome.

[0704] In some cases, the modification may be performed without the permanent introduction into the genome of the plant of any foreign gene, including those encoding components of the composition herein, so as to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous. Components which are transiently introduced into the plant cell are typically removed upon crossing.

[0705] For example, the modification may be performed by transient expression of the components of the compositions and systems. The transient expression may be performed by delivering the components of the compositions and systems with viral vectors, delivery into protoplasts, with the aid of particulate molecules such as nanoparticles or CPPs.

Generation of plants with desired traits

[0706] The compositions, systems, and methods herein may be used to introduce desired traits to plants. The approaches include introduction of one or more foreign genes to confer a trait of interest, editing or modulating endogenous genes to confer a trait of interest.

Agronomic traits

[0707] In one embodiment, crop plants can be improved by influencing specific plant traits. Examples of the traits include improved agronomic traits such as herbicide resistance, disease resistance, abiotic stress tolerance, high yield, and superior quality, pesticide-resistance, disease resistance, insect and nematode resistance, resistance against parasitic weeds, drought tolerance, nutritional value, stress tolerance, self-pollination avoidance, forage digestibility biomass, and grain yield.

[0708] In one embodiment, genes that confer resistance to pests or diseases may be introduced to plants. In cases there are endogenous genes that confer such resistance in a plants, their expression and function may be enhanced (e.g., by introducing extra copies, modifications that enhance expression and/or activity).

[0709] Examples of genes that confer resistance include plant disease resistance genes (e.g., Cf- 9, Pto, RSP2, SIDMR6-1), genes conferring resistance to a pest (e.g., those described in WO96/30517), *Bacillus thuringiensis* proteins, lectins, Vitamin-binding proteins (e.g., avidin), enzyme inhibitors (e.g., protease or proteinase inhibitors or amylase inhibitors), insect-specific hormones or pheromones (e.g., ecdysteroid or a juvenile hormone, variant thereof, a mimetic based thereon, or an antagonist or agonist thereof) or genes involved in the production and regulation of such hormone and pheromones, insect-specific peptides or neuropeptide, Insect-specific venom (e.g., produced by a snake, a wasp, etc., or analog thereof), Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity, Enzymes involved in the modification of biologically active molecule (e.g., a glycolytic enzyme, a proteolytic enzyme, a lipolytic enzyme, a nuclease, a cyclase, a transaminase, an esterase, a hydrolase, a phosphatase, a kinase, a phosphorylase, a polymerase, an elastase, a chitinase and a glucanase, whether natural or synthetic), molecules that stimulates signal transduction, Viral-invasive proteins or a complex toxin derived therefrom, Developmental-arrestive proteins produced in nature by a pathogen or a parasite, a developmental-arrestive protein produced in nature by a plant, or any combination thereof.

[0710] The compositions, systems, and methods may be used to identify, screen, introduce or remove mutations or sequences lead to genetic variability that give rise to susceptibility to certain pathogens, e.g., host specific pathogens. Such approach may generate plants that are non-host resistance, e.g., the host and pathogen are incompatible or there can be partial resistance against all races of a pathogen, typically controlled by many genes and/or also complete resistance to some races of a pathogen but not to other races.

[0711] In one embodiment, the compositions, systems, and methods may be used to modify genes involved in plant diseases. Such genes may be removed, inactivated, or otherwise regulated or modified. Examples of plant diseases include those described in [0045]-[0080] of US20140213619A1, which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins disclosed herein.

[0712] In one embodiment, genes that confer resistance to herbicides may be introduced to plants. Examples of genes that confer resistance to herbicides include genes conferring resistance to herbicides that inhibit the growing point or meristem, such as an imidazolinone or a sulfonylurea, genes conferring glyphosate tolerance (e.g., resistance conferred by, e.g., mutant 5-enolpyruvylshikimate-3- phosphate synthase genes, *aroA* genes and glyphosate acetyl transferase (GAT) genes, respectively), or resistance to other phosphono compounds such as by glufosinate (phosphinothricin acetyl transferase (PAT) genes from *Streptomyces* species, including *Streptomyces hygroscopicus* and *Streptomyces viridichromogenes*), and to pyridinoxy or phenoxy propionic acids and cyclohexones by ACCase inhibitor-encoding genes), genes conferring resistance to herbicides that inhibit photosynthesis (such as a triazine (*psbA* and *gs+* genes) or a benzonitrile (nitrilase gene), and glutathione S-transferase), genes encoding enzymes detoxifying the herbicide or a mutant glutamine synthase enzyme that is resistant to inhibition, genes encoding a detoxifying enzyme is an enzyme encoding a phosphinothricin acetyltransferase (such as the *bar* or *pat* protein from *Streptomyces* species), genes encoding hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, e.g., naturally occurring HPPD resistant enzymes, and genes encoding a mutated or chimeric HPPD enzyme.

[0713] In one embodiment, genes involved in Abiotic stress tolerance may be introduced to plants. Examples of genes include those capable of reducing the expression and/or the activity of poly(ADP-ribose) polymerase (PARP) gene, transgenes capable of reducing the expression and/or the activity of the PARG encoding genes, genes coding for a plant-functional enzyme of the nicotineamide adenine dinucleotide salvage synthesis pathway including

nicotinamidase, nicotinate phosphoribosyltransferase, nicotinic acid mononucleotide adenylyl transferase, nicotinamide adenine dinucleotide synthetase or nicotine amide phosphoribosyltransferase, enzymes involved in carbohydrate biosynthesis, enzymes involved in the production of polyfructose (e.g., the inulin and levan-type), the production of alpha-1,6 branched alpha-1,4-glucans, the production of alternan, the production of hyaluronan.

[0714] In one embodiment, genes that improve drought resistance may be introduced to plants. Examples of genes Ubiquitin Protein Ligase protein (UPL) protein (UPL3), DR02, DR03, ABC transporter, and DREB1A.

Nutritionally improved plants

[0715] In one embodiment, the compositions, systems, and methods may be used to produce nutritionally improved plants. In some examples, such plants may provide functional foods, e.g., a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains. In certain examples, such plants may provide nutraceuticals foods, e.g., substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. The nutraceutical foods may be useful in the prevention and/or treatment of diseases in animals and humans, e.g., cancers, diabetes, cardiovascular disease, and hypertension.

[0716] An improved plant may naturally produce one or more desired compounds and the modification may enhance the level or activity or quality of the compounds. In some cases, the improved plant may not naturally produce the compound(s), while the modification enables the plant to produce such compound(s). In some cases, the compositions, systems, and methods used to modify the endogenous synthesis of these compounds indirectly, e.g. by modifying one or more transcription factors that controls the metabolism of this compound.

[0717] Examples of nutritionally improved plants include plants comprising modified protein quality, content and/or amino acid composition, essential amino acid contents, oils and fatty acids, carbohydrates, vitamins and carotenoids, functional secondary metabolites, and minerals. In some examples, the improved plants may comprise or produce compounds with health benefits. Examples of nutritionally improved plants include those described in Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939–953.

[0718] Examples of compounds that can be produced include carotenoids (e.g., α -Carotene or β -Carotene), lutein, lycopene, Zeaxanthin, Dietary fiber (e.g., insoluble fibers, β -Glucan, soluble fibers, fatty acids (e.g., ω -3 fatty acids, Conjugated linoleic acid, GLA,), Flavonoids

(e.g., Hydroxycinnamates, flavonols, catechins and tannins), Glucosinolates, indoles, isothiocyanates (e.g., Sulforaphane), Phenolics (e.g., stilbenes, caffeic acid and ferulic acid, epicatechin), Plant stanols/sterols, Fructans, inulins, fructo-oligosaccharides, Saponins, Soybean proteins, Phytoestrogens (e.g., isoflavones, lignans), Sulfides and thiols such as diallyl sulphide, Allyl methyl trisulfide, dithiolthiones, Tannins, such as proanthocyanidins, or any combination thereof.

[0719] The compositions, systems, and methods may also be used to modify protein/starch functionality, shelf life, taste/aesthetics, fiber quality, and allergen, antinutrient, and toxin reduction traits.

[0720] Examples of genes and nucleic acids that can be modified to introduce the traits include stearyl-ACP desaturase, DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid, Tf RAP2.2 and its interacting partner SINAT2, Tf Dof1, and DOF Tf AtDof1.1 (OBP2).

Regulation of fruit-ripening

[0721] The compositions, systems, and methods may be used to regulate ripening of fruits. Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it may render a fruit or vegetable inedible, which can bring significant losses to both farmers and consumers.

[0722] In one embodiment, the compositions, systems, and methods are used to reduce ethylene production. In some examples, the compositions, systems, and methods may be used to suppress the expression and/or activity of ACC synthase, insert an ACC deaminase gene or a functional fragment thereof, insert a SAM hydrolase gene or functional fragment thereof, suppress ACC oxidase gene expression.

[0723] Alternatively or additionally, the compositions, systems, and methods may be used to modify ethylene receptors (e.g., suppressing ETR1) and/or Polygalacturonase (PG). Suppression of a gene may be achieved by introducing a mutation, an antisense sequence, and/or a truncated copy of the gene to the genome.

Increasing storage life of plants

[0724] In one embodiment, the compositions, systems, and methods are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. The modification may be in a gene that prevents the accumulation of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino

acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In particular embodiments, the methods provided herein are used to reduce or inhibit expression of the vacuolar invertase gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose.

Reducing allergens in plants

[0725] In one embodiment, the compositions, systems, and methods are used to generate plants with a reduced level of allergens, making them safer for consumers. To this end, the compositions, systems, and methods may be used to identify and modify (e.g., suppress) one or more genes responsible for the production of plant allergens. Examples of such genes include Lol p5, as well as those in peanuts, soybeans, lentils, peas, lupin, green beans, mung beans, such as those described in Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222), which is incorporated by reference herein in its entirety and can be adapted for use with the engineered Type V Cas proteins disclosed herein.

Generation of male sterile plants

[0726] The compositions, systems, and methods may be used to generate male sterile plants. Hybrid plants typically have advantageous agronomic traits compared to inbred plants. However, for self-pollinating plants, the generation of hybrids can be challenging. In different plant types (e.g., maize and rice), genes have been identified which are important for plant fertility, more particularly male fertility. Plants that are as such genetically altered can be used in hybrid breeding programs.

[0727] The compositions, systems, and methods may be used to modify genes involved male fertility, e.g., inactivating (such as by introducing mutations to) genes required for male fertility. Examples of the genes involved in male fertility include cytochrome P450-like gene (MS26) or the meganuclease gene (MS45), and those described in Wan X et al., Mol Plant. 2019 Mar 4;12(3):321-342; and Kim YJ, et al., Trends Plant Sci. 2018 Jan;23(1):53-65.

Increasing the fertility stage in plants

[0728] In one embodiment, the compositions, systems, and methods may be used to prolong the fertility stage of a plant such as of a rice. For instance, a rice fertility stage gene such as Ehd3 can be targeted in order to generate a mutation in the gene and plantlets can be selected for a prolonged regeneration plant fertility stage.

Production of early yield of products

[0729] In one embodiment, the compositions, systems, and methods may be used to produce early yield of the product. For example, flowering process may be modulated, e.g., by mutating flowering repressor gene such as SP5G. Examples of such approaches include those described in Soyk S, et al., Nat Genet. 2017 Jan;49(1):162-168.

Oil and biofuel production

[0730] The compositions, systems, and methods may be used to generate plants for oil and biofuel production. Biofuels include fuels made from plant and plant-derived resources. Biofuels may be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. Biofuels include bioethanol and biodiesel. Bioethanol can be produced by the sugar fermentation process of cellulose (starch), which may be derived from maize and sugar cane. Biodiesel can be produced from oil crops such as rapeseed, palm, and soybean. Biofuels can be used for transportation.

Generation of plants for production of vegetable oils and biofuels

[0731] The compositions, systems, and methods may be used to generate algae (e.g., diatom) and other plants (e.g., grapes) that express or overexpress high levels of oil or biofuels.

[0732] In some cases, the compositions, systems, and methods may be used to modify genes involved in the modification of the quantity of lipids and/or the quality of the lipids. Examples of such genes include those involved in the pathways of fatty acid synthesis, e.g., acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl_acyl- carrier protein synthase III, glycerol-3-phosphate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phosphatidate phosphatase, fatty acid thioesterase such as palmitoyl protein thioesterase, or malic enzyme activities.

[0733] In further embodiments, it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolization. Examples of genes include those involved in the activation of triacylglycerol and free fatty

acids, β -oxidation of fatty acids, such as genes of acyl-CoA synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucomutase.

[0734] In some examples, algae may be modified for production of oil and biofuels, including fatty acids (e.g., fatty esters such as acid methyl esters (FAME) and fatty acid ethyl esters (FAEE)). Examples of methods of modifying microalgae include those described in Stovicek et al. *Metab. Eng. Comm.*, 2015; 2:1; US Patent No. 8,945,839; and International Patent Publication No. WO 2015/086795.

[0735] In some examples, one or more genes may be introduced (e.g., overexpressed) to the plants (e.g., algae) to produce oils and biofuels (e.g., fatty acids) from a carbon source (e.g., alcohol). Examples of the genes include genes encoding acyl-CoA synthases, ester synthases, thioesterases (e.g., *tesA*, *'tesA*, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA1*, or *fatA*), acyl-CoA synthases (e.g., *fadD*, *JadK*, *BH3103*, *pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa39*), ester synthases (e.g., synthase/acyl-CoA:diacylglycerol acyltransferase from *Simmondsia chinensis*, *Acinetobacter* sp. ADP, *Alcanivorax borkumensis*, *Pseudomonas aeruginosa*, *Fundibacter jadensis*, *Arabidopsis thaliana*, or *Alkaligenes eutrophus*, or variants thereof).

[0736] Additionally or alternatively, one or more genes in the plants (e.g., algae) may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. Examples of such genes include genes encoding acyl-CoA dehydrogenases (e.g., *fade*), outer membrane protein receptors, and transcriptional regulator (e.g., repressor) of fatty acid biosynthesis (e.g., *fabR*), pyruvate formate lyases (e.g., *pflB*), lactate dehydrogenases (e.g., *IdhA*).

Organic acid production

[0737] In one embodiment, plants may be modified to produce organic acids such as lactic acid. The plants may produce organic acids using sugars, pentose, or hexose sugars. To this end, one or more genes may be introduced (e.g., and overexpressed) in the plants. An example of such genes includes LDH gene.

[0738] In some examples, one or more genes may be inactivated (e.g., expression of the genes is decreased). For examples, one or more mutations may be introduced to the genes. The genes may include those encoding proteins involved an endogenous metabolic pathway which produces a metabolite other than the organic acid of interest and/or wherein the endogenous metabolic pathway consumes the organic acid.

[0739] Examples of genes that can be modified or introduced include those encoding pyruvate decarboxylases (pdc), fumarate reductases, alcohol dehydrogenases (adh), acetaldehyde dehydrogenases, phosphoenolpyruvate carboxylases (ppc), D-lactate dehydrogenases (d-ldh), L-lactate dehydrogenases (l-ldh), lactate 2-monooxygenases, lactate dehydrogenase, cytochrome-dependent lactate dehydrogenases (e.g., cytochrome B2-dependent L-lactate dehydrogenases).

Enhancing plant properties for biofuel production

[0740] In one embodiment, the compositions, systems, and methods are used to alter the properties of the cell wall of plants to facilitate access by key hydrolyzing agents for a more efficient release of sugars for fermentation. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, lignin biosynthesis may be downregulated in the plant so as to increase fermentable carbohydrates.

[0741] In some examples, one or more lignin biosynthesis genes may be down regulated. Examples of such genes include 4-coumarate 3-hydroxylases (C3H), phenylalanine ammonia-lyases (PAL), cinnamate 4-hydroxylases (C4H), hydroxycinnamoyl transferases (HCT), caffeic acid O-methyltransferases (COMT), caffeoyl CoA 3-O-methyltransferases (CCoAOMT), ferulate 5- hydroxylases (F5H), cinnamyl alcohol dehydrogenases (CAD), cinnamoyl CoA-reductases (CCR), 4- coumarate-CoA ligases (4CL), monolignol-lignin-specific glycosyltransferases, and aldehyde dehydrogenases (ALDH), and those described in WO 2008064289.

[0742] In some examples, plant mass that produces lower level of acetic acid during fermentation may be reduced. To this end, genes involved in polysaccharide acetylation (e.g., Cas1L and those described in WO 2010096488) may be inactivated.

Other microorganisms for oils and biofuel production

[0743] In one embodiment, microorganisms other than plants may be used for production of oils and biofuels using the compositions, systems, and methods herein. Examples of the microorganisms include those of the genus of *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Synechococcus*, *Synechoystis*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Khuyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophamonas*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*. This is also further discussed elsewhere herein.

Detecting modifications in the plant genome-selectable markers

[0744] When the compositions, systems, and methods are used to modify a plant, suitable methods may be used to confirm and detect the modification made in the plant. In some examples, when a variety of modifications are made, one or more desired modifications or traits resulting from the modifications may be selected and detected. The detection and confirmation may be performed by biochemical and molecular biology techniques such as Southern analysis, PCR, Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR, enzymatic assays, ribozyme activity, gel electrophoresis, Western blot, immunoprecipitation, enzyme-linked immunoassays, in situ hybridization, enzyme staining, and immunostaining.

[0745] In some cases, one or more markers, such as selectable and detectable markers, may be introduced to the plants. Such markers may be used for selecting, monitoring, isolating cells and plants with desired modifications and traits. A selectable marker can confer positive or negative selection and is conditional or non-conditional on the presence of external substrates. Examples of such markers include genes and proteins that confer resistance to antibiotics, such as hygromycin (hpt) and kanamycin (nptII), and genes that confer resistance to herbicides, such as phosphinothricin (bar) and chlorosulfuron (als), enzyme capable of producing or processing a colored substances (e.g., the β -glucuronidase, luciferase, B or C1 genes).

[0746] Other methods and compositions of detection and screening using the systems and compositions of the present invention are described in greater detail elsewhere herein and can be applied to plants and plant tissues and cells.

Further exemplary applications in plants

[0747] Further applications of the compositions, systems, and methods on plants and fungi include visualization of genetic element dynamics (e.g., as described in Chen B, et al., Cell. 2013 Dec 19;155(7):1479-91), targeted gene disruption positive-selection *in vitro* and *in vivo* (as described in Malina A et al., Genes Dev. 2013 Dec 1;27(23):2602-14), epigenetic modification such as using fusion of Cas polypeptide and histone-modifying enzymes (e.g., as described in Rusk N, Nat Methods. 2014 Jan;11(1):28), identifying transcription regulators (e.g., as described in Waldrip ZJ, Epigenetics. 2014 Sep;9(9):1207-11), anti-virus treatment for both RNA and DNA viruses (e.g., as described in Price AA, et al., Proc Natl Acad Sci U S A. 2015 May 12;112(19):6164-9; Ramanan V et al., Sci Rep. 2015 Jun 2;5:10833), alteration of genome complexity such as chromosome numbers (e.g., as described in Karimi-Ashtiyani R et

al., Proc Natl Acad Sci U S A. 2015 Sep 8;112(36):11211-6; Anton T, et al., Nucleus. 2014 Mar-Apr;5(2):163-72), self-cleavage of the composition for controlled inactivation/activation (e.g., as described Sugano SS et al., Plant Cell Physiol. 2014 Mar;55(3):475-81), multiplexed gene editing (as described in Kabadi AM et al., Nucleic Acids Res. 2014 Oct 29;42(19):e147), development of kits for multiplex genome editing (as described in Xing HL et al., BMC Plant Biol. 2014 Nov 29;14:327), starch production (as described in Hebelstrup KH et al., Front Plant Sci. 2015 Apr 23;6:247), targeting multiple genes in a family or pathway (e.g., as described in Ma X et al., Mol Plant. 2015 Aug;8(8):1274-84), regulation of non-coding genes and sequences (e.g., as described in Lowder LG, et al., Plant Physiol. 2015 Oct;169(2):971-85), editing genes in trees (e.g., as described in Belhaj K et al., Plant Methods. 2013 Oct 11;9(1):39; Harrison MM, et al., Genes Dev. 2014 Sep 1;28(17):1859-72; Zhou X et al., New Phytol. 2015 Oct;208(2):298-301), introduction of mutations for resistance to host-specific pathogens and pests.

[0748] Additional examples of modifications of plants and fungi that may be performed using the compositions, systems, and methods include analogous modifications described in International Patent Publication Nos. WO2016/099887, WO2016/025131, WO2016/073433, WO2017/066175, WO2017/100158, WO 2017/105991, WO2017/106414, WO2016/100272, WO2016/100571, WO 2016/100568, WO 2016/100562, and WO 2017/019867.

Applications in fungi

[0749] The compositions, systems, and methods disclosed herein can be used to perform efficient and cost effective gene or genome interrogation or editing or manipulation in fungi or fungal cells, such as yeast. The approaches and applications in plants may be applied to fungi as well.

[0750] A fungal cell may be any type of eukaryotic cell within the kingdom of fungi, such as phyla of *Ascomycota*, *Basidiomycota*, *Blastocladiomycota*, *Chytridiomycota*, *Glomeromycota*, *Microsporidia*, and *Neocallimastigomycota*. Examples of fungi or fungal cells include yeasts, molds, and filamentous fungi.

[0751] In one embodiment, the fungal cell is a yeast cell. A yeast cell refers to any fungal cell within the phyla *Ascomycota* and *Basidiomycota*. Examples of yeasts include budding yeast, fission yeast, and mold, *S. cerevisiae*, *Kluyveromyces marxianus*, *Issatchenkia orientalis*, *Candida* spp. (e.g., *Candida albicans*), *Yarrowia* spp. (e.g., *Yarrowia lipolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and

Kluyveromyces marxianus), *Neurospora* spp. (e.g., *Neurospora crassa*), *Fusarium* spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., *Issatchenkia orientalis*, *Pichia kudriavzevii* and *Candida acidothermophilum*).

[0752] In one embodiment, the fungal cell is a filamentous fungal cell, which grow in filaments, e.g., hyphae or mycelia. Examples of filamentous fungal cells include *Aspergillus* spp. (e.g., *Aspergillus niger*), *Trichoderma* spp. (e.g., *Trichoderma reesei*), *Rhizopus* spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

[0753] In one embodiment, the fungal cell is of an industrial strain. Industrial strains include any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains include, without limitation, JAY270 and ATCC4124.

[0754] In one embodiment, the fungal cell is a polyploid cell whose genome is present in more than one copy. Polyploid cells include cells naturally found in a polyploid state, and cells that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may be a cell whose entire genome is polyploid, or a cell that is polyploid in a particular genomic locus of interest. In some examples, the abundance of guide molecule may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the composition disclosed herein may take advantage of using certain fungal cell types.

[0755] In one embodiment, the fungal cell is a diploid cell, whose genome is present in two copies. Diploid cells include cells naturally found in a diploid state, and cells that have been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest.

[0756] In one embodiment, the fungal cell is a haploid cell, whose genome is present in one copy. Haploid cells include cells naturally found in a haploid state, or cells that have been

induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0757] The compositions and systems, and nucleic acid encoding thereof may be introduced to fungi cells using the delivery systems and methods herein. Examples of delivery systems include lithium acetate treatment, bombardment, electroporation, and those described in Kawai et al., 2010, *Bioeng Bugs*. 2010 Nov-Dec; 1(6): 395–403.

[0758] In some examples, a yeast expression vector (e.g., those with one or more regulatory elements) may be used. Examples of such vectors include a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2 μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

Biofuel and materials production by fungi

[0759] In one embodiment, the compositions, systems, and methods may be used for generating modified fungi for biofuel and material productions. For instance, the modified fungi for production of biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. Foreign genes required for biofuel production and synthesis may be introduced into fungi. In some examples, the genes may encode enzymes involved in the conversion of pyruvate to ethanol or another product of interest, degrade cellulose (e.g., cellulase), endogenous metabolic pathways which compete with the biofuel production pathway.

[0760] In some examples, the compositions, systems, and methods may be used for generating and/or selecting yeast strains with improved xylose or cellobiose utilization, isoprenoid biosynthesis, and/or lactic acid production. One or more genes involved in the metabolism and synthesis of these compounds may be modified and/or introduced to yeast cells. Examples of the methods and genes include lactate dehydrogenase, PDC1 and PDC5, and those described in Ha, S.J., et al. (2011) *Proc. Natl. Acad. Sci. USA* 108(2):504-9 and

Galazka, J.M., et al. (2010) Science 330(6000):84-6; Jakočiūnas T et al., Metab Eng. 2015 Mar;28:213-222; Stovicek V, et al., FEMS Yeast Res. 2017 Aug 1;17(5).

Improved plants and yeast cells

[0761] The present disclosure further provides improved plants and fungi. The improved and fungi may comprise one or more genes introduced, and/or one or more genes modified by the compositions, systems, and methods herein. The improved plants and fungi may have increased food or feed production (e.g., higher protein, carbohydrate, nutrient or vitamin levels), oil and biofuel production (e.g., methanol, ethanol), tolerance to pests, herbicides, drought, low or high temperatures, excessive water, etc.

[0762] The plants or fungi may have one or more parts that are improved, e.g., leaves, stems, roots, tubers, Seeds, endosperm, ovule, and pollen. The parts may be viable, nonviable, regeneratable, and/or non-regeneratable.

[0763] The improved plants and fungi may include gametes, Seeds, embryos, either zygotic or somatic, progeny and/or hybrids of improved plants and fungi. The progeny may be a clone of the produced plant or fungi, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly plants.

Exemplary Applications in Non-Human Animals

[0764] The compositions, systems, and methods may be used to study, diagnose, treat, and modify non-human animals, e.g., introducing desirable traits, minimizing or eliminating undesirable traits, and evaluating disease resilience and/or susceptibility, treating and/or preventing diseases, facilitating breeding, etc. In one embodiment, the compositions, systems, and methods may be used to improve breeding and introducing desired traits, e.g., increasing the frequency of trait-associated alleles, introgression of alleles from other breeds/species without linkage drag, and creation of de novo favorable alleles. Genes and other genetic elements that can be targeted may be screened and identified. Examples of application and approaches include those described in Tait-Burkard C, et al., Livestock 2.0 - genome editing for fitter, healthier, and more productive farmed animals. Genome Biol. 2018 Nov 26;19(1):204; Lillico S, Agricultural applications of genome editing in farmed animals. Transgenic Res. 2019 Aug;28(Suppl 2):57-60; Houston RD, et al., Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020 Apr 16. doi: 10.1038/s41576-020-0227-y, which are incorporated herein by reference in their entireties.

Applications described in other sections such as therapeutic, diagnostic, etc. can also be used on the animals herein.

[0765] The compositions, systems, and methods may be used on animals such as fish, amphibians, reptiles, mammals, and birds. The animals may be farm and agriculture animals, or pets. Examples of farm and agriculture animals include horses, goats, sheep, swine, cattle, llamas, alpacas, and birds, e.g., chickens, turkeys, ducks, and geese. The animals may be a non-human primate, e.g., baboons, capuchin monkeys, chimpanzees, lemurs, macaques, marmosets, tamarins, spider monkeys, squirrel monkeys, and vervet monkeys. Examples of pets include dogs, cats, horses, wolfs, rabbits, ferrets, gerbils, hamsters, chinchillas, fancy rats, guinea pigs, canaries, parakeets, and parrots.

[0766] In one embodiment, one or more genes may be introduced (e.g., overexpressed) in the animals to obtain or enhance one or more desired traits. Growth hormones, insulin-like growth factors (IGF-1) may be introduced to increase the growth of the animals, e.g., pigs or salmon (such as described in Pursel VG et al., J Reprod Fertil Suppl. 1990;40:235-45; Waltz E, Nature. 2017;548:148). Fat-1 gene (e.g., from *C elegans*) may be introduced for production of larger ratio of n-3 to n-6 fatty acids may be induced, e.g., in pigs (such as described in Li M, et al., Genetics. 2018;8:1747-54). Phytase (e.g., from *E coli*) xylanase (e.g., from *Aspergillus niger*), beta-glucanase (e.g., from *Bacillus licheniformis*) may be introduced to reduce the environmental impact through phosphorous and nitrogen release reduction, e.g., in pigs (such as described in Golovan SP, et al., Nat Biotechnol. 2001;19:741-5; Zhang X et al., elife. 2018). sgRNA molecule decoy may be introduced to induce avian influenza resilience e.g., in chicken (such as described in Lyall et al., Science. 2011;331:223-6). Lysozyme or lysostaphin may be introduced to induce mastitis resilience e.g., in goat and cow (such as described in Maga EA et al., Foodborne Pathog Dis. 2006;3:384-92; Wall RJ, et al., Nat Biotechnol. 2005;23:445-51). Histone deacetylase such as HDAC6 may be introduced to induce PRRSV resilience, e.g., in pig (such as described in Lu T., et al., PLoS One. 2017;12:e0169317). CD163 may be modified (e.g., inactivated or removed) to introduce PRRSV resilience in pigs (such as described in Prather RS et al., Sci Rep. 2017 Oct 17;7(1):13371). Similar approaches may be used to inhibit or remove viruses and bacteria (e.g., Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema) that may be transmitted from animals to humans.

[0767] In one embodiment, one or more genes may be modified or edited for disease resistance and production traits. Myostatin (e.g., GDF8) may be modified to increase muscle growth, e.g., in cow, sheep, goat, catfish, and pig (such as described in Crispo M et al., PLoS One. 2015;10:e0136690; Wang X, et al., Anim Genet. 2018;49:43–51; Khalil K, et al., Sci Rep. 2017;7:7301; Kang J-D, et al., RSC Adv. 2017;7:12541–9). Pc POLLED may be modified to induce horlessness, e.g., in cow (such as described in Carlson DF et al., Nat Biotechnol. 2016;34:479–81). KISS1R may be modified to induce boretaint (hormone release during sexual maturity leading to undesired meat taste), e.g., in pigs. Dead end protein (dnd) may be modified to induce sterility, e.g., in salmon (such as described in Wargelius A, et al., Sci Rep. 2016;6:21284). Nano2 and DDX may be modified to induce sterility (e.g., in surrogate hosts), e.g., in pigs and chicken (such as described Park K-E, et al., Sci Rep. 2017;7:40176; Taylor L et al., Development. 2017;144:928–34). CD163 may be modified to induce PRRSV resistance, e.g., in pigs (such as described in Whitworth KM, et al., Nat Biotechnol. 2015;34:20–2). RELA may be modified to induce ASFV resilience, e.g., in pigs (such as described in Lillico SG, et al., Sci Rep. 2016;6:21645). CD18 may be modified to induce Mannheimia (Pasteurella) haemolytica resilience, e.g., in cows (such as described in Shanthalingam S, et al., roc Natl Acad Sci U S A. 2016;113:13186–90). NRAMP1 may be modified to induce tuberculosis resilience, e.g., in cows (such as described in Gao Y et al., Genome Biol. 2017;18:13). Endogenous retrovirus genes may be modified or removed for xenotransplantation such as described in Yang L, et al. Science. 2015;350:1101–4; Niu D et al., Science. 2017;357:1303–7). Negative regulators of muscle mass (e.g., Myostatin) may be modified (e.g., inactivated) to increase muscle mass, e.g., in dogs (as described in Zou Q et al., J Mol Cell Biol. 2015 Dec;7(6):580-3).

[0768] Animals such as pigs with severe combined immunodeficiency (SCID) may generated (e.g., by modifying RAG2) to provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development. Examples of methods and approaches include those described Lee K, et al., Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5; and Schomberg et al. FASEB Journal, April 2016; 30(1):Suppl 571.1.

[0769] SNPs in the animals may be modified. Examples of methods and approaches include those described Tan W. et al., Proc Natl Acad Sci U S A. 2013 Oct 8;110(41):16526-31; Mali P, et al., Science. 2013 Feb 15;339(6121):823-6.

[0770] Stem cells (e.g., induced pluripotent stem cells) may be modified and differentiated into desired progeny cells, e.g., as described in Heo YT et al., Stem Cells Dev. 2015 Feb 1;24(3):393-402.

[0771] Profile analysis (such as Igenity) may be performed on animals to screen and identify genetic variations related to economic traits. The genetic variations may be modified to introduce or improve the traits, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain.

[0772] It will be appreciated that the systems and compositions of the present invention can be used on animals of agriculture importance as well as companion animals, particularly for the treatment and diagnosis of disease, disease susceptibility and/or risk, prevention, coat color, performance (e.g., athletic or reproductive), and lineage/parentage, and breed identity. Exemplary profile analysis for companion animals includes Embark Pet DNA tests, Wisdom Panel pet DNA tests, Basepaws DNA test, Optimal Selection DNA test, and/or the like. These screen for a variety of health and disease risk markers, breed identification markers, and lineage markers. The systems and compositions disclosed herein can be used in similar type profile analysis.

[0773] Compositions and methods for detection of markers relevant to these and other profile analysis of non-human animals are described in greater detail elsewhere herein.

[0774] 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 – General Structure of Exemplary Cas12 Proteins

[0775] The crystal structure of Type V Cas protein, C2c1 (also known as Cas 12b), reveals similarity with Type V Cas protein, Cpf1 (also known as Cas12a). Both C2c1 and Cpf1 consist of an α -helical recognition lobe (REC) and a nuclease lobe (NUC). The NUC lobe further contains a oligonucleotide-binding (WED/OBD) domain, a RuvC domain, a Nuc domain, and a bridge helix (BH), with structural shuffling and folding to form the intact 3D C2c1 structure (Liu et al. Mol. Cell 65, 310-322). Certain mutations (e.g., R1226A in AsCpf1) in the Nuc domain render Cpf1 into a nickase for non-target strand cleavage. Mutations of the catalytic residues (e.g., mutations at D908, E933, D1263 of AsCpf1) in the RuvC domain abolishes

catalytic activity of Cpf1 as a nuclease. Further, mutations in the PAM interaction (PI) domain of Cpf1 (e.g., mutations at S542, K548, N522, and K607 of AsCpf1), have been shown to alter Cpf1 specificities, potentially increasing or reducing off-target cleavage (See Gao et al. Cell Research (2016) 26, 901–913 (2016); Gao et al. Nature Biotechnology 35, 789–792 (2017)). The crystal structure of C2c1 also reveals that C2c1 lacks an identifiable PI domain; rather, it is suggested that C2c1 undergoes conformation adjustment to accommodate the binding of the PAM proximal double stranded DNA for PAM recognition and R-loop formation; C2c1 likely engages the WED/OBD and alpha helix domain to recognize the PAM duplex from both the major and the minor groove sides (Yang et al, Cell 167, 1814-1828 (2016)).

AsCas12a

[0776] Structural characteristics of AsCpf1 include those described in Yamano et al., Crystal structure of Cpf1 in complex with guide RNA and target DNA, Cell 165, 949-962, May 5, 2016.

Overall structure

[0777] AsCpf1 adopts a bilobed architecture consisting of an α -helical recognition (REC) lobe and a nuclease (NUC) lobe, with the crRNA-target DNA heteroduplex bound to the positively charged, central channel between the two lobes. The REC lobe consists of the REC1 and REC2 domains, whereas the NUC lobe consists of the RuvC domain and three additional domains, denoted A, B, and C. The REC1 domain comprises 13 α helices, and the REC2 domain comprises ten α helices and two β strands that form a small antiparallel sheet. Domains A and B play functional roles similar to those of the WED(Wedge) and PI (PAM-interacting) domains of Cas9 (Anders et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016). Domain C is involved in DNA cleavage. Thus, domains A, B, and C are referred to as theWED,PI, and Nuc domains, respectively. The WED domain is assembled from three separate regions (WED-I-III) in the Cpf1 sequence. The WED domain can be divided into a core subdomain comprising a nine-stranded, distorted antiparallel β sheet (β 1- β 8 and β 11) flanked by seven α helices (α 1- α 6 and α 9) and a subdomain comprising two β strands (β 9 and β 10) and two α helices (α 7 and α 8). The PI domain comprises seven α helices (α 1- α 7) and a β hairpin (β 1 and β 2) and is inserted between the WED-II and WED-III regions, whereas the REC lobe is inserted between the WED-I and WED-II regions. The RuvC domain contains the three motifs (RuvC-I-III) that form the endonuclease active center. A characteristic helix (referred to as the bridge helix) is

located between the RuvC-I and RuvC-II motifs and connects the REC and NUC lobes. The Nuc domain is inserted between the RuvC-II and RuvC-III motifs.

crRNA and target DNA recognition

[0778] The 5' handle of the crRNA is bound at the groove between the WED and RuvC domains. The U(-1),U(-16) base pair in the 5' handle is recognized by the WED domain in a base-specific manner. U(-1) and U(-16) hydrogen bond with His761 and Arg18/Asn759, respectively, while U(-1) stacks on His761. The N6 of A(-19) hydrogen bonds with Leu807 and Asn808, while the base moieties of A(-18) and A(-19) form stacking interactions with Ile858 and Met806, respectively. Moreover, the phosphodiester backbone of the 5' handle forms an extensive network of interactions with the WED and RuvC domains.

Recognition of the crRNA-target DNA heteroduplex

[0779] The crRNA-target DNA heteroduplex is accommodated within the positively charged, central channel formed by the REC1, REC2, and RuvC domains and is recognized by the protein in a sequence-independent manner. The PAM-distal and PAM-proximal regions of the heteroduplex are recognized by the REC1-REC2 domains and the WED-REC1-RuvC domains, respectively. Arg951 and Arg955 in the bridge helix interact with the sugar-phosphate backbone of the target DNA strand. Notably, the sugar-phosphate backbone of the nucleotides G1–A8 in the crRNA forms multiple contacts with the WED and REC1 domains, and the base pairing within the 5-bp PAM-proximal “seed” region is important for Cpf1-mediated DNA cleavage. In addition, the backbone phosphate group between dT(-1) and dC1 of the target DNA strand (referred to as the +1 phosphate) is recognized by the side chain of Lys780 and the main-chain amide group of Gly783. This interaction results in the rotation of the +1 phosphate group, thereby facilitating base pairing between dC1 in the target DNA strand and G1 in the crRNA. The residues involved in the heteroduplex recognition are conserved in most members of the Cpf1 family (Zetsche et al., 2015), and the R176A, R192A, G783P, and R951A mutants exhibited reduced activities, confirming their functional relevance.

Cpf1 recognizes the 20-bp RNA-DNA heteroduplex. The side chain of Trp382 in the REC2 domain forms a stacking interaction with the C20:dG20 base pair in the heteroduplex and thus prevents base pairing between A21 and dT21. Indeed, the W382A mutant showed reduced activity, highlighting its functional importance.

PAM recognition

[0780] The PAM duplex adopts a distorted conformation with a narrow minor groove and is bound to the groove formed by the WED, REC1, and PI domains. The PAM duplex is recognized by the WED-REC1 and PI domains from the major and minor groove sides, respectively. The dT(-1):dA(-1*) base pair in the PAM duplex does not form base-specific contacts with the protein, consistent with the lack of specificity in the fourth position of the 5'-TTTN-3' PAM. Lys607 in the PI domain is inserted into the narrow minor groove and plays critical roles in the PAM recognition. The O2 of dT(-2*) forms a hydrogen bond with the side chain of Lys607, whereas the nucleobase and deoxyribose moieties of dA(-2) form van der Waals interactions with the side chains of Lys607 and Pro599/Met604, respectively. The 5-methyl group of dT(-3*) forms a van der Waals interaction with the side-chain methyl group of Thr167, whereas the N3 and N7 of dA(-3) form hydrogen bonds with Lys607 and Lys548, respectively. The 5-methyl group of dT(-4*) is surrounded by the side-chain methyl groups of Thr167 and Thr539, whereas the O4' of dA(-4) forms a hydrogen bond with the side chain of Lys607. Notably, the N3 and O4 of dT(-4*) form hydrogen bonds with the N1 of dA(-4) and the N6 of dA(-3), respectively. The K548A and M604A mutants exhibited reduced activities, confirming that Lys548 and Met604 participate in the PAM recognition. Further, the K607A mutant showed almost no activity, indicating that Lys607 is critical for the PAM recognition.

RuvC domain

[0781] The RuvC domain comprises a typical RNase H fold, consisting of a five-stranded mixed β sheet ($\beta 1$ – $\beta 5$) flanked by three α helices ($\alpha 1$ – $\alpha 3$), and two additional helices and three β strands. The conserved, negatively charged residues Asp908, Glu993, and Asp1263 form an active site. The D908A and E993A mutants had almost no activity, whereas the D1263A mutant exhibited significantly reduced activity, confirming the roles of Asp908, Glu993, and Asp1263 in DNA cleavage. Notably, the bridge helix is inserted between strand $\beta 3$ and helix $\alpha 1$ in the RNase H fold and interacts with the REC2 domain. The main-chain carbonyl group of Gln956 in the bridge helix forms a hydrogen bond with the side chain of Lys468 in the REC2 domain. In addition, Trp958 in the RuvC domain is accommodated in the hydrophobic pocket formed by Leu467, Leu471, Tyr514, Arg518, Ala521, and Thr522 in the REC2 domain. These observations highlight the functional importance of the bridge helix-mediated interaction between the REC and NUC lobes.

Nuc domain

[0782] The crystal structure revealed the presence of the Nuc domain, which is inserted between the RuvC-II (strand $\beta 5$) and RuvC-III (helix $\alpha 3$) motifs in the RuvC domain. The Nuc domain is connected to the RuvC domain via two linker loops (referred to as L1 and L2). The Nuc domain comprises five α helices and nine β strands and lacks detectable structural or sequence similarity to any known nucleases or proteins. Notably, the conserved polar residues Arg1226 and Asp1235 and the partially conserved Ser1228 are clustered in the proximity of the active site of the RuvC domain. The S1228A mutant showed DNA cleavage activity comparable to that of wild-type AsCpf1. In contrast, the D1235A mutant exhibited reduced activity, whereas the R1226A mutant showed almost no activity, indicating that Arg1226 is critical for DNA cleavage. Further characterization revealed that the R1226A mutant acts as a nickase that cleaves the non-target DNA strand, but not the target strand, indicating that the Nuc and RuvC domains cleave the target and non-target DNA strands, respectively. The mutations of the catalytic residues in the AsCpf1 RuvC domain abolished the cleavage of both DNA strands, suggesting that the cleavage of the non-target strand by the RuvC domain is a prerequisite for the target strand cleavage by the Nuc domain.

FnCpf1

[0783] Structural characteristics of FnCpf1 include those described in (Swarts et al., Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a, Molecular Cell 66, 221-233, April 20, 2017).

Overall structure

[0784] FnCpf1 adopts a bilobed architecture with the two lobes connected by the wedge (WED) domain. The N-terminal REC lobe consists of two α -helical domains (REC1 and REC2) that have been shown to coordinate the crRNA-target DNA heteroduplex. The C-terminal NUC lobe consists of the C-terminal RuvC and Nuc domains involved in target cleavage, the arginine-rich bridge helix (BH), and the PAM-interacting (PI) domain. The repeat-derived segment of the crRNA forms a pseudoknot stabilized by intra-molecular base-pairing and hydrogen-bonding interactions. The pseudoknot is coordinated by residues from the WED, RuvC, and REC2 domains, as well as by two hydrated magnesium cations. Notably, nucleotides 1–5 of the crRNA are ordered in the central cavity of FnCas12a and adopt an A-form-like helical conformation. Conformational ordering of the seed sequence is facilitated by multiple interactions between the ribose and phosphate moieties of the crRNA backbone and

Fncas12a residues in the WED and REC1 domains. These include residues Thr16, Lys595, His804, and His881 from the WED domain and residues Tyr47, Lys51, Phe182, and Arg186 from the REC1 domain. The structure of the Fncas12a-crRNA complex further reveals that the bases of the seed sequence are solvent exposed and poised for hybridization with target DNA.

Pre-crRNA processing

[0785] Essential residues for crRNA processing include His843, Lys852, and Lys869. Structural observations are consistent with an acid-base catalytic mechanism in which Lys869 acts as the general base catalyst to deprotonate the attacking 2'-hydroxyl group of U(-19), while His843 acts as a general acid to protonate the 5'-oxygen leaving group of A(-18). In turn, the side chain of Lys852 is involved in charge stabilization of the transition state. Collectively, these interactions facilitate the intra-molecular attack of the 2'-hydroxyl group of U(-19) on the scissile phosphate and promote the formation of the 2',3'-cyclic phosphate product.

R-loop formation

[0786] The crRNA-target DNA strand heteroduplex is enclosed in the central cavity formed by the REC and NUC lobes and interacts extensively with the REC1 and REC2 domains. The PAM-containing DNA duplex comprises target strand nucleotides dT0–dT8 and non-target strand nucleotides dA(8)*–dA0* and is contacted by the PI, WED, and REC1 domains. The 5'-TTN-3' PAM is recognized in Fncas12a by a mechanism combining the shape-specific recognition of a narrowed minor groove, with base-specific recognition of the PAM bases by two invariant residues, Lys671 and Lys613. Directly downstream of the PAM, the duplex of the target DNA is disrupted by the side chain of residue Lys667, which is inserted between the DNA strands and forms a cation- π stacking interaction with the dA0–dT0* base pair. The phosphate group linking target strand residues dT(-1) and dT0 is coordinated by hydrogen-bonding interactions with the side chain of Lys823 and the backbone amide of Gly826. Target strand residue dT(-1) bends away from residue T0, allowing the target strand to interact with the seed sequence of the crRNA. The non-target strand nucleotides dT1*–dT5* interact with the Arg692-Ser702 loop in Fncas12a through hydrogen-bonding and ionic interactions between backbone phosphate groups and side chains of Arg692, Asn700, Ser702, and Gln704, as well as main-chain amide groups of Lys699, Asn700, and Ser702. Alanine substitution of Q704 or replacement of residues Thr698–Ser702 in Fncas12a with the sequence Ala-Gly3 substantially reduced DNA cleavage activity, suggesting that these residues

contribute to R-loop formation by stabilizing the displaced conformation of the nontarget DNA strand.

[0787] In the FnCas12a R-loop complex, the crRNA-target strand heteroduplex is terminated by a stacking interaction with a conserved aromatic residue (Tyr410). This prevents base pairing between the crRNA and the target strand beyond nucleotides U20 and dA(-20), respectively. Beyond this point, the target DNA strand nucleotides re-engage the non-target DNA strand, forming a PAM-distal DNA duplex comprising nucleotides dC(-21)–dA(-27) and dG21*–dT27*, respectively. The duplex is confined between the REC2 and Nuc domains at the end of the central channel formed by the REC and NUClobes.

Target DNA cleavage

[0788] FnCpf1 can independently accommodate both the target and non-target DNA strands in the catalytic pocket of the RuvC domain. The RuvC active site contains three catalytic residues (D917, E1006, and D1255). Structural observations suggest that both the target and non-target DNA strands are cleaved by the same catalytic mechanism in a single active site in Cas12a enzymes.

Example 2 – Example Methods for Identifying Reduced Immunogenicity Epitopes

Experimental Methods

[0789] Briefly, Cyrus/Rosetta (“C/R”), also known as Rosetta/Bench, software, was used to identify several epitope locations and sequences for modification. C/R implements a hybrid of a physics-based energy function and a trained statistical scoring function to model protein structure. This method avoids the need for explicit solvent models thereby reducing computational cost and increasing computational efficiency. The hybrid method is integrated into a modified Monte Carlo (MC) sampling process resulting in an exhaustive conformational landscape search for increased confidence in the results.

[0790] For example, a polypeptide of interest is first conformationally stabilized via the modified (MC) sampling method. While (MC) sampling methods of polypeptides are well known in the art, C/R is capable of rapidly searching rotamer and backbone space by further limiting angular and torsional rotations. These limitations are calibrated from angular and torsional restraints from databases of known polypeptides (e.g., from the RCSB Protein Data Bank (RCSB PDB)). Once the structure of the polypeptide of interest has been thoroughly

sampled, the amino acids of designated side chains can be modified for further testing. C/R will then re-sample the modified polypeptide and the stability of the proposed modification can be ranked. See, e.g., **FIG. 3**.

[0791] During the MC process, conformationally acceptable polypeptides will be further measured by the hybrid physical/statistical model. This hybrid model calculates the potential energy of the MC-derived polypeptide system using a molecular mechanics force field such as CHARMM and a potential energy term from the statistical scoring function. The potential energy from the statistical scoring function is derived, for example, from Bayesian and/or Potential of Mean Force models. These models are trained on known intra/inter-polypeptide interactions such as bond lengths, angles and dihedrals between atoms, and energy distributions. The force-field potential energy is combined with the statistical potential energy to form a relative energy, the structure with the lowest relative energy is the most likely structure of the polypeptide. See, e.g., **FIG. 4**.

[0792] Polypeptides designed with C/R may be further refined by taking into account host genome and epitope-like sequences. In one such example, the sequence of the polypeptide of interest are searched against the host genome. Polymers (such as 9-mers) in the sequence of the polypeptide of interest are aligned with the host genome and return a “reward” to the scoring function if the polymer and host genome match. In addition, the polymers corresponding to the sequence of the polypeptide of interest are searched against known epitope sequences. Polymers that match sequences in a known epitope sequence return a “penalty” to the scoring function. The polymers that are penalized are removed or modified to avoid immune response. The modified sequences are processed by a machine learning algorithm, which predicts whether the modified sequence will be recognized by the immune system. See, e.g., **FIGS. 5-7** and King, C.; *et al.* Removing T-Cell Epitopes with Computational Protein Design. Proceedings of the National Academy of Sciences, 2014, 111, 8577–8582.

[0793] The machine learning algorithm can be modeled many ways. For example, HLA Class I peptides can be categorized into supertypes based on their binding profile (See, e.g., **FIG. 8**). Accordingly, the machine learning algorithm can be a single model trained against all supertypes or the machine learning algorithm can be part of a system or network comprising individual models, wherein each model represents one of the supertypes. The advantage of the latter is a single model may reward an input sequence because it is predicted to not correspond to most supertypes. However, the input sequence may trigger an immune response for the

remaining subtypes. Therefore, multiple models may produce a more refined prediction across all supertypes and, consequently, produce a deimmunized polypeptide for more people. See, e.g., **FIGS. 9-11** and King *et al.* 2014, which illustrate a design process for a deimmunized polypeptide comprising:

- 1) Rosetta MHC PSSM design at selected positions with manual curation; and
- 2) Evaluate MHC binding across a panel of 14 human alleles using neural network-based NetMHCpan for external validation:
 - a. NetMHCpan
 - i. Calculate 4 metrics per design,
 - ii. Weak binder threshold 5%,
 - iii. Strong binder threshold 0.5%.
 - b. Single HLA Rank Score: *Higher is better.*
 - i. *What is the predicted binding strength of this peptide to the HLA molecule in the experimental cell line?*
 - c. #Binders: *Lower is better.*
 - i. *For this epitope, how many HLA molecules from the population panel are predicted to bind?*
 - d. #Strong Binders: *Lower is better.*
 - i. *For this epitope, how many HLA molecules from the population panel are predicted to bind very strongly?*
 - e. #Global Binders: *Lower is better.*
 - i. *For the whole protein, how many HLA molecules from the population panel are predicted to bind?*

Cas12 Epitopes Selected for Modification

[0794] **FIG. 1** illustrates the 3D structure of AsCas12a (AsCpf1) based on C/R and the several AsCas12a epitope locations and sequences identified by C/R for modification. Several of the tested AsCas12a single mutants (single epitope with a single amino acid substitution) displayed high single HLA rank scores (as described herein). **FIG. 12** shows the AsCas12a P210 epitope sequence (RLITAVPSL (SEQ ID NO: 57)), supertype (A2), RNA contact (none), and structure. **FIG. 13** shows the AsCas12a P210 epitope A2 rank scoring, #Binders, #Strong Binders, #Global Binders, and epitope structure for native AsCas12a P210 epitope vs. single mutants. **FIG. 14** shows the AsCas12a P277 epitope sequence (LNEVLNLAI (SEQ ID NO:

58)), supertype (C17), RNA contact (Yes, small), and structure. **FIG. 15** shows the AsCas12a P277 epitope C17 rank scoring, #Binders, #Strong Binders, #Global Binders and epitope structure of native AsCas12a P277 epitope vs. single mutants. **FIG. 16** shows the AsCas12a P972 epitope sequence (YLSQVIHEI (SEQ ID NO: 59)), supertype (A2), RNA contact (yes), and structure. **FIG. 17** shows the AsCas12a P972 A2 rank scoring, #Binders, #Strong Binders, #Global Binders and epitope structure for native AsCas12a P972 epitope vs. single mutants.

Modification and Evaluation of Identified Cas12 Epitopes

[0795] **FIG. 2A** illustrates the AsCas12a single mutant nuclease efficiency of various “single epitope mutants” - each epitope comprises a single amino acid substitution. **FIG. 2B** illustrates AsCas12a double mutant nuclease efficiency (two epitopes, each epitope comprises a single amino substitution).

Modification and Evaluation of Predicted asCas12a Epitopes

[0796] As described above, the Cyrus/Rosetta software was used to identify asCas12a epitope locations and sequences that are predicted to confer reduced immunogenicity.

[0797] **FIG. 18** shows the asCas12a p2 predicted epitope sequence (MTQFEGFTNLYQ (SEQ ID NO: 109)), predicted p2 mutants that remove P9 pocket residues (L->Q,N mutations), supertype (B39), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0798] **FIG. 19** shows the asCas12a p11 predicted epitope sequence (NLYQVSKTLRFELI (SEQ ID NO: 114)), predicted p11 mutants that target P9 pocket residues (F->N,D mutations), supertype (B62), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0799] **FIG. 20** shows the asCas12a p160 predicted epitope sequence (NALLRSFDKFTTYF (SEQ ID NO: 118)), predicted p160 mutants that target P9 pocket residues (Y->Q,G,M mutations), supertype (B62), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0800] **FIG. 21** shows the asCas12a p241 predicted epitope sequence (VFSFPFYNNQLLTQT (SEQ ID NO: 122)), predicted p241 mutants that remove P1, P3, and P9 residues, supertype (C17), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0801] **FIG. 22** shows the asCas12a p505 predicted epitope sequence (TGIKLEMEPSLSFYI (SEQ ID NO: 128)), predicted p505 mutants that remove P2 and P9

pocket residues (E->N,Q, F->N mutations), supertype (B44), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0802] FIG. 23 shows the asCas12a p557 predicted epitope sequence (LFVKNGLYYLGIMI (SEQ ID NO: 131)), predicted p557 mutants that remove P2 and P9 residues (V->Q, L->N,T mutations), supertype (B62), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0803] FIG. 24 shows the asCas12a p869 predicted epitope sequence (SDKFFFHVPITLN (SEQ ID NO: 136)), predicted p869 mutants that removes P2 and P9 pocket residues (F->A,M, L->G,A mutations), supertype (B39), structure and scoring (#Binders, #Strong Binders, #Global Binders).

[0804] FIG. 25 summarizes the scoring of the asCas12a predicted epitope mutants of FIGs. 18-24 based on number of Natural vs Mutant Binders; number of Natural strong vs Mutant strong binders and number of Natural vs Mutant Global Binders.

[0805] Table 6 below depicts the amino acid sequences of the predicted wild type and mutant asCas12a epitopes disclosed in **FIGs. 18-24**.

TABLE 6

FIG	SEQUENCE	SEQ ID NO
18	MTQFEGFTNLYQ	109
	MLVYEGFTNNYQ	110
	MLVYEGFTNQYQ	111
	MLFYEGFTNQYQ	112
	MTYYEGFTNNYQ	113
19	NLYQVSKTLRFELI	114
	NLMQQSKTLRNELI	115
	NLFQVSKTLRNELI	116
	NLMQQSKTLRDELI	117
20	NALLRSFDKFTTYF	118
	NALLQSNDKFTTQF	119
	NALLQSNDKFTTGF	120
	NALLQSNDKFTTMF	121
21	VFSFPFYNQLLTQT	122
	VNANPFYNQQLTQT	123
	VNSAPFYNQNLTTQT	124
	VNSAPFYNQQLTQT	125
	VNDLPFYNQLLTQT	126
	VQAAPFYNQQLTQT	127
22	TGIKLEMEPSLSFYI	128
	TGIKQNNEPSLSNYI	129
	TGIKQKMEPSLSNYI	130
23	LFVKNGLYYLGIMI	131
	LFQKNGLYYNGIMI	132
	LFQKNGLYYTGIMI	133
	LGQKNGLYYNGIMI	134
	LQQKNGLYYNGIMI	135
24	SDKFFFHVPITLN	136
	SDKAMNHVPITGN	137
	SDKFFNHVPITGN	138
	SDKFMNHVPITGN	139
	SDKFDNHVPITGN	140
	SDKFMNHVPITAN	141
	SDKFANHVPITGN	142
	SDKFANHVPITAN	143

[0806] Modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. An engineered Type V Cas comprising one or more amino acid substitutions in one or more epitopes corresponding to one or more MHC Class I and/or MHC Class II binding sites in a wild-type Type V Cas, wherein the engineered Type V Cas has reduced immunogenicity compared to a wild-type Type V Cas.
2. The engineered Type V Cas of claim 1, wherein the MHC Class I and/or MHC Class II binding sites are selected from HLA Class I binding sites, HLA Class II binding sites, or any combination thereof.
3. The engineered Type V Cas of any of the preceding claims, wherein the Type V Cas is a Cas12a, a Cas12b, a Cas12c, a Cas12d, a Cas 12e, a Cas12f (Cas14), Cas12k or a variant thereof.
4. The engineered Type V Cas of claim 3, wherein the Type V Cas is a Cas12a, and wherein the Cas12a is a *Lachnospiraceae bacterium* Cas12a (LbCas12a), an *Acidaminococcus* sp. (AsCas12a), a *Francisella novicida* Cas12a (FnCas12a), a *Moraxella bovoculi* Cas12a (MbCas12), an *Eubacterium rectale* Cas12a (MAD7), or a variant thereof.
5. The engineered Type V Cas of claim 4, wherein the Cas12 is an *Acidaminococcus* sp. (AsCas12a), or a variant thereof.
6. The engineered Type V Cas of any one of the preceding claims, wherein the one or more epitopes corresponding to one or more MHC Class I and/or MHC Class II binding sites are selected from AsCas12a residues 210-218, 227-285, and/or 971-979, or residues in another Type V Cas analogous thereto.
7. The engineered Type V Cas of claim 6, wherein the one or more substitutions are at AsCas12a residues 211, 218, 285, 972, and/or 979, or at residues in another Type V Cas analogous thereto.

8. The engineered Type V Cas of claim 6 or 7, wherein the one or more amino acid substitutions comprise substitution of one or more original residues selected from I, L, Y, and any combination thereof, with one or more naturally occurring residues selected from A, D, E, F, G, I, K, M, N, P, Q, R, S, T, V, W, Y, and any combination thereof.
9. The engineered Type V Cas of any one of claims 6-8, wherein the one or more amino acid substitutions comprise: L211V, L211A, L218A, L218D, L218E, L218F, L218G, L218I, L218K, L218M, L218N, L218P, L218Q, L218R, L218S, L218T, L218V, L218W, L218Y, L277D, L277F, L277K, L277M, L277N, L277P, L277Q, L277R, L277W, L281G, L281M, L281N, L281P, L281Q, L281T, L281W, I285A, I285S, I285T, Y971D, Y971M, Y971N, Y971P, Y971Q, Y971W, L972A, L972D, L972E, L972F, L972G, L972I, L972K, L972M, L972N, L972P, L972Q, L972R, or any combination thereof.
10. The engineered Type V Cas of any one of claims 6-9, wherein the one or more amino acid substitutions comprise two or more amino acid substitutions selected from:
- L218S and I285A,
 - L218S and I285S,
 - L218S and I285T,
 - L218S and L972A,
 - L218S and L972F,
 - L218T and I285A,
 - L218T and I285S,
 - L218T and L972A, and
 - L218T and L972F.
11. The engineered Type V Cas of any one of the preceding claims comprising one or more additional modifications that increase nuclease efficiency, RNA binding efficiency, or reduce off-target nuclease activity.
12. The engineered Type V Cas of anyone of the preceding claims, wherein the Cas is a nickase or catalytically inactive (dCas).

13. The engineered Type V Cas of anyone of the preceding claims, wherein the Cas is further linked to or otherwise capable of associating with a heterologous functional domain.
14. The engineered Type V Cas of anyone of the preceding claims, wherein functional domain is a nucleotide deaminase, a transposase, a reverse transcriptase, a recombinase, a methylase, a demethylase, an acetylase, or a deacetylase.
15. A nucleic acid molecule comprising a nucleotide sequence that encodes the engineered Type V Cas of any one of claims 1-14.
16. A composition comprising (i) the engineered Type V Cas of any one of claims 1-14 and (ii) at least one guide molecule capable of forming a complex with the Type V Cas and directing binding of the complex to a target sequence on a target polynucleotide.
17. The composition of claim 16, further comprising a donor template.
18. A nucleic acid molecule comprising a nucleotide sequence that encodes the Type V Cas and the at least one guide molecule of claim 16.
19. A vector comprising the nucleic acid molecule of claim 15 or 18.
20. The vector of claim 19, wherein the vector is a viral vector.
21. A delivery particle comprising the vector of claim 19 or 20 .
22. A delivery particle comprising the composition of claim 16 or 17.
23. A cell comprising the engineered Type V Cas polypeptide of any one of claims 1-11, the nucleic acid molecule of claim 15 or 18, the vector of claim 19 or 20, the composition of claim 16 or 14, the delivery particle of claim 18 or 19, or any combination thereof.

24. A method of modifying target polynucleotides comprising administering the composition of claim 16 or 17, the vector of claims 19 or 20, or the delivery particle of claim 21 or 22.

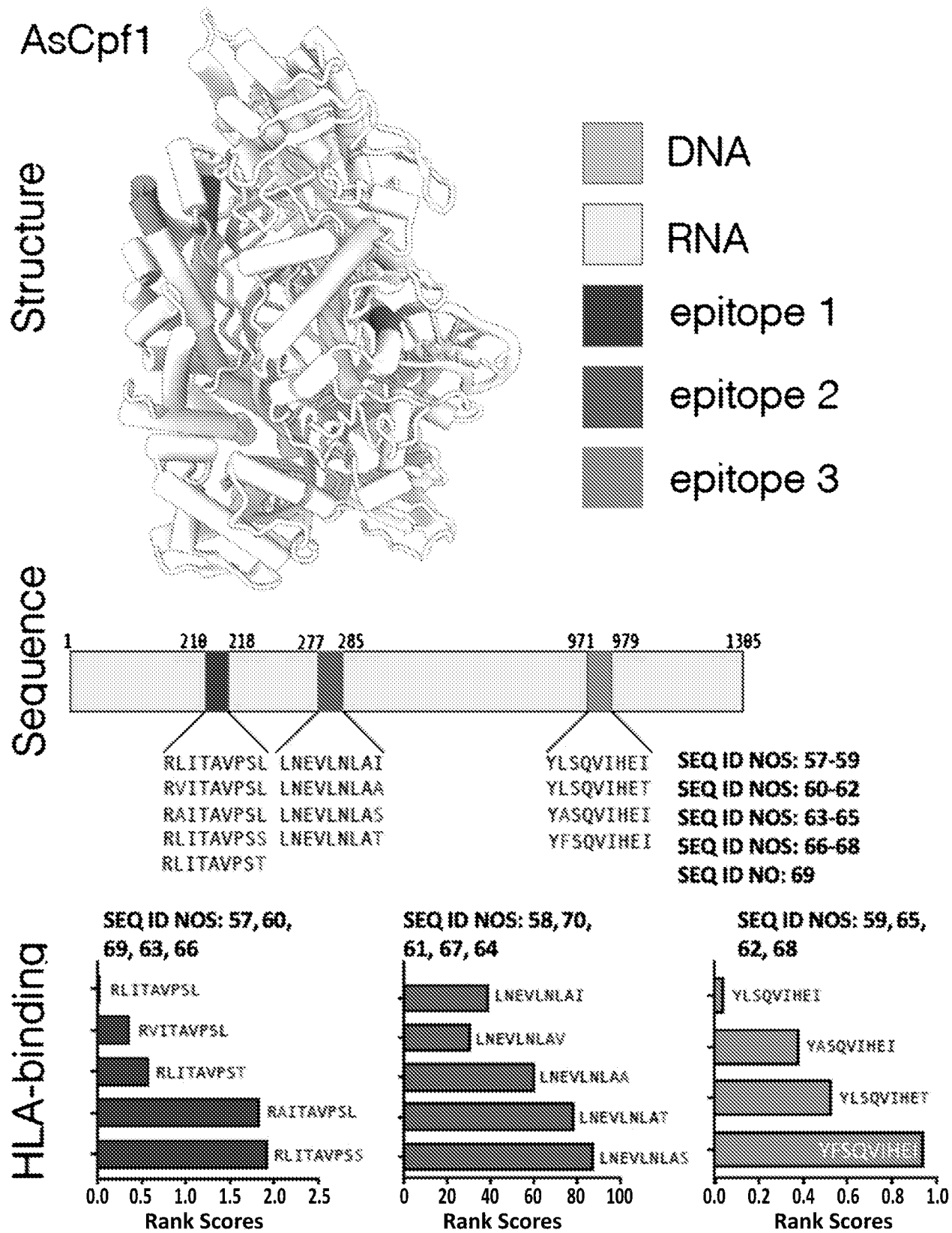


FIG. 1

2/25

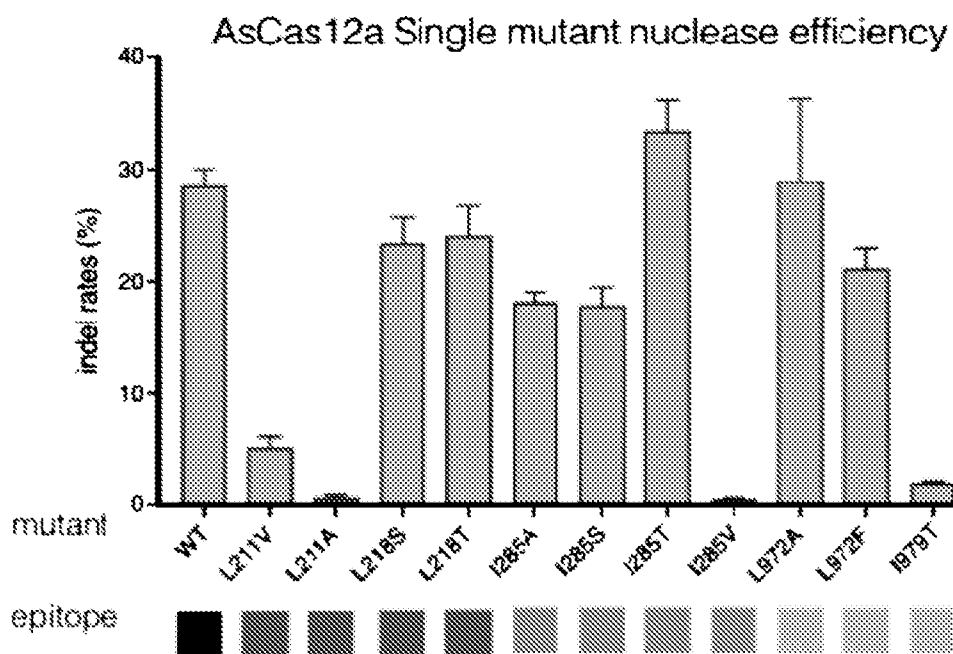


FIG. 2A

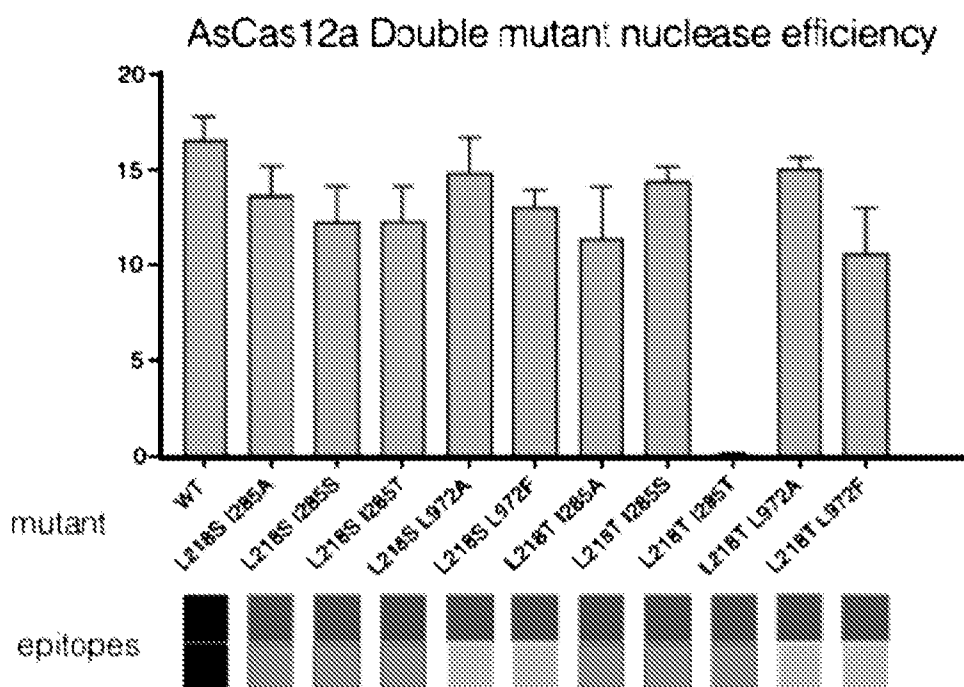


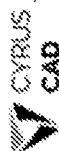
FIG. 2B

3/25

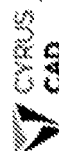
Sampling: Efficiency by Monte Carlo Rotamer+Backbone Search

- Sidechain Rotamer and Identity sampled by Monte Carlo ("MC")

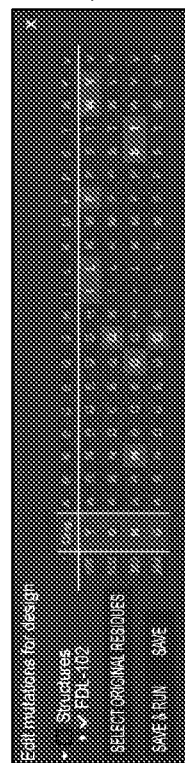
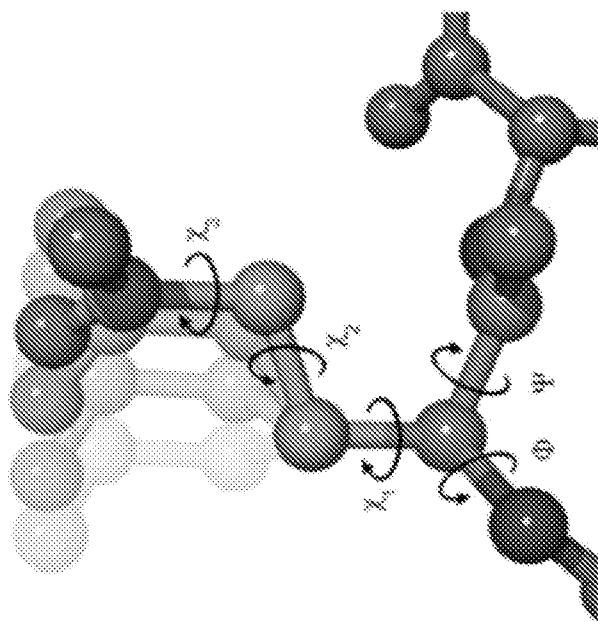
- Rapid Search of large range of sidechain space
- Other methods exist, but MC is highly efficient for realistically-sized proteins



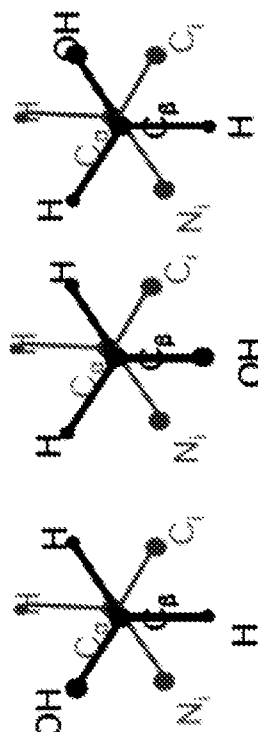
- "**Repack**" = MC sampling of rotamers in a given amino acid



- "**Design**" = MC sampling of both rotamers and new amino acids, as directed by user in the Design tab. All residues in the structure **Repack** in a Design - only Selector residues can mutate.



SEQ ID NOS: 71-75



- Proven since Top7⁽¹⁾, hundreds of published papers in Rosetta use this method

FIG. 3

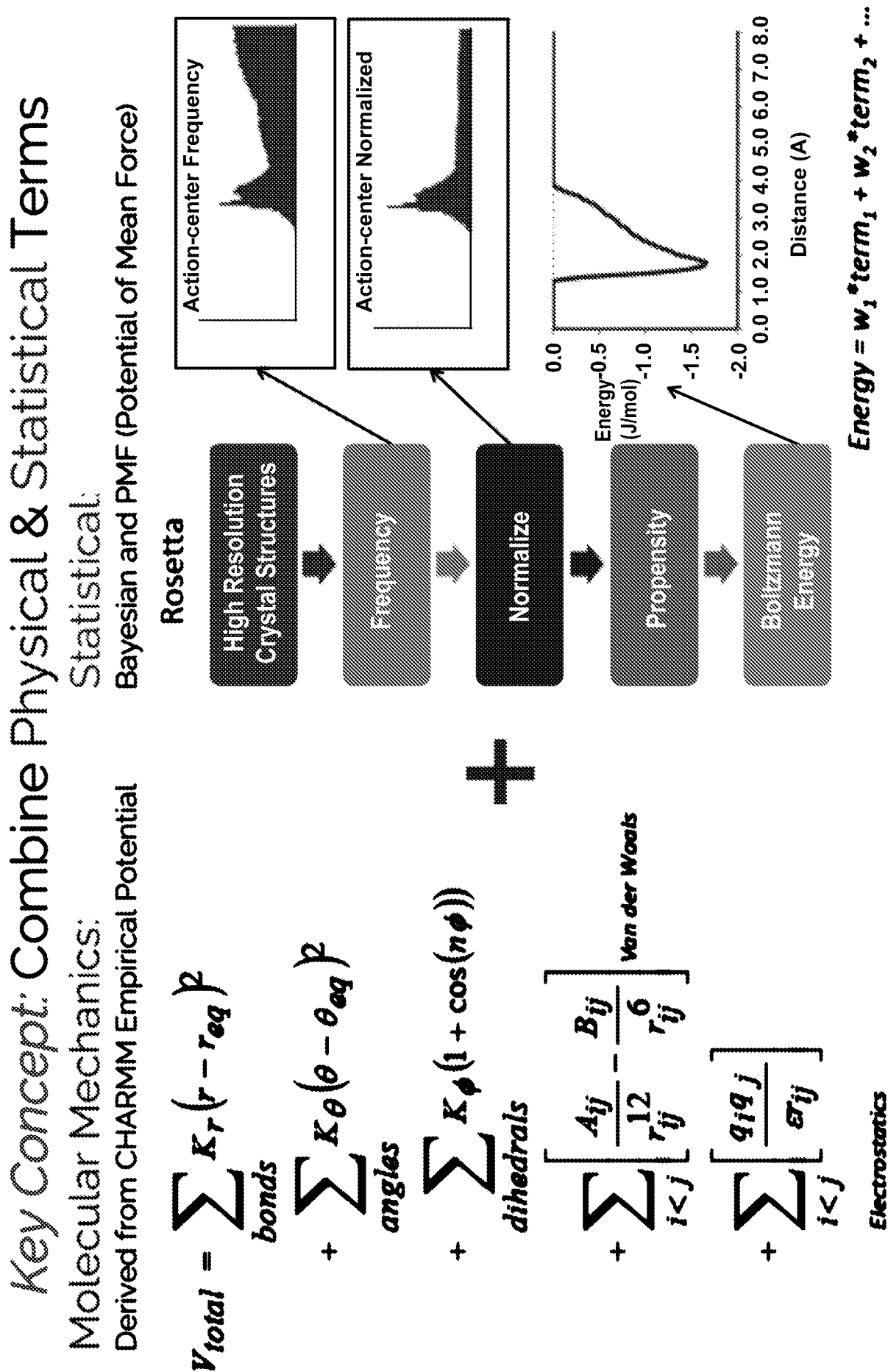


FIG. 4

Three additions to the Rosetta energy function penalize epitope-like sequences.

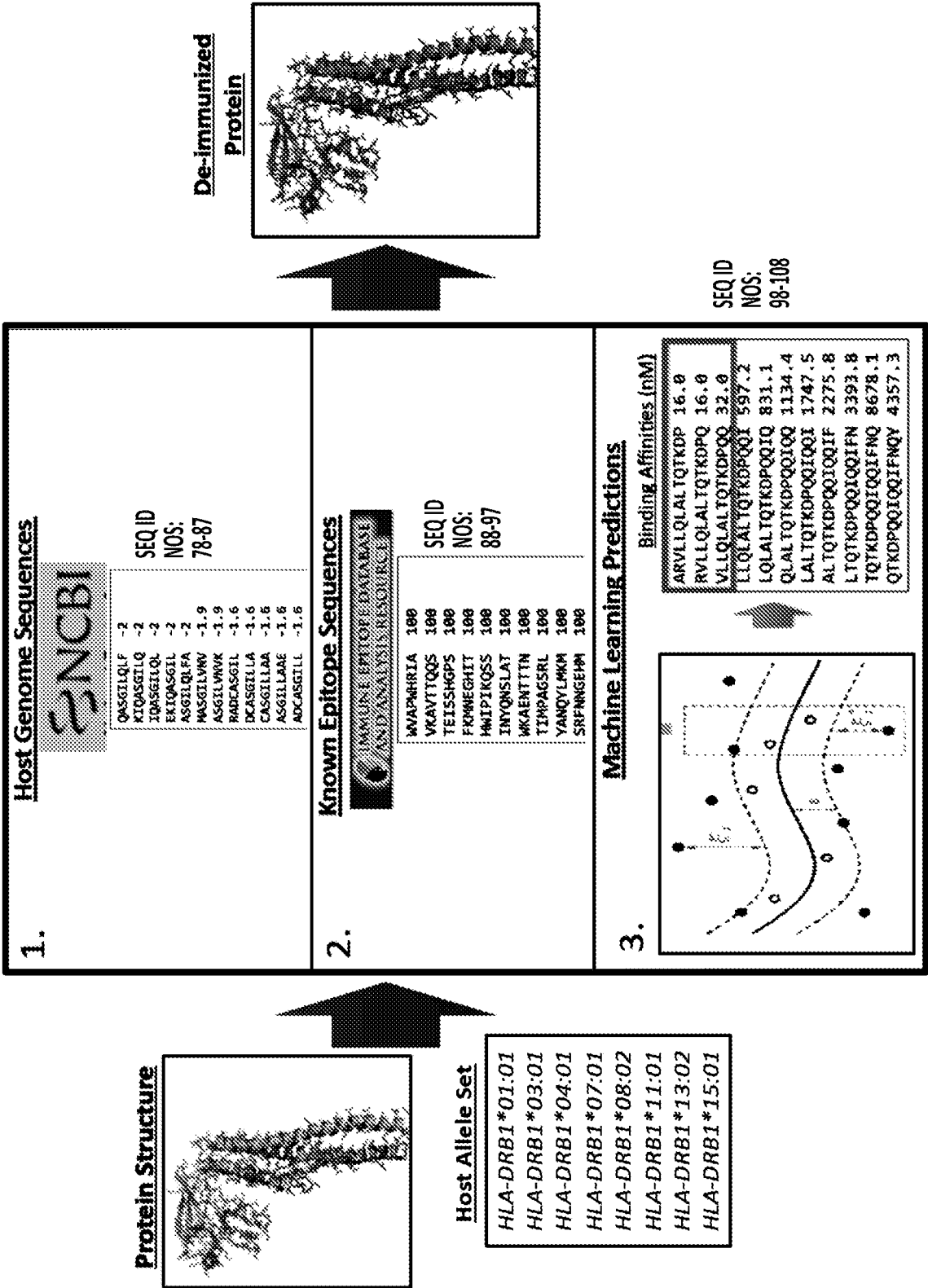


FIG. 5

1: Increase host genome sequence content by rewarding 9mers from host genome

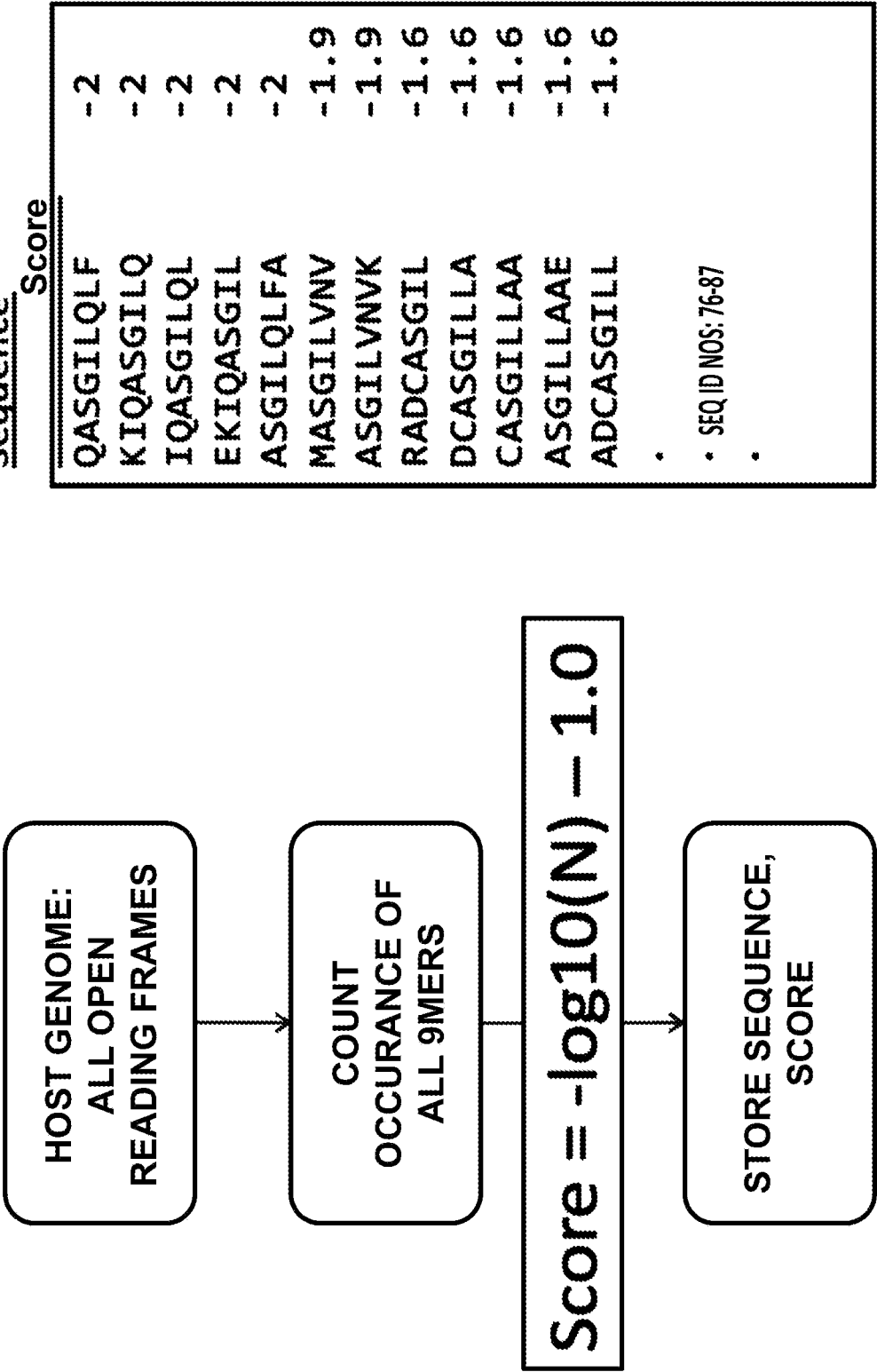


FIG. 6

7/25

Cyrus Matrices

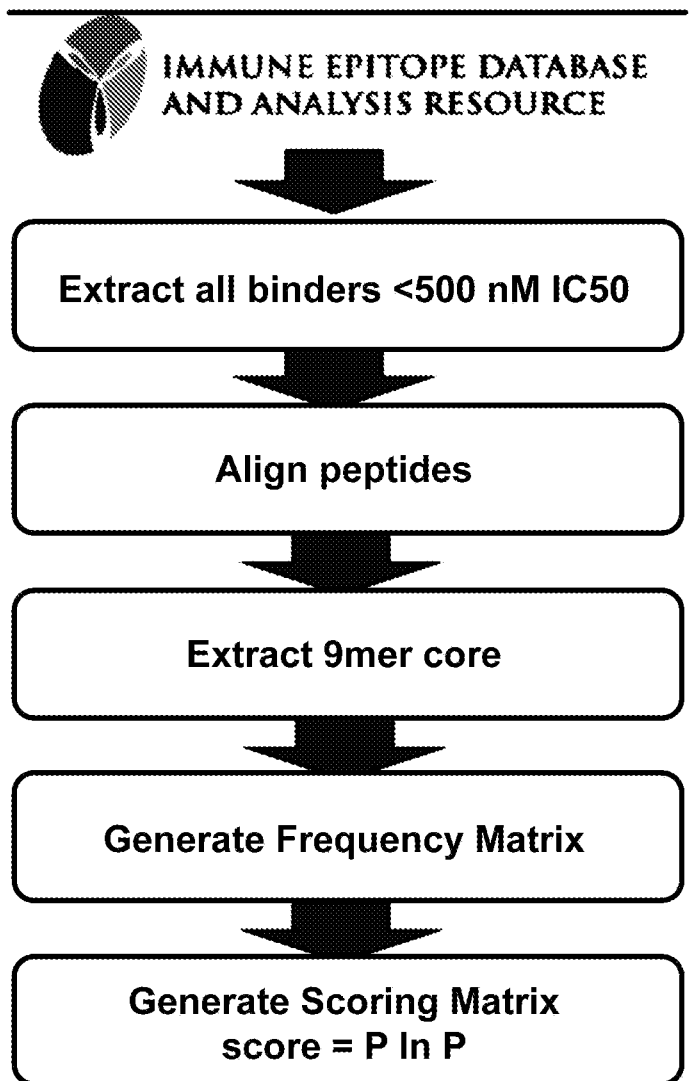


FIG. 7

HLA Class I peptide binding diversity can be clustered into a discrete set of supertypes based on peptide binding profiles

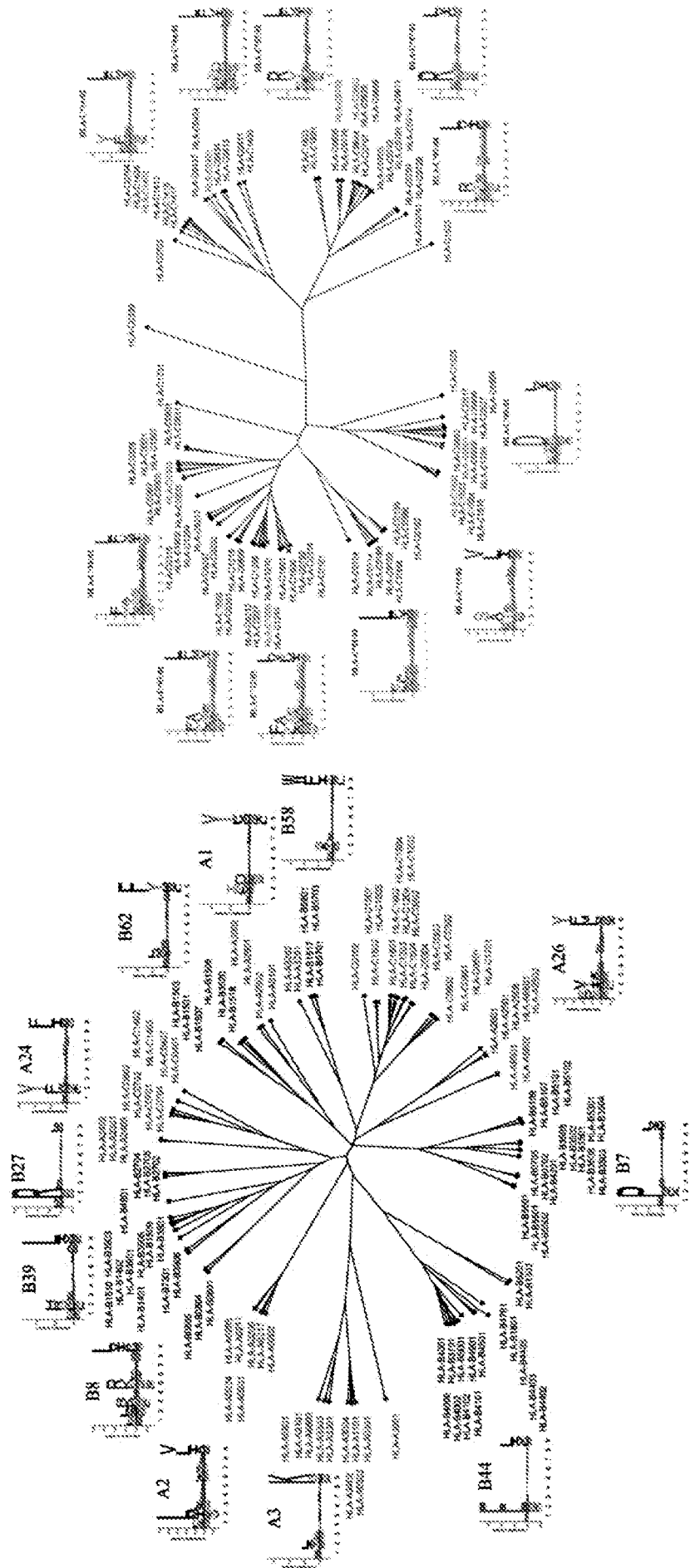


FIG. 8

9/25

Each supertype cluster is represented by a single allele ML model

Supertype	Representative Model
A3	A3101
A2	A0201
B8	B0801
B39	B3901
B27	B2705
A24	A2402
B62	B1501
A1	A0101
B58	B5801
A26	A2601
B7	B0702
B44	B4101
C17	C1701
C2	C0202

FIG. 9

10/25

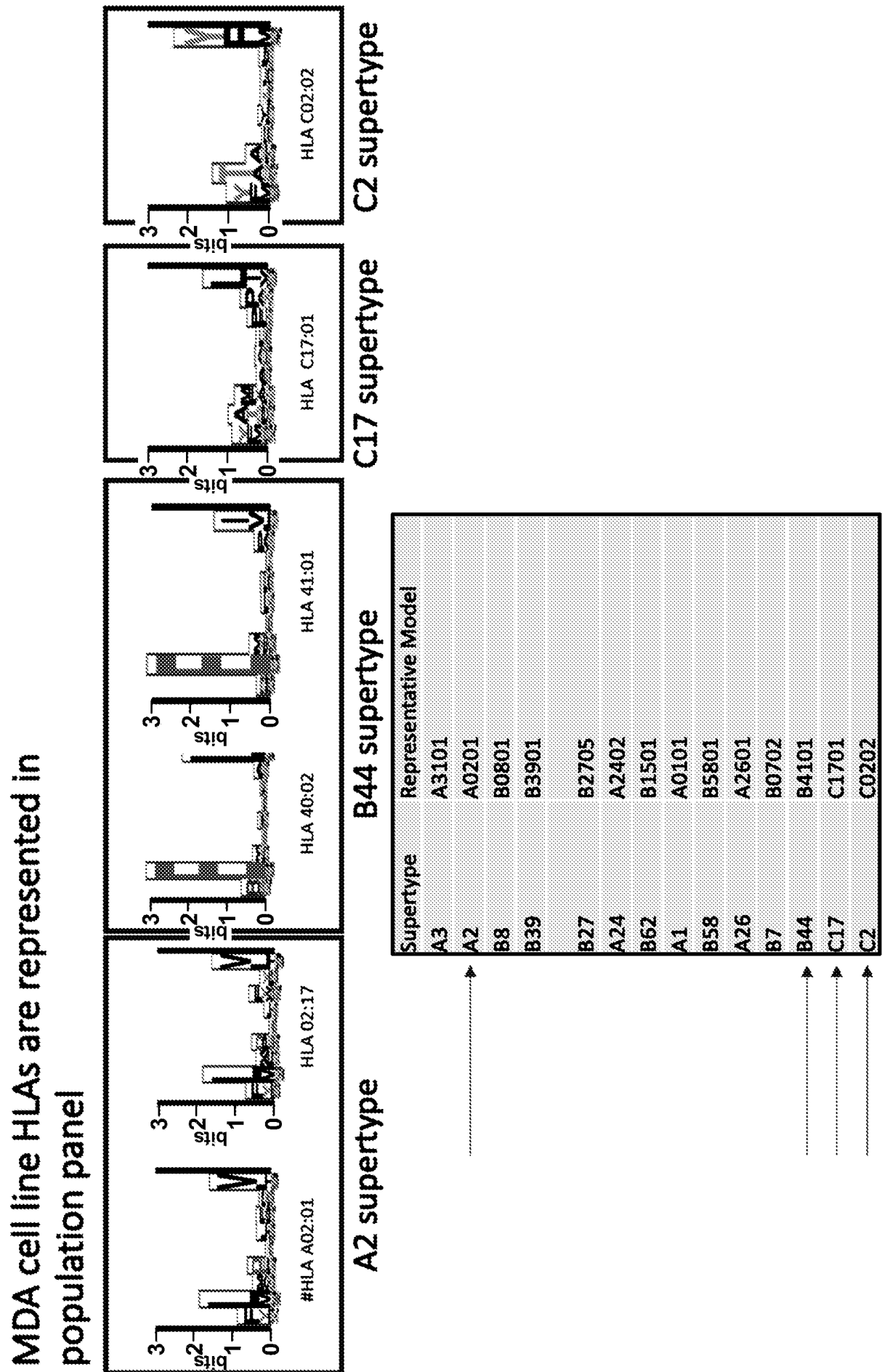


FIG. 10

11/25

Six HLA Class I alleles represented by cell line belong to 4 supertypes

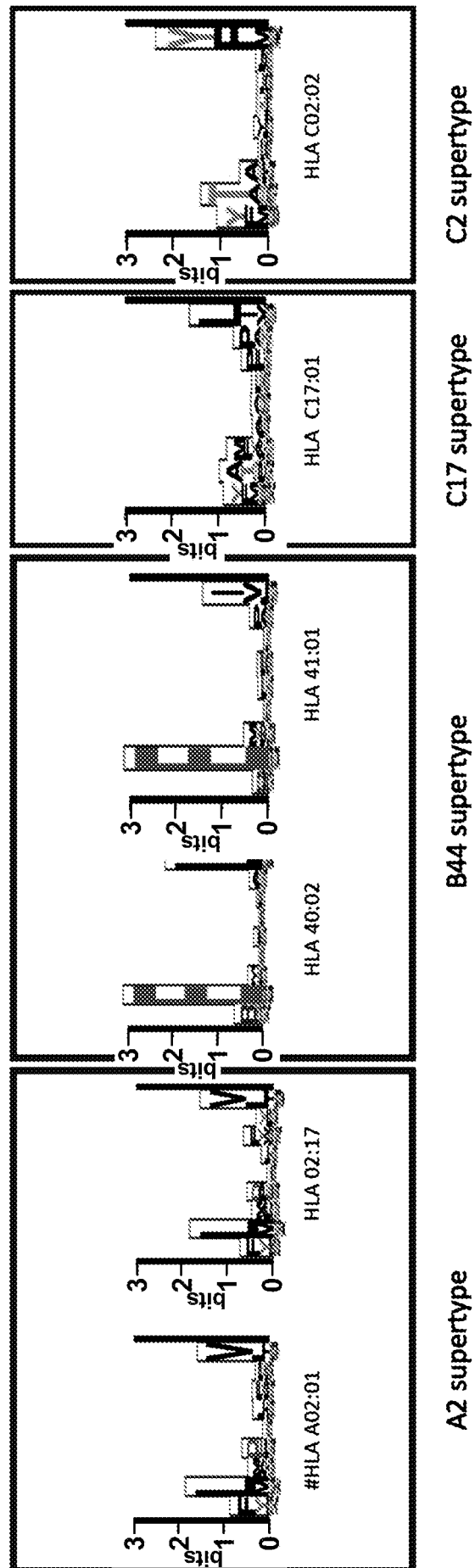


FIG. 11

12/25

P210 epitope RLITAVPSL SEQ ID NO:57

SEQ ID NO:57

- Sequence:
 - RLITAVPSL
- Supertype:
 - A2
- RNA contact?
 - No

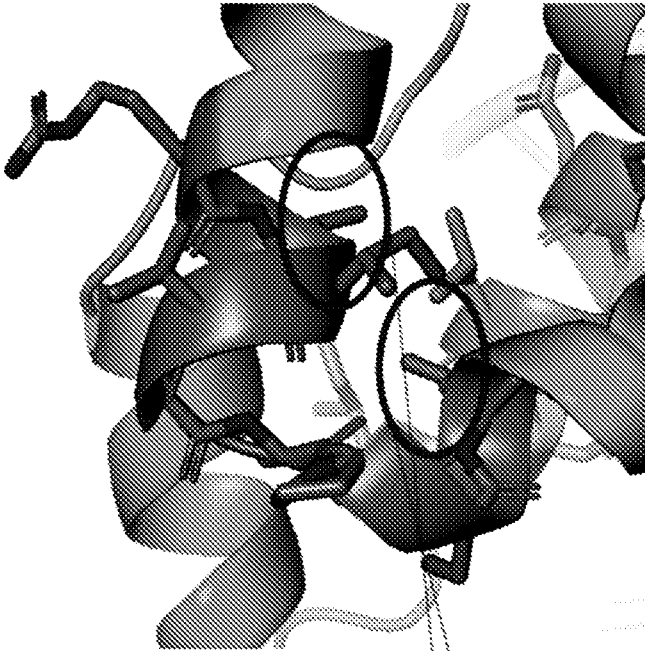
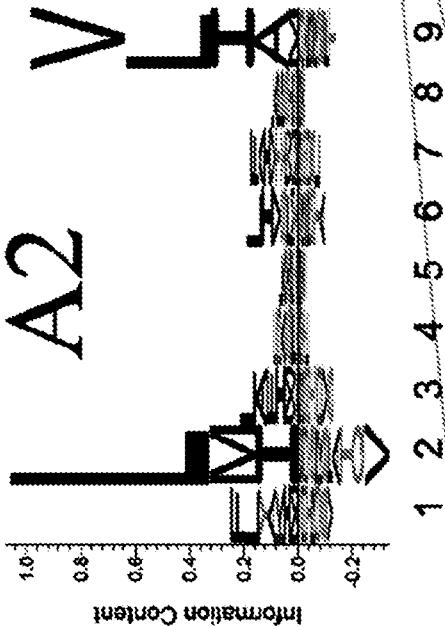
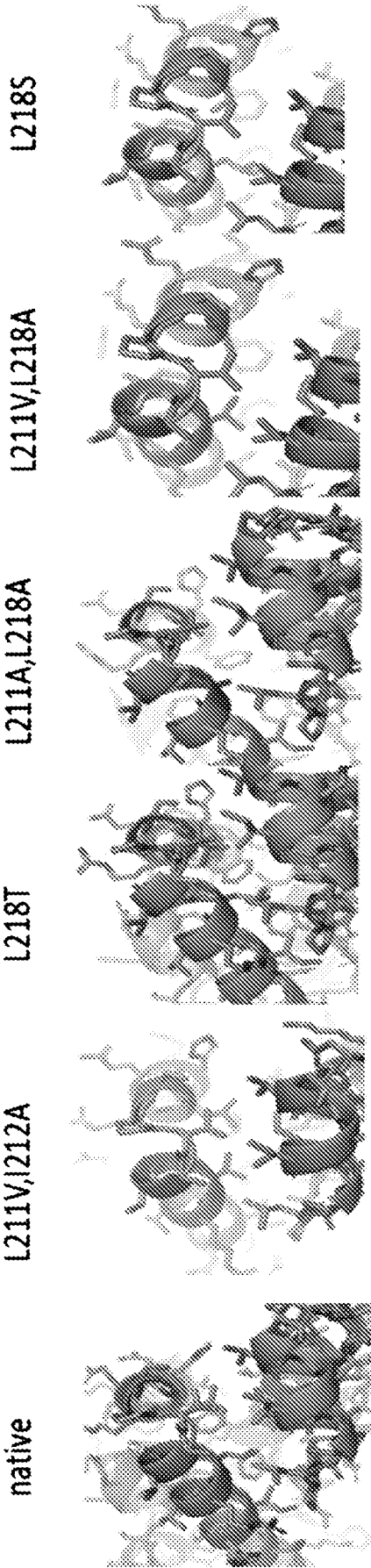


FIG. 12

P210 epitope RLITAVPSL SEQ ID NO: 57



A2 Rank Score	0.04	0.695	0.452	3.74	1.07	1.68
#Binders	13	11	4	7	9	2
#Strong Binders	4	3	1	0	0	0
#Global Binders	2231	2230	2224	2223	2227	2217

FIG. 13

P277 epitope LNEVLNLAI SEQ ID NO: 58

- Sequence:
 - LNEVLNLAI
SEQ ID NO: 58
- Supertype:
 - C17? (ambiguous)
- RNA contact?
 - Yes

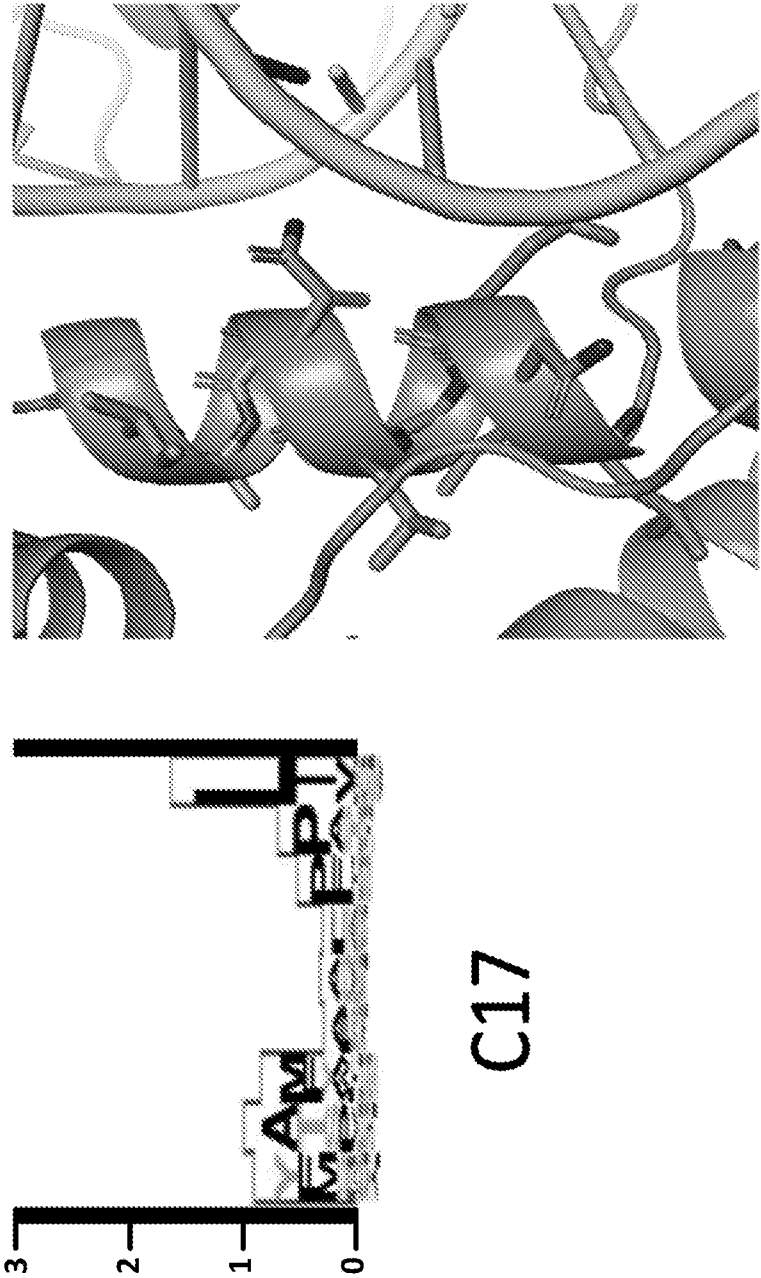


FIG. 14

P277 epitope LNEVLNLA| SEQ ID NO: 58
Epitope is RNA binding site, must be very careful.

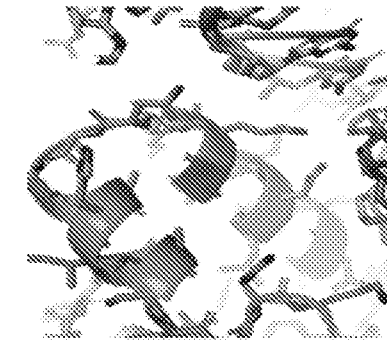
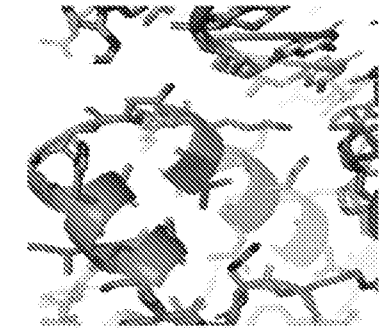
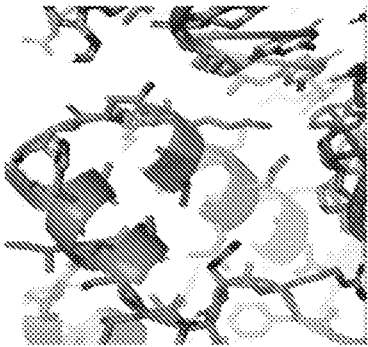
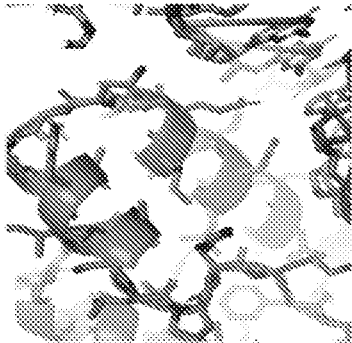
native

I285V

I285T

I285A

I285S



15/25

C17 Rank Score	13.32	14.17	53.91	32.60	67.5
#Binders	0	0	0	0	0
#Strong Binders	0	0	0	0	0
#Total	2231	2232	2230	2231	2230

FIG. 15

P972 epitope YLSQVIHEI SEQ ID NO:59

SEQ ID NO:59

- Sequence:
 - YLSQVIHEI
- Supertype:
 - A2
- RNA contact?
 - Yes

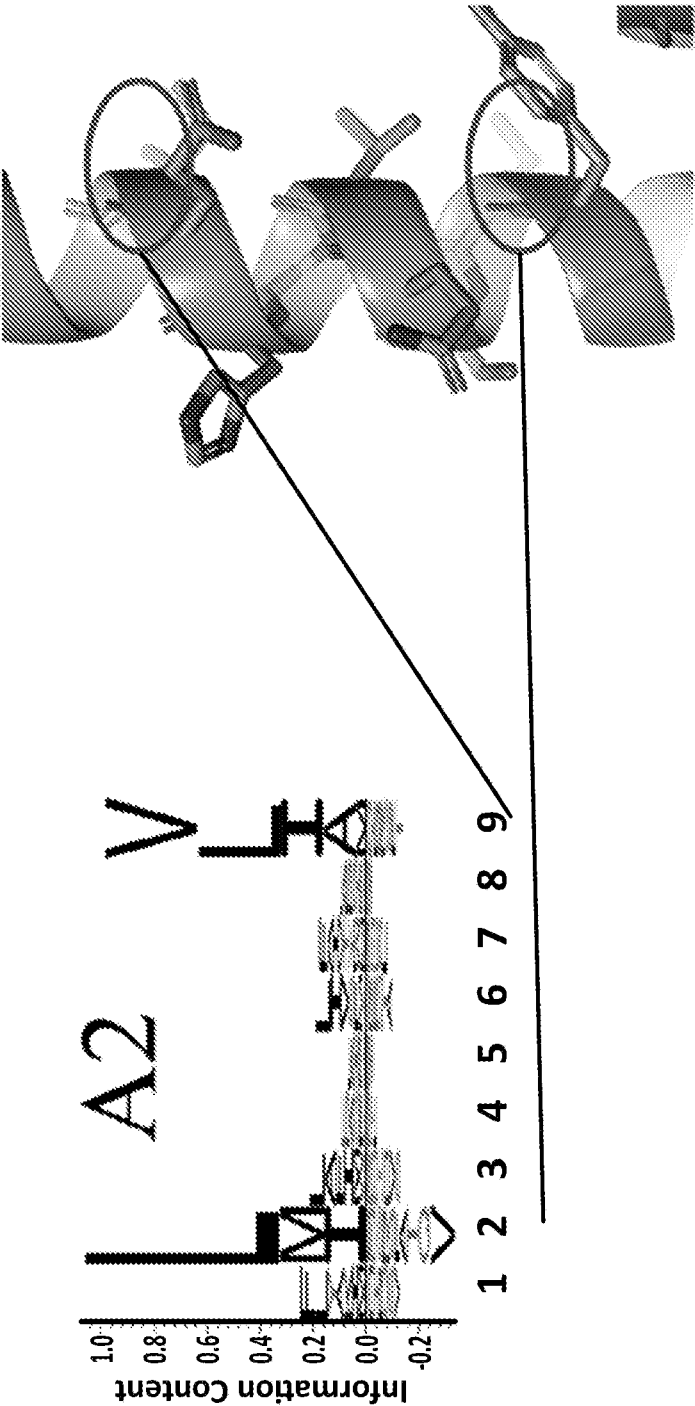


FIG. 16

P972 epitope YLSQVIHEI SEQ ID NO: 59

Epitope is RNA binding site, must be very careful.

	native	L972F	I979A	I979T	L972A,I979V	I979A,I907F	I979T,I907F
							
A2 Rank Score	0.01	1.25	0.02	0.09	0.36	0.09	0.09
#Binders	10	9	9	3	9	9	3
#Strong Binders	3	3	3	1	3	3	1
#Total	2231	2233	2229	2221	2228	2223	2225

FIG. 17

p2 mutants remove P9 pocket residues
(L->Q,N)

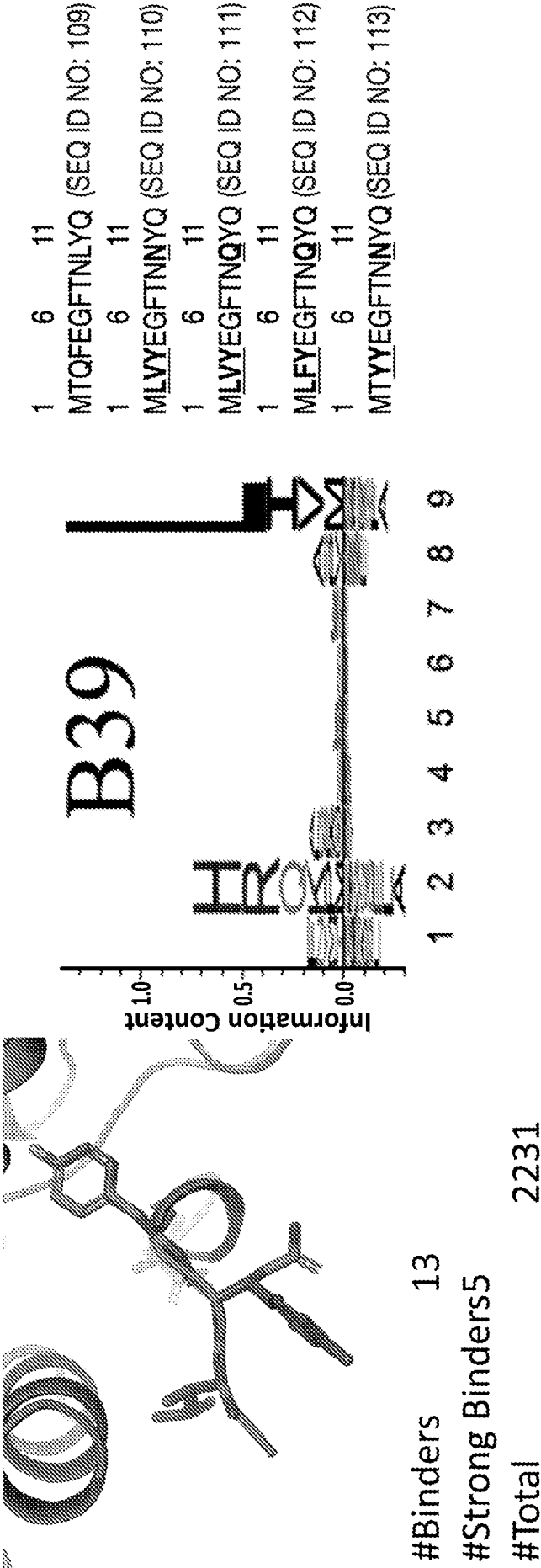


FIG. 18

p11 mutations target P9 with F->N,D

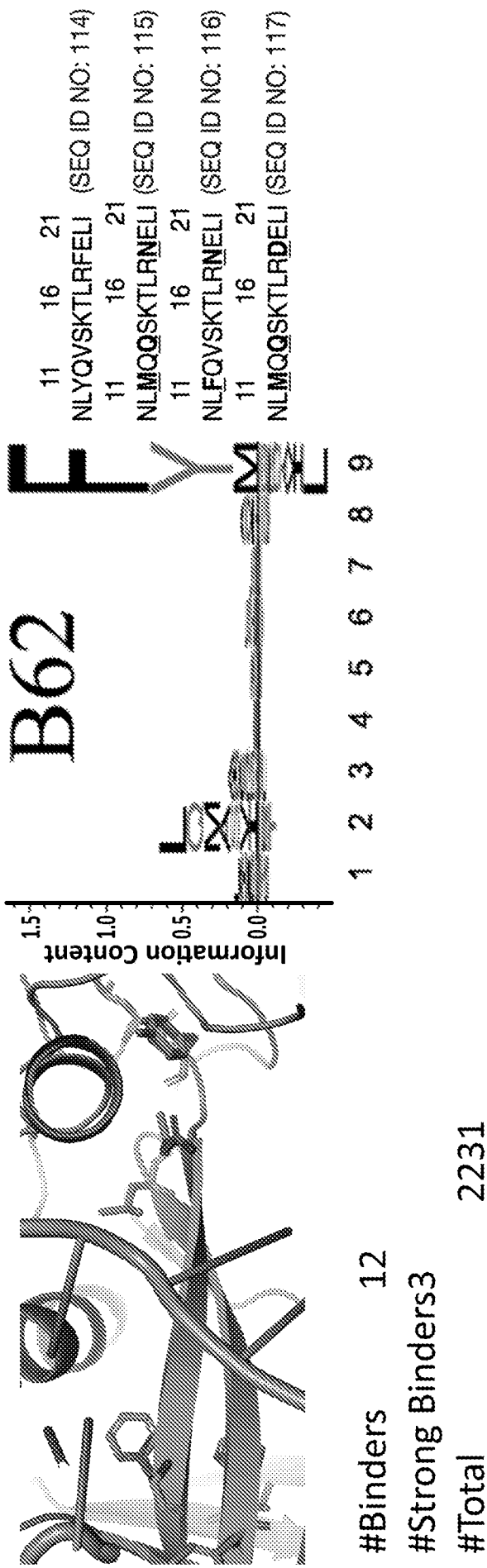


FIG. 19

p160 mutations target P9 with Y->Q,G,M

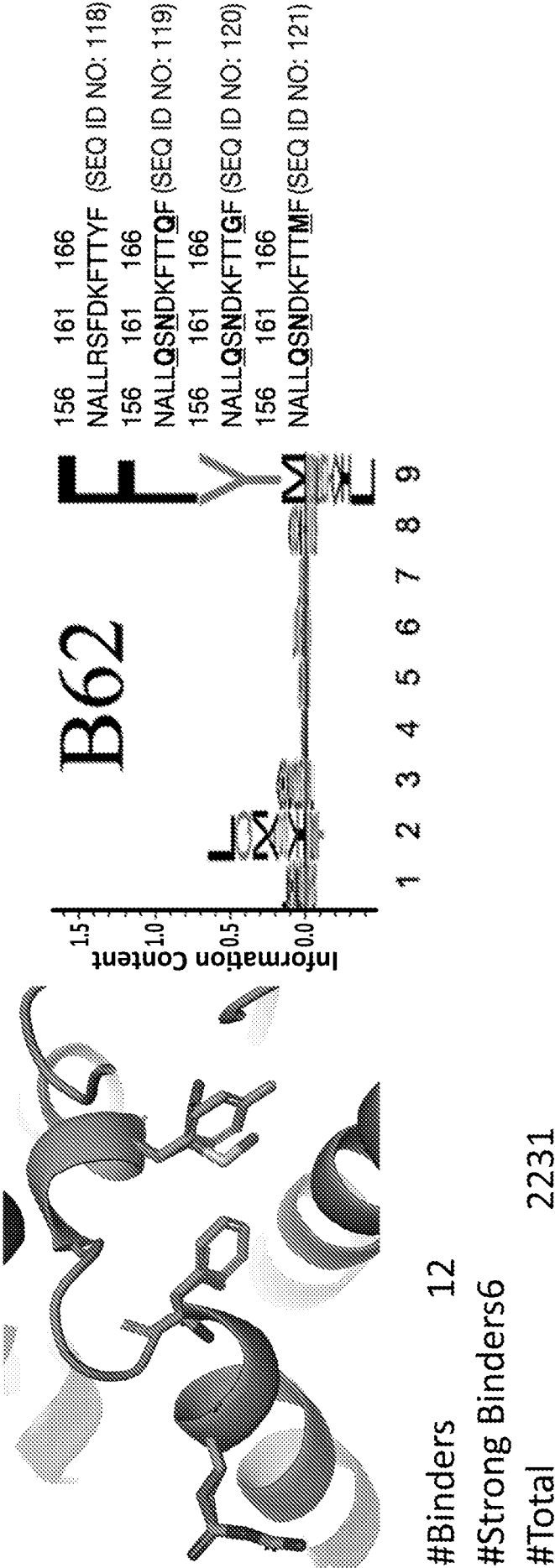
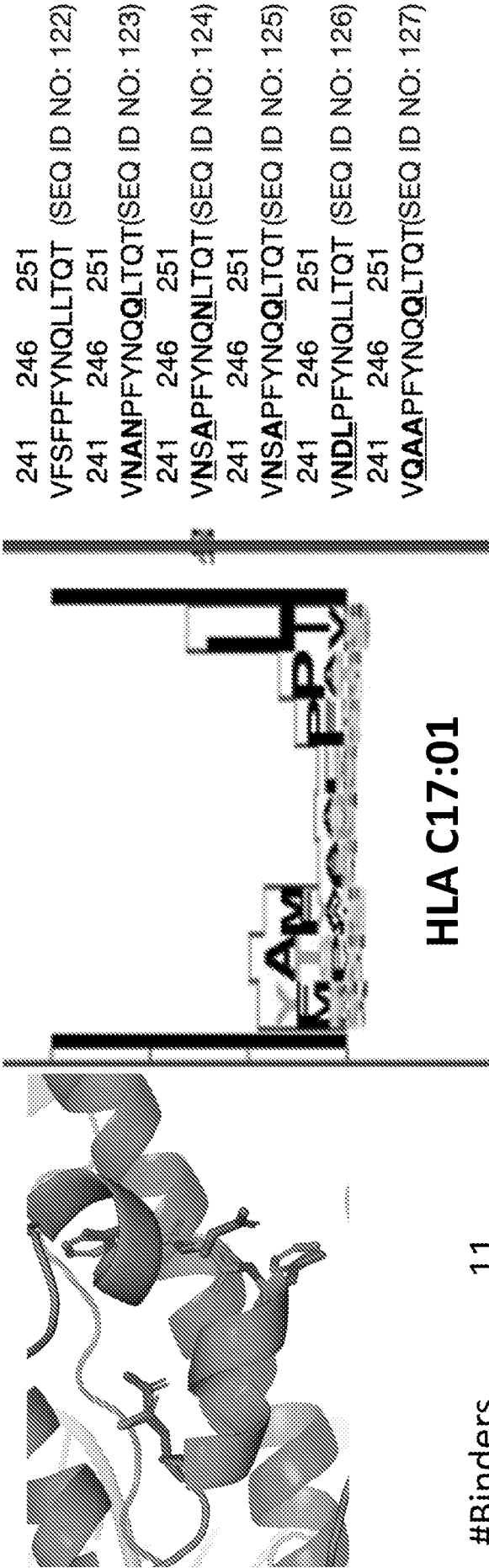


FIG. 20

p241 designs remove P1, P3, and P9 residues



#Binders	11
#Strong Binders2	
#Total	2231

FIG. 21

p505 mutants remove P2 and P9 pocket residues
(E->N,Q, F->N)

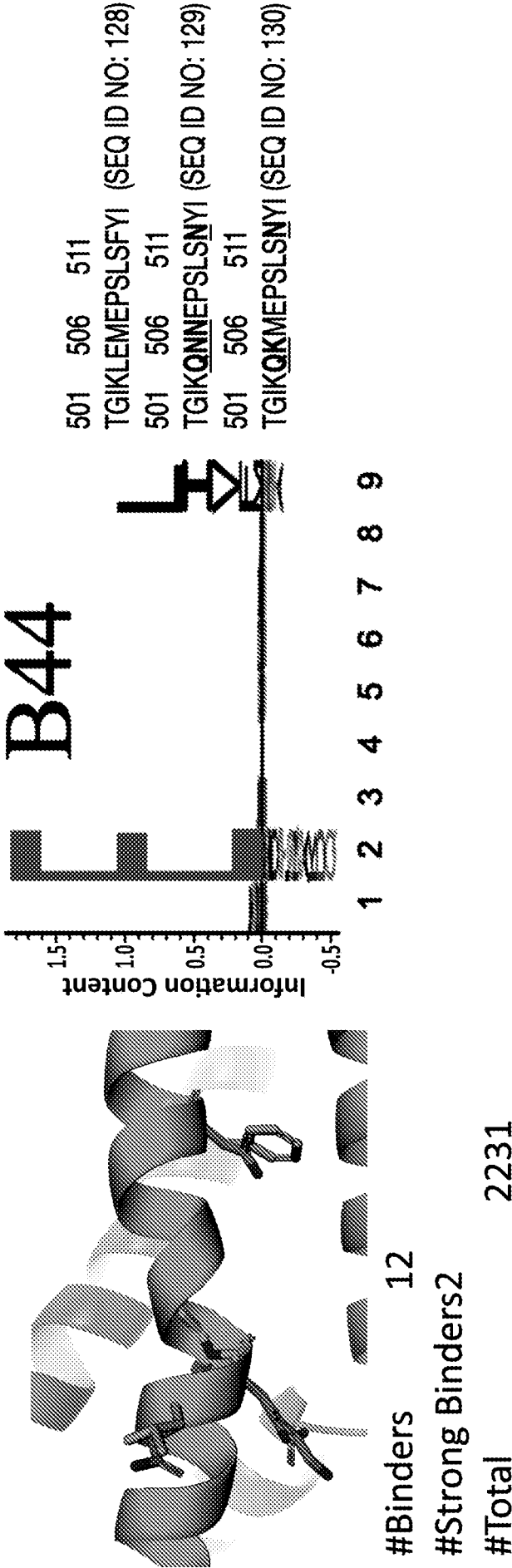
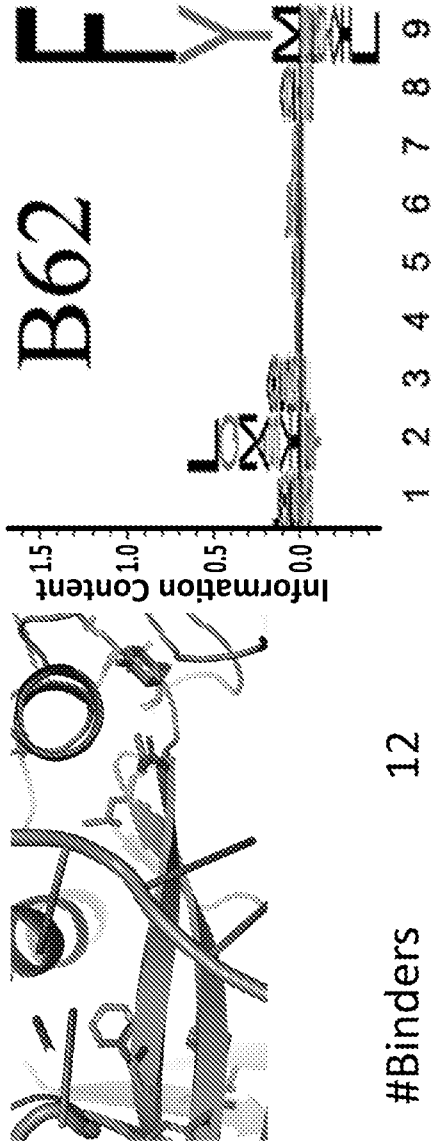


FIG. 22

p557 designs remove P2 and P9 residues
(V->Q, L->N,T)



#Binders 12
#Strong Binders3
#Total 2231

556	561	566	(SEQ ID NO: 131)
LFV	KNGLYY	LGIMI	
556	561	566	(SEQ ID NO: 132)
LFQ	KNGLYY	NGIMI	
556	561	566	(SEQ ID NO: 133)
LFQ	KNGLYY	TGIMI	
556	561	566	(SEQ ID NO: 134)
LGQ	KNGLYY	NGIMI	
556	561	566	(SEQ ID NO: 135)
LQQ	KNGLYY	NGIMI	

FIG. 23

p869 mutants remove P2, P9 pocket residues
(F->A,M, L->G,A)

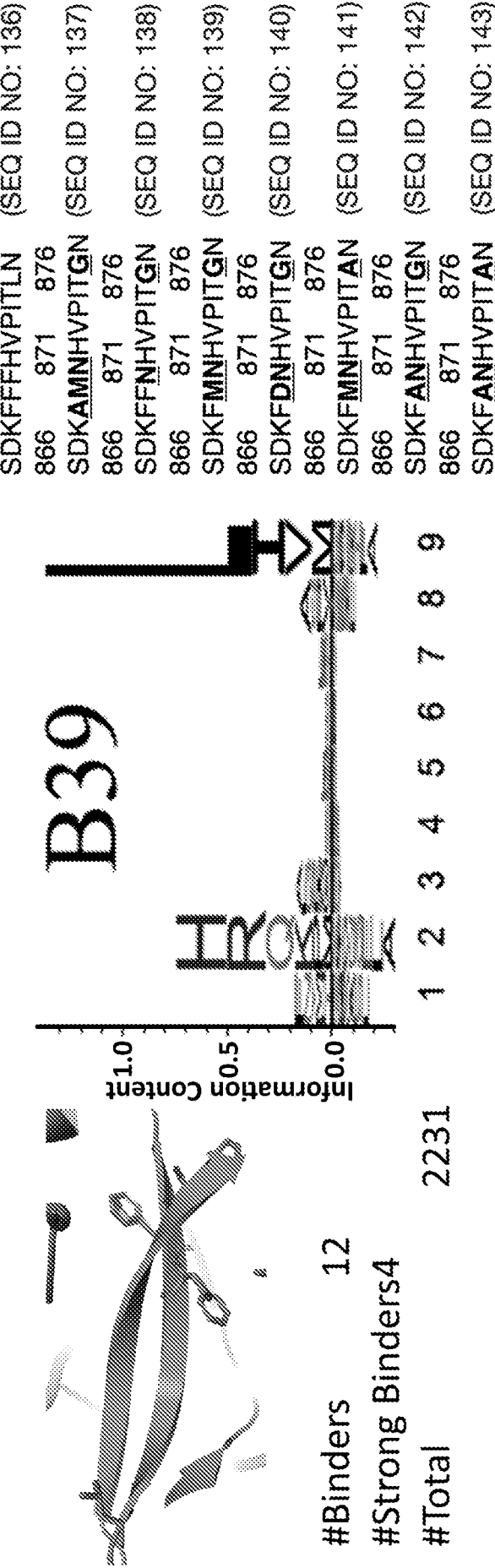


FIG. 24

asCas12a predicted epitope mutants

peptide	seq	Mutation	Nat Binders	Mut Binders	Nat Strong Binders	Mut Strong Binders	Nat Global Binders	Mut Global Binders	2231	2216
2	TQFE G FTNL	10N	13	0	5	0	0	2231	2216	2216
		SEQ ID NO: 144	13	1	5	0	0	2231	2221	2221
11	YQVSKTLRF	19D	12	0	3	0	0	2231	2230	2230
		SEQ ID NO: 145	12	0	3	0	0	2231	2224	2224
160	RSFDKFTTY	168G	12	3	6	0	0	2231	2221	2221
		SEQ ID NO: 146	12	4	6	0	0	2231	2216	2216
241	FSFPFY N QL	249S	11	2	2	0	0	2231	2226	2226
		SEQ ID NO: 147	11	2	2	0	0	2231	2222	2222
505	LEM E PSLSF	513S	12	1	2	1	1	2231	2230	2230
		SEQ ID NO: 148	12	1	2	0	0	2231	2220	2220
557	FVKNG L IY L	565A	12	6	3	0	0	2231	2226	2226
		SEQ ID NO: 149	12	4	3	0	0	2231	2224	2224
869	FFFHV P ITA	877A	12	6	4	0	0	2231	2224	2224
		SEQ ID NO: 150	12	0	4	0	0	2231	2219	2219

FIG. 25

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/022895

A. CLASSIFICATION OF SUBJECT MATTER

IPC: **A61K 48/00** (2024.01); **C07K 14/195** (2024.01); **C12N 15/52** (2024.01); **C12N 15/90** (2024.01); **C12N 9/22** (2024.01)
CPC: **C12N 9/22**; **A61K 48/00**; **C07K 14/195**; **C12N 15/52**; **C12N 15/907**; **C12N 2310/20**

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

See Search History Document

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

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2020/0199555 A1 (THE BROAD INSTITUTE INC. et al.) 25 June 2020 (25.06.2020) entire document	1-5
A	US 2020/0377871 A1 (ARIZONA BOARD OF REGENTS ON BEHALF OF ARIZONA STATE UNIVERSITY) 03 December 2020 (03.12.2020) entire document	1-5
A	US 2018/0319850 A1 (LONZA LTD) 08 November 2018 (08.11.2018) entire document	1-5



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
“D” document cited by the applicant in the international application
“E” earlier application or patent but published on or after the international filing date
“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
“O” document referring to an oral disclosure, use, exhibition or other means
“P” document published prior to the international filing date but later than the priority date claimed

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

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

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

“&” document member of the same patent family

Date of the actual completion of the international search

29 May 2024 (29.05.2024)

Date of mailing of the international search report

12 June 2024 (12.06.2024)

Name and mailing address of the ISA/US

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

Facsimile No. **571-273-8300**

Authorized officer

**MATOS
TAINA**

Telephone No. **571-272-4300**

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/022895

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)),
☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2024/022895

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: **6-24**
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).